

Towards Spatial Reuse in Future Wireless Local Area Networks: a Sequential Learning Approach

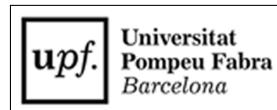
Francesc Wilhelmi Roca

TESI DOCTORAL UPF / ANY 2020

Director de la tesi

Boris Bellalta, Cristina Cano & Anders Jonsson

Departament of Information and Communication Technologies



Write here your dedication

Acknowledgments

This work has been partially supported by grants MDM-2015-0502, 2017-SGR-11888, by WINDMAL PGC2018-099959-B-I00 (MCIU/AEI/FEDER,UE), and by a Gift from the Cisco University Research Program (CG#890107) Fund.

Abstract

The Spatial Reuse (SR) operation is gaining momentum in the latest IEEE 802.11 family of standards due to the overwhelming requirements posed by next-generation wireless networks. In particular, the increasing traffic capacity and the number of concurrent devices compromise the efficiency of increasingly dense Wireless Local Area Networks (WLANs) and throw into question their decentralized nature. The SR operation, initially introduced by the IEEE 802.11ax-2021 amendment and further studied in IEEE 802.11be-2024, aims to increase the number of concurrent transmissions in an Overlapping Basic Service Set (OBSS), thus improving spectral efficiency. SR was initially defined as a distributed mechanism, but it is evolving towards coordinated schemes where different BSSs cooperate. Nevertheless, coordination entails communication and synchronization overhead, which impact the performance of WLANs that remains unknown. Moreover, the coordinated scheme is incompatible with devices using previous IEEE 802.11 versions, potentially leading to degrading the performance of legacy networks. For those reasons, in this thesis, we assess the viability of decentralized mechanisms for SR, and thoroughly examine the main impediments and shortcomings that may result from it. We aim to shed light on the future shape of WLANs concerning SR optimization and whether their decentralized nature should be kept or preferable to evolve towards coordinated and centralized settings. Given the challenges posed by the decentralized SR problem, we focus on Artificial Intelligence (AI) and propose using a class of sequential learning-based methods, namely Multi-Armed Bandits (MABs). The MAB setting suits the decentralized SR problem because it addresses the uncertainty caused by the concurrent operation of multiple devices (i.e., multi-player setting) and the lack of information resulting from it. MABs can potentially overcome the complex and varying spatial interactions among devices that result from modifying their sensitivity and transmit power. Nevertheless, the application of multi-agent sequential learning to communications raises several concerns that may compromise network devices' performance (definition of joint goals, time-horizon convergence, scalability aspects, non-stationarity, etc.). Our analysis of decentralized SR encompasses an in-depth study of the SR technology, infrastructure aspects for AI-enabled networking, and the suitability of the sequential learning approach.

Resum

To be written once the English version is definitive.

List of Publications

1. Wilhelm, F., Bellalta, B., Cano, C., & Jonsson, A. (2017, October). *Implications of decentralized Q-learning resource allocation in wireless networks*. In 2017 IEEE 28th annual international symposium on personal, indoor, and mobile radio communications (PIMRC) (pp. 1-5). IEEE.
2. Wilhelm, F., Cano, C., Neu, G., Bellalta, B., Jonsson, A., & Barrachina-Muñoz, S. (2019). *Collaborative spatial reuse in wireless networks via selfish multi-armed bandits*. Ad Hoc Networks, 88, 129-141.
3. Wilhelm Roca, F., Barrachina Muñoz, S., Bellalta, B., Cano Sandín, C., Jonsson, A., & Neu, G. (2019). *Potential and pitfalls of multi-armed bandits for decentralized spatial reuse in WLANs*. Journal of Network and Computer Applications, 2019, 127.
4. Barrachina-Muñoz, S., Wilhelm, F., Selinis, I., & Bellalta, B. (2019, April). *Komondor: a wireless network simulator for next-generation high-density WLANs*. In 2019 Wireless Days (WD) (pp. 1-8). IEEE.
5. Wilhelm, F., Muñoz, S. B., Cano, C., Selinis, I., & Bellalta, B. (2019). *Spatial Reuse in IEEE 802.11 ax WLANs*. arXiv preprint arXiv:1907.04141.
6. Wilhelm, F., Barrachina-Muñoz, S., & Bellalta, B. (2019, October). *On the Performance of the Spatial Reuse Operation in IEEE 802.11 ax WLANs*. In 2019 IEEE Conference on Standards for Communications and Networking (CSCN) (pp. 1-6). IEEE.
7. Wilhelm, F., Barrachina-Munoz, S., Bellalta, B., Cano, C., Jonsson, A., & Ram, V. (2020). *A Flexible Machine-Learning-Aware Architecture for Future WLANs*. IEEE Communications Magazine, 58(3), 25-31.
8. Wilhelm, F., Carrascosa, M., Cano, C., Ram, V., & Bellalta, B. (2020). *Usage of Network Simulators in Machine-Learning-Assisted 5G/6G Networks*.

Contents

1	INTRODUCTION	1
1.1	Motivation	1
1.2	Contributions	2
1.2.1	Contributions to the Research Community	2
1.2.2	Contributions to Standardization	3
1.3	Document Structure	3
2	SPATIAL REUSE IN IEEE 802.11 WLANS	5
2.1	Related Work	5
2.2	Spatial Reuse in IEEE 802.11ax	6
2.2.1	OBSS/PD-based Spatial Reuse	7
2.2.2	Parametrized Spatial Reuse	7
2.3	Spatial Reuse in IEEE 802.11be	8
3	MACHINE LEARNING IN WLANS	11
3.1	Architectural Aspects for ML-Enabled Communications	12
3.1.1	Computation Paradigms	12
3.1.2	Machine-Learning-Aware Network Architecture	13
3.2	Multi-Armed Bandits in Communications	14
3.3	Multi-Armed Bandits for Decentralized Spatial Reuse: Between ML and Game Theory	18
4	METHODOLOGY AND ENABLERS	21
4.1	Spatial Reuse through Continuous Time Markov Networks	21
4.1.1	IEEE 802.11ax OBSS/PD-based Spatial Reuse	23
4.1.2	IEEE 802.11be Coordinated Spatial Reuse	23
4.2	Simulation of Sequential Learning for Spatial Reuse	24
4.2.1	Spatial Reuse Implementation	24
4.2.2	Agents-based Implementation	26
5	MAIN FINDINGS	29
6	CONCLUDING REMARKS	37

Chapter 1

INTRODUCTION

1.1 Motivation

The Institute of Electrical and Electronics Engineers (IEEE) 802.11 family of protocols for wireless local area networks (WLANs) was first released in 1997 as a novel solution for physical (PHY) and medium access control (MAC) layers. Since that date, the standard has evolved to sustain the increasing user requirements in terms of capacity, load, and coverage, as well as serve for different purposes (e.g., mesh networking, security-enhanced communications, channel measurement, etc.). The set of new and improved capabilities have been captured along the time in the plethora of amendments that followed the initial 802.11-1997 standard (e.g., 802.11b, 802.11g, 802.11h, etc.).

Looking ahead, the next generation of WLAN standards is expected to revolutionize the telecommunications and converge along with 5G systems and beyond to expand to multiple domains, such as light communications (IEEE 802.11bb), Internet of Things (IEEE 802.11ah), vehicle-to-everything (IEEE 802.11bd), or next-generation positioning (IEEE 802.11az). One of the most influential amendments is the IEEE 802.11ax-2021 (11ax) amendment for High Efficiency (HE) WLANs [1–3], which primary goal is to enhance network efficiency in ultra-dense deployments to grant high capacity (up to 10 Gbps). In particular, the 11ax (commercially known as Wi-Fi 6) includes a set of unprecedented techniques, such as Orthogonal Frequency Division Multiple Access (OFDMA), Downlink/Uplink Multi-User Multiple-Input-Multiple-Output (DL/UL MU-MIMO), and Spatial Reuse (SR), for addressing the broad range of issues arisen from high-density scenarios [4].

This thesis focuses on the SR operation, which started with IEEE 802.11ax and is now evolving in the IEEE 802.11be. SR aims to enhance spectral efficiency by increasing the number of parallel transmissions in high-dense deployments. To this end, SR proposes a mechanism to ignore transmissions whose source is a device belonging to a different Basic Service Sets (aka inter-BSS transmissions). As a result, a less restrictive carrier sense threshold is used for inter-BSS transmissions, referred to as Overlapping BSS Packet Detect (OBSS/PD) threshold. To promote fairness, SR also incorporates a mechanism that limits the transmit power of the new transmissions resulting from a less restrictive OBSS/PD threshold (so that primary transmissions are not affected). Table 1.1 summarizes the potential effects and implications of adjusting the sensitivity threshold and the transmit power in WLANs.

To address the underlying complexity of SR, we study the application of Artificial Intelligence (AI) mechanisms for automatically adjusting both the sensitivity and the transmit power

Table 1.1: Effects and implications of adjusting the sensitivity threshold and the transmit power in IEEE 802.11 WLANs.

	Data rate	Channel access probability	Generate starvation probability	Hidden-node probability	Exposed-node probability
Sensitivity ↑	-	↑	↑	↑	↓
Tx. power ↑	↑	-	↑	↓	↑

of wireless devices. AI is gaining momentum in telecommunications – it is in fact expected to be pervasively included as part of the network operation in 6G systems [5–7] – due to its ability to exploit complex characteristics from data, thus allowing to solve problems that are hard to solve by hand-programming.

1.2 Contributions

1.2.1 Contributions to the Research Community

In light of the importance of SR for future wireless networks and the current evolution of communications towards AI-enabled systems, in this thesis, we study the potential application of Machine Learning (ML) for addressing the challenges raised by SR. In particular, we aim to shed light on the potential gains of SR and devise its intersection with AI. The contributions of this thesis are summarized next:

1. We study state-of-the-art solutions for improving spectral efficiency in wireless networks. In particular, we focus on methods based on sensitivity adjustment and power control. Then, we narrow the scope to IEEE 802.11-based SR solutions, mostly oriented to 11ax systems.
2. We provide an in-depth overview of the SR operation included in the IEEE 802.11ax amendment. Besides, we devise the potential evolution path of the SR technology in IEEE 802.11be and beyond.
3. We analytically model the SR operation and study the new kind of inter-networks interactions resulting from it. Thanks to the model, we can deeply understand the implications of applying SR.
4. We provide a simulation-based implementation of the SR operation, from which we extract results of its performance gains in future dense WLANs. The presented simulation tool allows for the inclusion of new blocks cost-effectively and serves to devise the potential of new technologies such as coordinated spatial reuse. Besides, our simulator allows characterizing high-density deployments in affordable simulation time.
5. We propose several ML-based solutions to address the SR problem in decentralized WLAN deployments. The simulation-based implementation of these methods allows us to study the performance gains compared to default carrier sensing approaches. In particular, we show that the concurrent learning operation may lead to suboptimal configurations in aggregate performance. Nevertheless, we also show that decentralized learning can help to mitigate unfairness in wireless networks when cooperation is enabled.

6. We study implications that decentralized ML solutions have on the operation of WLANs. In particular, we focus on the game-theoretic setting caused by the concurrent devices attempting to learn the best SR configuration. Besides, we delve into practical implementation aspects related to the communication limitations for cooperative approaches, the amount of information available for learning, and the dynamism of wireless networks.
7. We delve into architectural aspects to enable future ML-aware networks. Because of the promising performance gains that ML can provide to networking systems, its actual integration is currently a topic that is attracting a lot of attention. Special emphasis is being put on data handling and flexible interfaces, which are meant to address the issues related to data storage, data exchange, and data processing.

1.2.2 Contributions to Standardization

Apart from the contributions related to the research on SR and AI, we have participated in many standardization initiatives. Specifically, the output contributions in the context of International Telecommunications Union Telecommunication Standardization Sector (ITU-T) are as follows:

1. As part of the ITU's architectural framework for ML-enabled networks, we developed a specification on an isolated domain for training, testing, and evaluating ML models for communications.
2. We contributed to the group through the following topics: *i*) considerations when applying ML in heterogeneous environments in which both intelligent and non-intelligent (legacy) devices coexist, *ii*) practical usage of simulators for generating synthetic data sets that serve for training ML models, and *iii*) a use-case realization of the ITU's ML-aware architecture for IEEE 802.11 WLANs.
3. We provided a data set to address the problem of channel bonding in future WLANs through ML.¹ This data set has been developed in the context of the ITU-T AI Challenge - 2020, which aims to boost innovation in the integration of AI/ML into 5G networks and beyond.

1.3 Document Structure

This thesis is a compendium of articles resulting from the research activity on ML's application to address SR in IEEE 802.11 WLANs. We refer to the publications on page ix as Paper #1 through Paper #8. Besides the list of publications (attached at the end of this document), a monograph is provided to introduce the research topic and give some background on the same. This document is structured as follows. Chapter 2 surveys SR techniques in wireless networks, overviews the IEEE 802.11ax SR operation, and discusses the evolution of SR in future amendments. Chapter 3 provides insights on the intersection between ML and wireless communications, including architectural aspects and state-of-the-art applications. Then, the SR problem

¹All the details on the problem statement can be found at https://www.upf.edu/web/wnrg/ai_challenge

is formulated through Multi-Armed Bandits. Chapter 4 introduces the analytical and simulation tools used for performance evaluation. The main finding of this thesis are summarized in Chapter 5, and concluding remarks are provided in Chapter 6.

Chapter 2

SPATIAL REUSE IN IEEE 802.11 WLANS

In this Chapter, we describe the SR operation and survey the related work, ranging from solutions for sensitivity and transmit power adjustment in wireless networks, to specific IEEE 802.11 mechanisms. Then, we overview the IEEE 802.11ax SR operation and discuss the next steps being taken by the Task Group 802.11be (TGbe) to make this technology evolve.

2.1 Related Work

Improving medium utilization through SR has been extensively studied for sensitivity and transmit power adjustment in different domains such as multi-hop networks [8, 9], cellular networks [10], and IEEE 802.11 WLANs [11]. SR can be realized through beamforming/null-steering [12], OFDMA-based multi-user scheduling [13], MU-MIMO transmissions [14], sensitivity adjustment [11], and transmission power control [15]. In this thesis, we focus on sensitivity and transmit power adjustment, which does not require additional infrastructure and promises substantial performance gains concerning the lack of efficiency of the current IEEE 802.11 channel sharing methods [9].

Sensitivity and transmit power adjustment has been applied in different manners to address multiple problems (improve capacity, boost fairness, save energy, etc.). Figure 2.1 shows a categorization of these SR techniques according to the optimization goal and their implementation. In particular, we find CS/CCA adaptation mechanisms, power control, and joint CS/CCA adaptation and power control. While some of the works focus their attention on modeling and analyzing the effects of tuning sensitivity and transmit power in wireless networks, others provide different optimization mechanisms. Iterative methods have been typically applied to decentralized and varying deployments (see, for instance, the work in [16]). Apart from that, pre-defined solutions are based on a preliminary analysis of the implications of sensitivity adjustment and power control in particular network settings. In [17], for instance, the authors provide a Carrier Sense Threshold (CST) adaptation mechanism based on the location of APs. In particular, a position-CST table is derived from a preliminary analysis of the conditions that allow APs to transmit concurrently without suffering losses. Finally, global solutions are provided for deployments that can be fully controlled at a single point (availability of complete information or coordinated setting). This is the case of the work in [18], where DL is used to maximize the sum-rate under Quality of Service (QoS) constraints in multi-user power control.

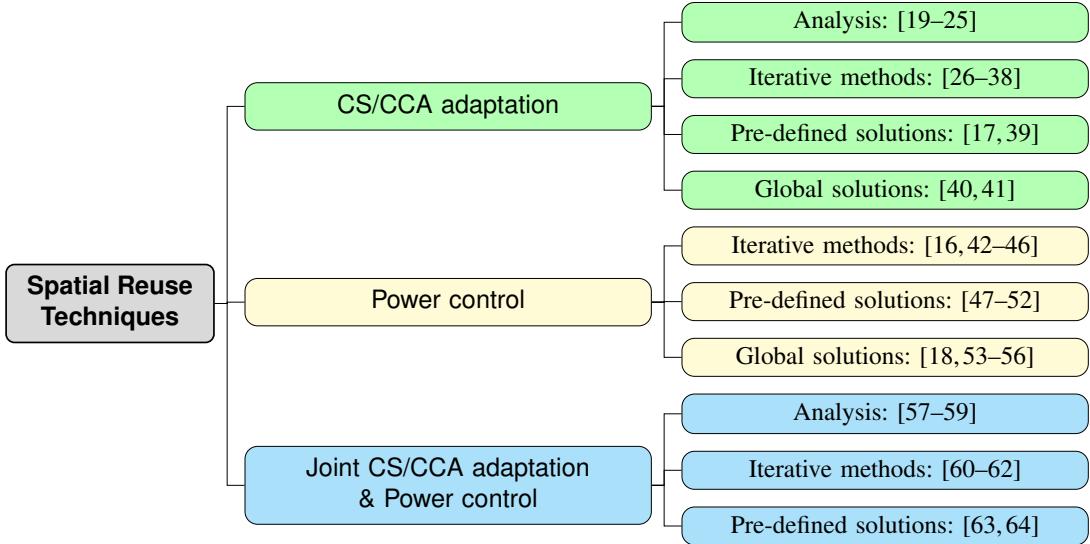


Figure 2.1: Spatial reuse techniques in wireless networks.

Concerning IEEE 802.11ax WLANs, the Dynamic Sensitivity Control (DSC) scheme [34] was the first proposal for adapting the sensitivity of devices in an OBSS, but it was never incorporated in any amendment. Roughly, the DSC mechanism iteratively increases or reduces the sensitivity of a Station (STA) in a decentralized manner, based on the average perceived RSSI. Intuitively, DSC aims to increase the sensitivity level at STAs that are close to the Access Point (AP) to avoid contention, while reducing this threshold for STAs at the cell edge, which is useful for lowering collisions by hidden nodes. While DSC was initially proposed for tuning the Physical Carrier Sense (PCS) threshold, it was later introduced as a method for optimizing the OBSS/PD [65]. Due to its promising potential, DSC’s performance has been extensively studied in multiple scenarios and combined with other mechanisms [66–76].

Apart from DSC, other solutions for tuning the sensitivity threshold have been proposed in [35–38]. First, [35] proposed tuning the transmission power based on the Expected Transmission Count (ETX) metric, which has been widely used in wireless sensor networks. The authors in [36] provided an iterative method whereby the OBSS/PD is progressively updated, based on the Received Signal Strength Indicator (RSSI) at STAs. Similarly, [37] proposed the RSSI to OBSS threshold (RTOT) method, whereby the OBSS/PD value used by an STA is derived from the RSSI received from its AP (which is used as an indicator of the distance). Although this method is designed to deal with network dynamics (the OBSS/PD varies according to the RSSI), a static margin value is used for selecting the OBSS/PD. As for DSC, the rigidity of the margin value may lead to not finding the optimal solution in some scenarios. Finally, the Interference-based Dynamic Channel Algorithm (IB-DCA) was proposed in [38], whereby STAs exchange the expected RSSI so that the transmit power is globally adjusted, rather than applying the OBSS/PD.

2.2 Spatial Reuse in IEEE 802.11ax

The IEEE 802.11ax SR operation includes two different mechanisms: *i*) OBSS/PD-based SR, for decentralized settings, and *ii*) Parametrized SR (PSR), for scheduled uplink transmissions.

Both mechanisms are based on BSS coloring, whereby HE devices can quickly determine the source of the detected transmissions. In **Paper #5**, we provided an exhaustive overview and tutorial of the IEEE 802.11ax SR operation, which we briefly describe next.

2.2.1 OBSS/PD-based Spatial Reuse

In OBSS/PD-based SR, an HE STA can use a less restrictive OBSS/PD threshold when detecting inter-BSS transmissions, thus increasing the probability of ignoring them and accessing the channel. In case of initiating a transmission due to OBSS/PD-based SR (an SR-based TXOP is gained), an HE STA must regulate the transmit power it uses. The maximum allowed transmission power (TX_PWR_{\max}) depends on the selected OBSS/PD threshold, and is given by:

$$\text{TX_PWR}_{\max} = \text{TX_PWR}_{\text{ref}} - (\text{OBSS/PD} - \text{OBSS/PD}_{\min}),$$

where $\text{TX_PWR}_{\text{ref}}$ is the reference transmit power (depends on the transmitter's antenna capabilities) and OBSS/PD_{\min} is the minimum OBSS/PD (set to -82 dBm).

Figure 2.2 sketches an example of the OBSS/PD-based SR mechanism in which an HE device (i.e., AP_A) ignores inter-BSS transmissions (i.e., AP_B) by applying the OBSS/PD threshold, which allows initiating a simultaneous transmission with limited transmit power.

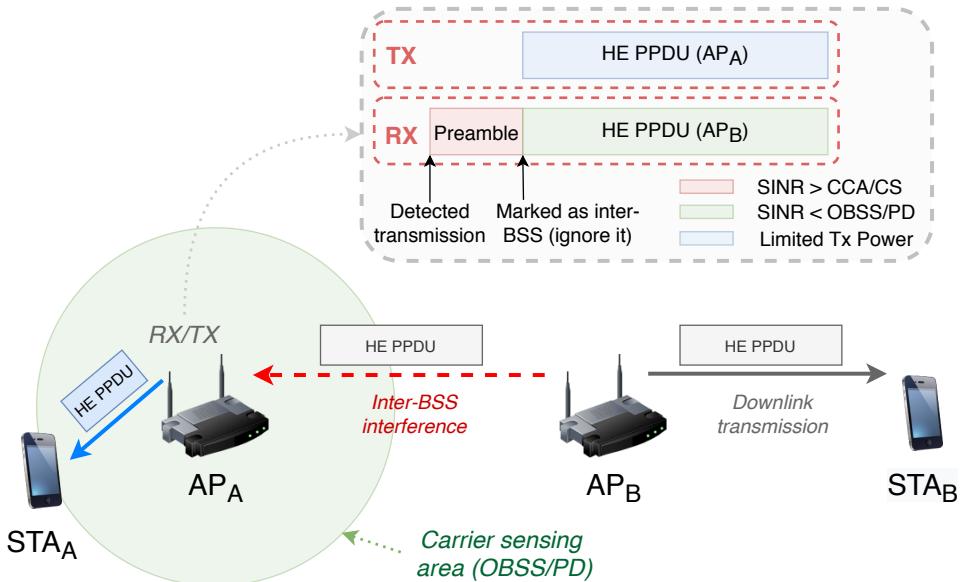


Figure 2.2: Example of OBSS/PD-based SR in a toy scenario.

The performance gains of OBSS/PD-based SR operation have been previously analyzed in [77–80]. In **Paper #5** and **Paper #6** we also provide a performance evaluation of 11ax SR. Unlike the previous literature, our performance evaluation is focused on the DL rather than the UL and puts particular emphasis on the analysis of the delay. Moreover, our works target the latest draft (D4.0) of the IEEE 802.11ax amendment.

2.2.2 Parametrized Spatial Reuse

Unlike for OBSS/PD-based SR, PSR aims to exploit Triggered-Based (TB) UL transmissions to carry out SR. Depending on the role of nodes participating in the PSR operation, we find two

types of devices: *sharing* (the ones initiating TB transmissions and indicating support for the PSR operation) and *shared* (the ones taking advantage of the PSR opportunities from detected TB transmissions).

For the sake of detecting PSR opportunities, shared devices must check whether their intended transmit power meets the requirements indicated in TB Physical Layer Conformance Procedure (PLCP) Protocol Data Unit (PPDU) from sharing devices. These requirements are based on the maximum level of interference supported by the sharing device. In particular, the intended transmit power cannot exceed the following value:

$$\text{TX PWR}_{\max} = \text{TX PWR}_{\text{AP}} + I_{\text{AP}}^{\max} - RPL,$$

where $\text{TX PWR}_{\text{AP}}$ is the normalized transmit power in dBm at the output of the antenna connector, I_{AP}^{\max} is a normalized value in dB that captures the maximum allowed interference at the sharing device,¹ and Received Power Level (RPL) is measured from the legacy portion of the TF (i.e., from PHY headers).

The PSR operation is sketched in Figure 2.3 for a toy scenario. As shown, the sharing device (i.e., AP_B) schedules a UL TB transmission by sending a TF, inspected by the shared device (i.e., AP_A) to detect a PSR-based TXOP.

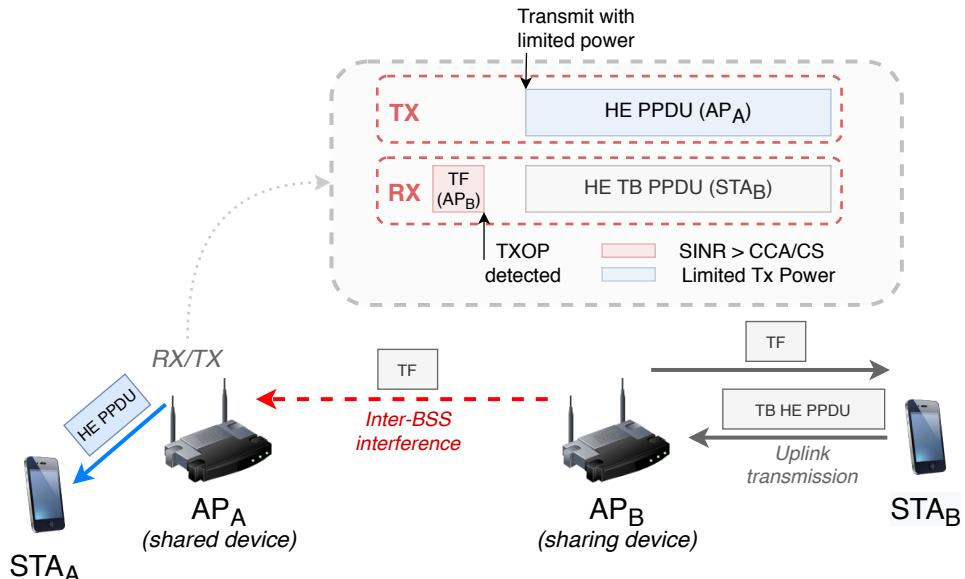


Figure 2.3: Example of PSR in a toy scenario.

The potential latency gains of PSR have been analyzed in [81].

2.3 Spatial Reuse in IEEE 802.11be

The SR operation included in the 11ax has been shown to provide significant gains for cell-center devices (i.e., devices close to the AP) but lacks applicability in cell-edge users (i.e., devices far from the AP) [82]. As a result, the 11be is working on Coordinated SR (CSR)

¹ I_{AP}^{\max} is computed as the target RSSI indicated in the TF minus the minimum SNR granting a 10% PER (a safety margin is also included not to exceed 5 dB).

[83], a cooperative scheme whereby BSSs exchange information (e.g., the acceptable level of interference supported by the different devices) to further enhance the quality of the parallel transmissions achieved through SR. Apart from that, SR is considered to converge with other technologies such as OFDMA [84] and beamforming/null steering [85]. Notice that multi-AP coordination (e.g., coordinated and joint transmission) is one of the main topics that have been so far discussed by IEEE 802.11 task groups [86]. Besides CSR, the main applications of multi-AP coordination are coordinated beamforming (CBF) [87] and coordinated OFDMA [88].

Concerning CSR (or Co-SR), it aims to improve the quality of the simultaneous transmissions that can occur due to the SR operation. In particular, the transmit power of secondary transmissions takes into account the maximum level of interference of the target devices whose transmissions are sought to be held. Co-SR is a natural extension of the SR scheme under the multi-AP operation framework and can be implemented with relatively low added complexity. At this stage, the CSR operation is considered to be built upon the following phases:

- **Preparation:** this phase includes capability announcement (via Beacon or management frames) and monitoring. The AP can request STAs to report the measured RSSI from the neighboring APs and the associated AP so that the Downlink Acceptable Receiver Interference Level (DL ARIL) can be computed. The DL ARIL is used to assess the feasibility of coordinated transmissions and the required transmit power limitations to carry them out.
- **Setup:** the AP winning a TXOP selects the candidate shared AP and indicates the maximum DL ARIL for the following scheduled DL transmission. Then, the shared AP responds to the sharing AP if the channel state is idle.
- **Communication:** the exchange of packets under CSR starts with a trigger frame sent by sharing APs, which choose the set of shared APs to transmit concurrently. The trigger includes information such as the transmission duration, the maximum transmit power allowed or the resource allocation for ACKs. Then, data and ACK packets are exchanged between nodes belonging to the authorized BSSs. The exchange of packets allows the shared APs to update measurements such as the DL ARIL. Finally, the ACK transmission from the STAs received CSR data frame can be performed by UL-OFDMA. However, this approach requires establishing a pre-agreement on the division of the frequency resources.

The abovementioned phases are illustrated in Figure 2.4 for two APs belonging to different BSSs. As it can be noted, the main challenge of CSR lies in data acquisition, which is mainly achieved during the monitoring phase. This is a critical aspect, especially for highly dynamic deployments. The trade-off between the necessary overhead and the potential gains of coordination hinders CSR's actual performance gains.

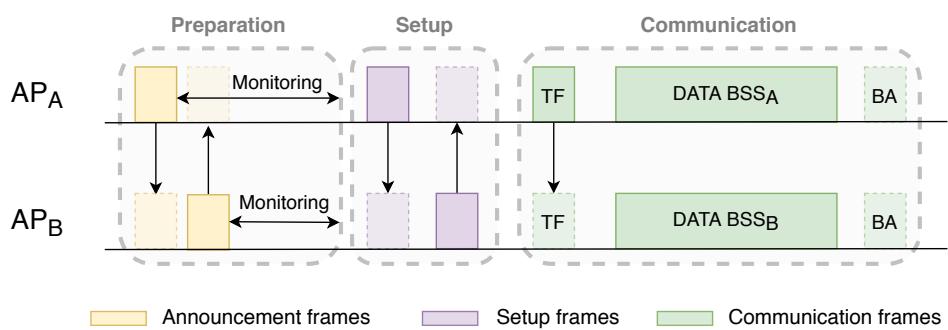


Figure 2.4: Example of the phases considered for IEEE 802.11be CSR.

Chapter 3

MACHINE LEARNING IN WLANS

ML is meant to empower a computational system for learning, based on experience, so that unseen situations can be properly handled without explicitly being programmed. Concerning wireless communications, the application of ML reveals a big potential because of the following aspects:

- First, there is a huge amount of unexploited data generated at both infrastructure and user levels, which could be extremely useful for identifying and exploiting patterns that help at improving network performance.
- Second, the increasing complexity of wireless communications problems (e.g., a massive number of concurrent users, ultra low latency applications, very high bandwidth requirements, etc.) complicates the design of hand-crafted solutions. Moreover, complex non-linear phenomena of communications systems (e.g., channel effects, varying traffic requirements, hardware imperfections, etc.) add further complexity. In this regard, ML can learn sophisticated solution strategies from data (model-based vs. data-driven approach), thus granting the ability to adapt to wireless communications problems and characteristics.
- Apart from the underlying complex characteristics of problems (e.g., wireless channel), communications systems are built based on functional blocks, each executing well defined and isolated functions (e.g., rate selection, channel allocation, etc.). While individual functions can be separately optimized, their joint operation may further increase end-to-end complexity, thus hindering globally optimized solutions. ML can therefore help to achieve optimal control by addressing the added combinatorial complexity of modularized communications systems.

Henceforth, ML is expected to overcome the systemic complexity inherited from novel use cases like Vehicle to Everything (V2X) communications, Machine Type Communications (mMTC), and Ultra-Reliable Low-Latency Communication (uRLLC). In particular, the inherent flexibility of ML for automatically learning diverse situations can address heterogeneous scenarios, including mobility, a massive number of devices, and varying throughput and latency requirements. Because of its high potential for solving complex problems in communications, ML has been applied to a plethora of fields. We address the interested reader to the surveys in [89–97] and references therein.

3.1 Architectural Aspects for ML-Enabled Communications

3.1.1 Computation Paradigms

Most popular ML approaches work with batch data, which allows providing powerful solutions and drawing insightful conclusions from typically large datasets. Batch learning methods require certain perennity of data, thus lacking responsiveness and not suiting real-time applications. Notice that large training datasets need significant computational resources to carry out time-consuming tasks. In this regard, online learning mechanisms emerge as a suitable tool to address the non-stationarity of problems whose underlying patterns cannot be fully learned on time. Online learning allows providing fast solutions rather than seeking for optimality (fast and moderate improvement vs. slow optimization). The adoption of either batch or online learning mechanisms in telecommunications requires the network infrastructure to accommodate ML-oriented tasks such as data collection, data processing, data analysis, and decision-making [98–100]. The procedures mentioned above can be held at different parts of the network, which depends on the network architecture. Batch learning mostly suits to centralized settings where training tasks are held at a single point (e.g., a data center). Centralization allows deriving global ML models encompassing data acquired from multiple sources (e.g., nodes in a network) and even from different domains (e.g., inter-operator data). Furthermore, online learning matches better with decentralized settings, which typically lack powerful equipment but, in turn, allow reducing the complexity of centralized problems by casting them into individual learning problems. Figure 3.1 depicts the centralized and decentralized network architectures for ML.

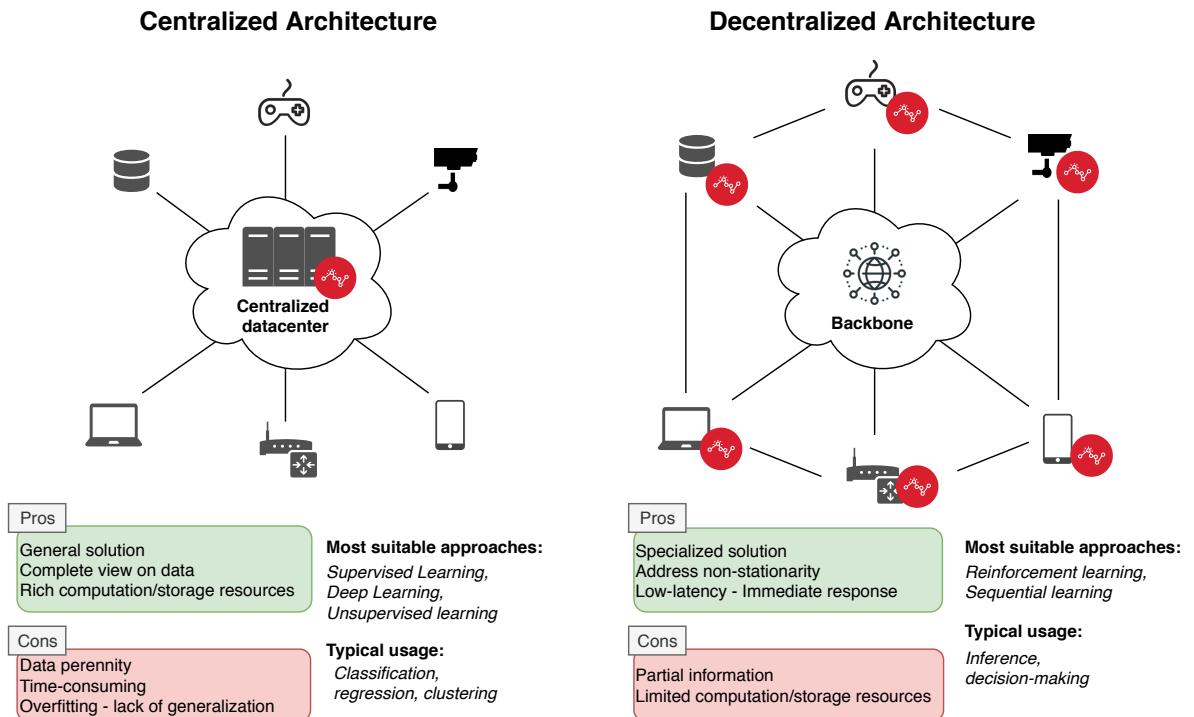


Figure 3.1: High-level representation of centralized and decentralized architectures for the application of ML mechanisms in networks.

In-between centralized and decentralized settings, we find mixed architectures where other learning mechanisms can be applied. For instance, transfer learning (storing knowledge gained while solving one problem and applying it to a different but related problem) [101] and federated learning (collaborative training starting from a general model to better fit different contexts and situations) [102, 103] can accelerate decentralized approaches by sharing some information among agents (e.g., common solution strategies that can then be personalized for each agent). However, the successful application of these kinds of approaches is tightly tied to the communication capabilities of the implied devices, which defines the degree of cooperation among nodes in a network.

In communications, batch learning has been shown to suit problems related to the core network or involving higher layers of the protocols stack. For instance, Deep Learning (DL) has been broadly applied for predicting periodical patterns of network traffic [104–106] or user mobility [107–109]. Alternatively, online learning suits better to problems related to the access network and PHY/MAC layers. In this regard, techniques such as Reinforcement Learning (RL) or sequential learning are widely employed for PHY/MAC optimization. Some examples are resource allocation [110], edge computing [111, 112], or MIMO optimization [113]. Notice that time-consuming mechanisms requiring a heavy workload such as Neural Networks (NNs) can be barely applicable to non-stationary problems, which would make trained ML models obsolete very fast.

3.1.2 Machine-Learning-Aware Network Architecture

The first steps towards ML-enabled networking have been recently materialized through the virtualization of networks, i.e., Network Function Virtualization (NFV) and Software-Defined Networks (SDN) paradigms. The fact is that network virtualization provides an unprecedented elasticity for the management and operation of network resources, which were previously limited by traditional hardware-based components. Besides, inter-operator coordination is boosted for bringing the ML operation to a macro-scale level, allowing them to share a vast amount of information and computation resources. These improvements favor the implementation of ML approaches such as federated learning, whereby network infrastructure requires elements to communicate for carrying out a joint decentralized training procedure in an efficient and scalable way.

Currently, the main standardization bodies are progressing on the definition of future ML-aware network architectures. In particular, the following progress has been made:

- The 3rd Generation Partnership Project (3GPP) is currently working on the integration of data analytics to network functions [114].
- ETSI groups on Experiential Networked Intelligence (ENI) and Zero-touch network and Service Management (ZSM) are actively studying the integration of AI to networks [115].
- ITU-T has released a set of specifications on a *Unified architecture for 5G and beyond* [116, 117]. Remarkably, ITU's standardized architecture provides a common nomenclature for ML-related mechanisms so that interoperability with other networking systems is achieved.

In **Paper #7**, we proposed a realization of the ITU-T ML-aware architecture for IEEE 802.11 WLANs. Through the definition of ML components and management functions, the

ITU-T architecture provides the necessary flexibility for fulfilling different use case requirements, ranging from centralized solutions to local-learning approaches. This is particularly suitable for Wi-Fi deployments, which may lack powerful centralized equipment for gathering data and processing it (e.g., residential scenarios). In these cases, it is imperative to flexibly instantiate ML pipeline nodes, thus adapting to the set of available resources and capabilities from each use case.

The ITU-T ML-aware architecture defines a set of components and procedures to enable the usage of ML models in networking operations. In particular, the following main components are defined:

- **ML pipeline:** set of logical nodes that are combined to form an ML application in a network. ML pipeline nodes are responsible for data collection, data processing, ML model application, and output distribution.
- **ML management and orchestration:** logical node to manage and orchestrate ML pipelines according to use case specifications.
- **ML sandbox:** isolated domain for training, testing, and evaluating ML pipeline nodes before being deployed in a production environment.
- **Data handling blocks:** framework to handle ML data collection, ML data processing, and ML data output.

Figure 3.2 shows the high-level architectural components defined in [116]. As illustrated, the modularized ML operation may take place at any point in the network. As a result, the chaining and deployment of ML pipeline nodes are flexible and depend on the use case.

Among the architectural components of the ITU-T architecture, we paid particular attention to the ML sandbox. In this regard, **Paper #8** delves into the potential usage of network simulators to enhance the reliability of ML for communications. In particular, we showcased that network simulators can be used to validate the performance of ML methods before applying them to live networking systems. To that purpose, we provided a proof-of-concept testbed implementation of a residential WLAN deployment, which was empowered with an ML model trained in a network simulator. Besides validation purposes, simulators are useful for generating synthetic data sets that can be used for training (e.g., predicting unforeseen situations from which there are no real data).

3.2 Multi-Armed Bandits in Communications

The *learning by experience* characteristic of sequential learning suits WLANs well because it allows addressing complex partial information problems. The fact is that WLANs pose a set of specific challenges resulting from their multiple deployment modes (e.g., campus network, residential usage) and their typical decentralized nature. Although WLANs can count with plenty of data to be used by ML methods in large and planned deployments, we find other types of scenarios (e.g., residential deployments) where data may fail to be integrated at a single point due to potential computation, storage, or communication limitations (e.g., end devices may have low-throughput connections and be intermittently available). Besides, the varying nature of WLAN deployments (e.g., due to STAs mobility) leads to a high non-stationarity.

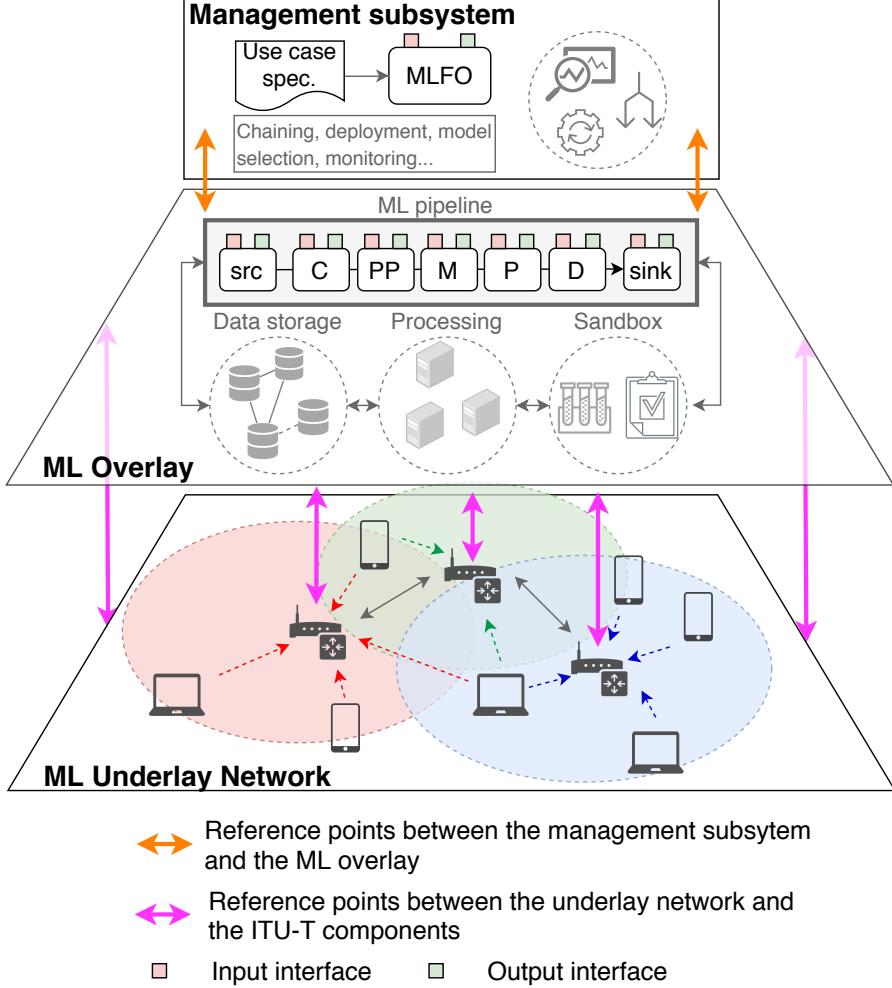


Figure 3.2: ITU-T’s high-level architecture for future ML-aware networks.

In the MAB problem, [118, 119], and as for classical RL formulations, an agent (or learner) interacts with the environment to accumulate knowledge that allows adapting to underlying changes in the reward distribution over time. The maximization of a long-term goal [120] requires finding an equilibrium between exploration (improve your knowledge) and exploitation (obtain the maximum profit based on your experience).

Formally, a learner sequentially picks actions $a_t \in K$ and observes their reward vector r_t over a time horizon T . Typically, the reward is granted by what is known as the environment, which may be of diverse nature (e.g., stochastic distribution, adversarial payoff). In the bandits setting, rewards are generated by hidden distributions, and their value is only revealed once the corresponding arm is played. Bandits differ from *partial* and *full* information settings that reveal the reward of a set or all the possible actions, respectively.

The performance of a given action-selection strategy is typically measured by the regret R_T , which compares the performance achieved by the selected actions with the best fixed action in hindsight. In general, an algorithm is said to learn if its regret grows sublinearly. Typical good performance is achieved for $R_T \in \mathcal{O}(\sqrt{T})$ or even $R_T \in \mathcal{O}(\log T)$.

Despite its simplicity, the bandits framework stands as a compelling solution for decision-making problems. The main reason is that bandit feedback (i.e., the agent only gets information

of actions as it plays them) is easier to be provided in practice than full or partial feedback (i.e., information of other actions are provided even if not being played). The most typical bandits applications are web advertising, sales optimization, online recommendations, resource allocation, and packet routing, among many others. A plethora of bandits formulations exist according to multiple assumptions that extend the basic bandits game. The different MAB formulations are based on multiple aspects: the statistics behind rewards (e.g., stochastic vs. non-stochastic bandits), the availability of actions (e.g., sleeping bandits, mortal bandits), the type of Markovian settings (e.g., rested vs. restless bandits), the nature of the environment (e.g., adversarial bandits), and a very long etcetera of variations. The bandits problem has been treated in detail by several books and surveys. Table 3.1 provides a high-level categorization of the most popular types of bandits. We encourage the interested reader to delve into the works in [121–125].

Table 3.1: High-level categorizations of most popular bandits types.

Categorization Criteria	Bandits models
Reward generation process	Stochastic, adversarial, Markovian (rested/restless)
Reward function	Discrete, continuum (linear/nonlinear), Lipschitz, Gaussian
Feedback type	Full information, bandit, semi-bandit, partial monitoring
State-awareness	Contextual bandit

In wireless communications, many problems have statistical characteristics that can be approximated with mathematical models. In this regard, MAB-based applications have shown great potential for optimizing a plethora of problems. Some examples are channel selection [126], spectrum access [127], transmission scheduling [128], or AP selection [129]. Table 3.2 provides an overview of some popular MAB-based applications in communications.

Concerning decentralized SR in wireless networks, it can be naturally defined as a multi-agent (or multi-player) problem. Each individual (e.g., a BSS) has player-specific goals and rewards. The maximum achievable performance of a node depends on its transmission capabilities, the interference it senses, the traffic load it needs to serve and/or receive, etc. The multi-agent approach allows capturing the distributed nature of IEEE 802.11 WLANs and keeping

Table 3.2: Overview of bandit-based applications for communications

Problem	Modeling	Goals	Baseline algorithms	References
Opportunistic spectrum access & Channel selection	Stochastic, non-stochastic, restless, contextual, Markovian bandits	Decentralized optimal allocation, optimize number of secondary transmissions, ϵ -correct ranking	UCB, ϵ -greedy, calibrated forecasting	[126, 127, 130–133]
Power control	Non-stochastic bandits	Optimize SINR	Follow the perturbed leader, exponential weighted average	[134]
User association	Sleeping, Bernoulli, non-stochastic bandits	Energy saving, improve the throughput	UCB, ϵ -greedy	[135, 136]
Inter-cell coordination	Adversarial, stochastic, non-stochastic, contextual bandits	Optimize inter-cell frequency resources, energy saving, map SON configurations and operator objectives	EXP3, UCB, ϵ -greedy	[137–140]
Dynamic rate selection	Structured, Markovian bandits	Maximize the number of packets successfully transmitted, learn changes in the channel	UCB	[141, 142]
LTE/Wi-Fi coexistence	Convex bandits	Fair channel sharing	Online gradient descent	[143]

Table 3.3: State-of-the-art multi-player MAB solutions for channel access in cognitive radio.

Work	Approach	Requirements	Results
[131]	Distributed mechanism that combines sensing with randomized access to learn channel statistics and the activity of other users	<ul style="list-style-type: none"> - The number of users is fixed and known - Channel sensing is perfect - All the players use the same strategy - Binary reward (free/collision) 	Order-optimal regret with logarithmic lower bound
[154]	Decentralized time-division fair sharing of the best arms	<ul style="list-style-type: none"> - i.i.d. reward - Conditions of linearity, continuity, and density for unknown parameters - Binary reward (free/collision) - The number of users is fixed and known 	Same logarithmic regret order as for collaborative approach where nodes exchange observations and make decisions jointly
[146]	Collaborative mechanism based on slotted periods (decision, sensing, transmission, communication)	<ul style="list-style-type: none"> - Channel sensing is done - CSMA/CA is used by secondary users - Rewards are broadcasted - Same channel conditions for all users 	Linear regret improving random and greedy channel access schemes
[133]	Distributed no-regret learning with calibrated forecaster	<ul style="list-style-type: none"> - The joint action profile is known - All the players use the same strategy - Time-invariant average channel gains 	Global optimal solution and convergence to correlated equilibria
[150]	Non-cooperative selfish approach based on ϵ -greedy exploration and CSMA/CA	<ul style="list-style-type: none"> - $K \geq N$ (K: channels, N: users) - The number of users is fixed and known 	Sub-linear regret and convergence to system-optimal solution
[126]	Centralized method for combinatorial bandits with user-channel pairs	<ul style="list-style-type: none"> - $K \geq N$ (K: channels, N: users) - Throughput as an i.i.d. random variable - Coordination/synchronization 	Upper bound regret that grows polynomially with the combinatorial number of users and channels

dimensionality low for the SR problem. However, it may lead to a competition among players, thus revealing a nexus with game theory. A player attempts to maximize a long-term reward in a single-agent system by interacting with an environment (which can be stochastic or non-stochastic) in isolation. Under this setting, performance guarantees can be straightforwardly provided, even if dealing with adversarial [144] or dynamic environments [145]. In contrast, weaker performance guarantees can be provided for multi-agent systems. Notice that, in the multi-player setting, the reward of an agent depends on what all (or some) of the other agents do, which leads to non-stationarity. As a result, the knowledge acquired by agents becomes quickly outdated.

Most of the current literature in multi-player MABs for wireless communications is based on the channel access problem in cognitive radio [90, 126, 130, 131, 146–150]. The cognitive radio characteristics make it a suitable and attractive problem to be modeled with the bandits framework. In particular, each node attempts to access the channel represents a player, and channels are arms (or bandits). In general, rewards are granted to players in a binary fashion, “1” if the channel can successfully be accessed, or “0” otherwise (two or more nodes select the same channel). Accordingly, each player has the same view on actions (different players playing the same action obtain the same payoff), which makes the game smooth, i.e., the player’s reward function is continuous concerning the entire strategy set. Table 3.3 analyzes the state-of-the-art approaches taken for modeling channel access in cognitive radio through multi-player MABs, which make use of popular baseline algorithms such as ϵ -greedy [120], upper confidence bound (UCB) [151, 152], exponential weight algorithm for exploration and exploitation (EXP3) [120, 153], and Thompson sampling [118].

3.3 Multi-Armed Bandits for Decentralized Spatial Reuse: Between ML and Game Theory

In **Paper #1**, **Paper #2**, and **Paper #3**, we have addressed decentralized SR through multi-player MABs, which allowed us to reduce the combinatorial complexity of the problem.¹ In particular, each agent (or player) $p \in \mathcal{P}$ represents a BSS in an OBSS. The actions $a \in \mathcal{A}_p$ that each agent can choose are defined as combinations of sensitivity and transmit power values, which lead to agent-specific rewards $r \in \mathcal{R}_p$. Notice that individual rewards depend on the joint action profile, i.e., the global configuration in BSSs, due to the spatial interactions between nodes. Algorithm 1 describes the decentralized SR game in general terms. The procedure mentioned above is enabled by a monitoring phase (time between iterations), whereby agents collect information of the performance granted by the actions being selected, which depend on the joint action profile (see Figure 3.3).

Algorithm 1: Decentralized SR game through MABs

```

1 Function MAB ( $\mathcal{A}$ );
  Input :  $\mathcal{A}_p$ : set of possible actions in  $\{1, \dots, K\}$  to be selected by each player  $p \in \mathcal{P}$ 
2 initialize:  $t = 0$ , for each arm  $k \in \mathcal{A}_p$ , set  $r_{p,k} = 0$  and  $n_{p,k} = 0$ 
3 while active do
4   for each  $p \in \mathcal{P}$  do
5     Play arm  $k \in \mathcal{A}_p = \operatorname{argmax}_{i=1,\dots,K} \theta_i$ 
6     Observe the reward obtained  $r_{p,k}(t)$ 
7     Compute the reward  $r_{k,t}$ 
8      $n_{k,t} \leftarrow n_{k,t} + 1$ 
9   end
10  Update the estimation function  $\theta$ 
11   $t \leftarrow t + 1$ 
12 end

```

In **Paper #1** and **Paper #2**, we have proposed a selfish method whereby agents learn towards maximizing local reward. Specifically, several concurrent agents attempt to improve their performance, based on local information. The effectiveness of this method is tightly coupled to the global reward. In potential games [155], a class of problems for which individual agents are connected to a global function with a unique maximum, optimizing local rewards also brings the whole system closer to the unique global optimum. For instance, the work in [149] addresses transmission power control and channel access in a distributed manner. According to a simplified model of opportunistic spectrum access, the authors can provide distributed no-regret strategies that lead to the set of correlated equilibria. However, some assumptions are made to define a unique maximum in the global performance function so that convergence guarantees to the optimum can be provided. In particular, the reward function used by the players is continuous with respect to the strategy set, which is also bounded. Intuitively, this means that the social cost that any action incurs to the other players can be linearly quantified.

¹Notice that the number of configurations in an OBSS grows exponentially with the number of BSS to be optimized.

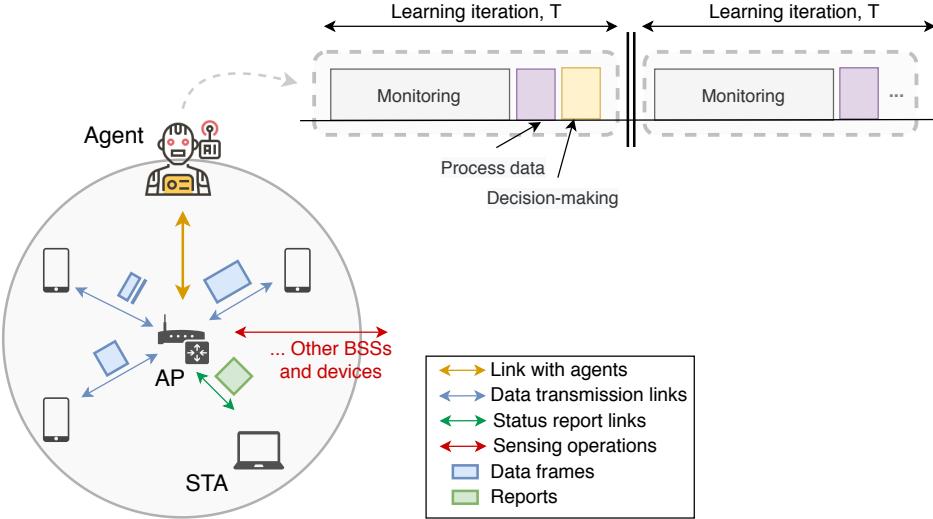


Figure 3.3: Sequential learning procedure carried out in IEEE 802.11 WLANs.

Therefore, in multi-player MABs, optimal solutions can typically be provided only to tractable problems that are generally linear, stationary, and generated by independent stochastic processes (e.g., with underlying Gaussian statistics). This is not the case of the proposed decentralized SR solution, in which it is not possible to provide a distributed no-regret strategy that converges to an optimal equilibrium. The fact is that the joint reward function of the SR problem with multiple concurrent players does not have a unique optimum. The set of correlated equilibria cannot be then characterized. In particular, the following properties differing from distributed channel access prevent to do so:

1. Spatial interactions inflict abrupt changes to the reward obtained by a BSS, which can be based, for instance, on the throughput. To put an example, increasing the sensitivity contributes to reducing contention, but it may lead to noticing a higher amount of interference during transmissions.
2. Besides, considering the worst-case interference (devices transmitting continuously) is an unrealistic assumption. Therefore, the social-cost of actions varies with time and according to the transmissions done on a per-packet-basis (where certain randomness is added due to many causes such as channel effects, retransmissions, the randomized backoff procedure, etc.).

In these situations, defining a shared learning goal in multi-agent systems is not trivial because rewards are not equally assigned to agents. Each reward depends on the joint action profile. Nonetheless, the MABs framework remains a powerful solution to address real-world problems with complex and unpredictable phenomena behind reward distributions. The multi-player setting (i.e., multiple agents attempt to learn concurrently) allows for tackling smaller individual problems rather than attempting to solve the full centralized problem. While this leads to non-stationary, on-time improvements prevail over long-term optimality.

As an alternative to the selfish approach, a collaborative setting was also proposed in **Paper #3**, whereby agents attempt to maximize a shared reward. In particular, we defined the max-min throughput to be improved by a set of overlapping BSSs. Note, as well, that the learning

procedure is kept decentralized (each agent is responsible for selecting its actions). The collaborative approach is meant to boost fairness, but as for the selfish setting, its effectiveness towards optimality is limited by the shape of the global performance function.

Chapter 4

METHODOLOGY AND ENABLERS

The use of analytical models and simulation tools is crucial for studying and understanding novel technologies such as SR. In this Chapter, we present the analytical and simulation tools that have been used for that purpose. In particular, we developed an analytical model to address the complex inter-BSS interactions posed by the SR operation. Besides, given the lack of simulation tools to characterize such a new technology, we developed an IEEE 802.11ax-based network simulator. The purpose of this simulator is twofold. First, we need it to evaluate the performance of SR in dense deployments at a macro-scale level, so that we can provide a first taste of the potential advantages and implications of SR. Notice that current well-known simulation tools like ns-3 can be complex to extend due to the high level of detail put in both MAC and PHY layers. Besides, simulating highly dense deployments can be costly or even intractable in terms of time and computational resources. Second, due to the primary purpose of this thesis on studying the application of ML models for SR, the simulator was developed for supporting the operation of intelligent agents.

4.1 Spatial Reuse through Continuous Time Markov Networks

The Bianchi model [156] has been widely adopted by the research community as a reference for analyzing the throughput of IEEE 802.11 WLANs. However, this model only focuses on the MAC layer and requires all the analyzed nodes to be in the same coverage area. To model SR, it is important to capture the dynamic PHY interactions when tuning the sensitivity threshold and the transmit power.

The analysis of SR has been previously addressed in multiple ways. Most of the models are based on signal-to-interference-plus-noise ratio (SINR) [157] and include the concept of physical carrier sensing to capture inter-device interactions. We find a plethora of works that build upon SINR-based models for meeting several purposes (model RTS/CTS, etc.) [22, 158–160]. Typically, SINR-based models use the concepts described in Table 4.1. First, the device’s sensitivity allows defining the transmission range of a given transmitter-receiver pair based on fixed transmit power. Second, the carrier sensing threshold determines the idle/busy channel condition from a transmitter node’s point of view. The carrier sensing threshold has been widely used for defining the carrier sense area (alternatively, the set of carrier sense nodes) or the silence set (i.e., the set of nodes that detect the channel busy if a given node transmits).

Table 4.1: Concepts to model spatial reuse in wireless networks.

Concept	Description	Function
Sensitivity	A receiver can detect a transmission if its received power is greater than the sensitivity threshold	- Determine the transmission range
Carrier sensing	A node cannot initiate a transmission if the power it senses is above its carrier sense threshold	- Determine the carrier sense set - Determine the carrier sense range - Determine the silence set
Capture effect	A receiver can decode a signal if the SINR is above a given threshold	- Define the interference range - Determine the interference set

Finally, the capture effect includes interfering nodes' operation, thus allowing to define the interference range (alternatively, the set of interfering devices). While SINR-based methods are useful to derive interactions among devices in the spatial dimension, they consider the worst-case interference (i.e., nodes are assumed to transmit permanently). This assumption entails neglecting spectrum access coordination and hence losing insights on the MAC operation.

Another field that is attracting a lot of attention is stochastic geometry (SG), which allows modeling the random nature of dense wireless networks. In particular, SG allows deriving statistical properties of the underlying interactions among a random set of nodes (typically, set according to random point processes). In telecommunications, SG has been widely applied to model the behavior of users and to estimate metrics such as the outage probability or the throughput per area [161]. Concerning SR, we highlight the works in [162–165], which provided models based on SG to capture the effect of tuning the sensitivity threshold in WLANs. However, SG models mainly focus on PHY layer effects and fail to capture the asymmetries that may occur when applying the SR operation, which also involves the tuning of the transmit power.

In **Paper #5**, we introduced Continuous Time Markov Networks (CTMNs) [166, 167] to analytically model the SR operation, thus capturing both MAC and PHY layers interactions. This model allowed us to provide insights into the inter-BSS interactions resulting from SR, which has been key to develop the mechanisms proposed in this thesis.

The CTMN model captures the CSMA/CA operation used in IEEE 802.11 WLANs through states representing the set of BSSs that are transmitting at a given moment. Transitions between states occur when BSSs start a transmission (i.e., they gain access to the medium) or when the transmission finishes (i.e., they abandon the channel). It is important to highlight that the CTMN model considers additive interference, which results from the combination of different simultaneous interfering transmissions. Accordingly, we are able to characterize real deployments where spatially-distributed interactions occur. Concerning this model, the following assumptions are considered:

1. The backoff procedure for accessing the medium is continuous in time, which is useful for modeling the attempt rate of transmitters as a Poisson process. However, this assumption prevents modeling collisions due to backoff expiring at the same instant.
2. Data transmissions are downlink only, which relaxes the complexity of the problem and allows us to focus only on inter-AP interactions. Notice that a state in the Markov chain represents a set of nodes transmitting in an OBSS. As a result, the complexity of the model grows in a combinatorial manner.

3. To be consistent with the previous assumption, uplink transmissions of control packets (e.g., ACKs) are only considered for computing the total transmission time. This leads not to consider uplink transmissions for modeling inter-BSS interactions.
4. Full-buffer traffic is considered for the sake of analyzing the interactions resulting from SR more lucidly.

The CTMN model is very useful to understand the new kind of inter-AP interactions resulting from the 11ax SR operation. Besides, it allows us to validate the implementation of SR done at the simulator, which is later introduced in Section 4.2. A disadvantage of the CTMN model is its lack of scalability for characterizing crowded deployments. Modeling dense scenarios becomes intractable since the number of feasible states increases in a combinatorial manner with the number of BSSs.

4.1.1 IEEE 802.11ax OBSS/PD-based Spatial Reuse

To model the 11ax SR operation, we consider new states for capturing the different situations created by the sensitivity levels and the corresponding transmit power limitation to be employed by each BSS in a given scenario. First, tuning the OBSS/PD of a BSS allows finding new types of inter-BSS interactions that could not exist without applying SR (i.e., a new set of states modeling concurrent transmissions). Moreover, the transmit power limitation in new states has implications on the capabilities of BSSs. Notice that a lower transmission power entails a more robust Modulation and Coding Scheme (MCS) and, as a consequence, a lower data rate.

Figure 4.1 illustrates the application of the proposed CTMN model for a simple deployment. In particular, we consider two BSSs that can transmit concurrently, provided that one of them (namely, BSS_A) applies OBSS/PD-based SR. Figure 4.1(a) shows the considered deployment, where the carrier sense area of AP_A is illustrated for two different carrier sensing approaches. First, the default CCA/CS threshold (shown in red) makes AP_A sense the channel busy when AP_B is transmitting, thus preventing parallel transmissions to occur. In contrast, the OBSS/PD threshold (shown in green) allows AP_A ignoring AP_B 's transmissions and therefore access the channel. Note, as well, that the lower OBSS/PD threshold entails reducing the transmit power, which leads to a lower MCS index. These interactions are captured by the corresponding CTMN (see Figure 4.1(b)). As shown, BSS_A can transmit in two different manners (noted by A and A_{SR} , respectively), which extend legacy interactions with SR channel access rules and corresponding nodes' capabilities. Notice that the transmit power limitation when applying the OBSS/PD threshold makes BSS_A to use a more robust MCS, thus sacrificing capacity. In exchange, it is possible to hold simultaneous transmissions when BSS_B occupies the channel.

4.1.2 IEEE 802.11be Coordinated Spatial Reuse

To the date of publishing this thesis, the C-SR operation is currently being defined, studied, and characterized. However, we started devising its implications through the CTMN model. In particular, we have extended the 11ax-based model by taking the following assumptions:

1. Coordinated transmissions always have priority when defining transitions between states. This means that joint transmission will occur over individual ones.

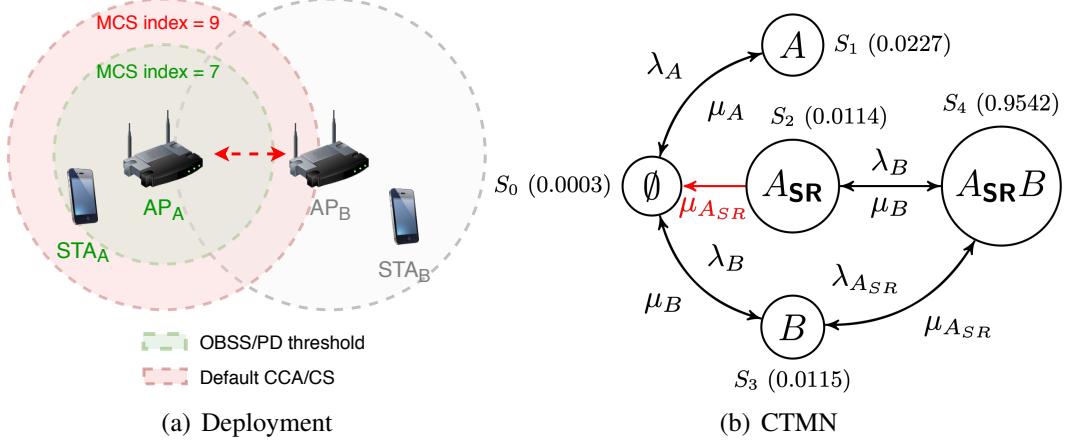


Figure 4.1: Example of the CTMN model for IEEE 802.11ax OBSS/PD-based SR.

2. The feasibility of a coordinated transmission is only assessed from the perspective of the sharing AP. Therefore, it is not assessed whether the transmission of shared APs will fail or succeed in transitioning to coordinated states.
3. Sharing and shared transmission duration are assumed to be equal.

The C-SR operation through CTMNs is exemplified in Figure 4.2. We address a deployment in which flow-in-the-middle starvation is prone to occur in a 3-BSS deployment if applying the default channel sensing mechanisms. As shown in Figure 4.2(a), AP_B would mark the channel as busy when detecting transmissions from either AP_A or AP_C. However, AP_A and AP_C are not in range of each other and can, therefore, transmit simultaneously. The C-SR operation can solve the flow-in-the-middle situation. BSS_A and BSS_B can be coordinated to transmit simultaneously, thus leading to the new kind of inter-BSS interactions depicted in 4.2(b). As shown, the introduced coordinated transmissions entail new inter-BSS interactions, which is, for instance, the simultaneous transmission of all the BSSs in the state AB_{SR}C. Similarly to 11ax OBSS/PD-based SR, the coordinated transmission requires adjusting the transmit power in order not to affect the primary transmissions (i.e., the ones started by the sharing nodes). In this case, the transmit power of both sharing and shared APs is negotiated based on the DL ARIL of the intended receivers. The fact of keeping track of the interference support by each STA allows improving further the quality of the transmissions resulting from the SR operation.

4.2 Simulation of Sequential Learning for Spatial Reuse

4.2.1 Spatial Reuse Implementation

To devise the potential gains that SR can provide to next-generation dense wireless networks, we introduce the 11ax SR operation in the Komondor simulator, which is presented in **Paper #4**. Komondor has been conceived to *i*) allow the low-cost integration of novel mechanisms included in new IEEE 802.11 standards, *ii*) simulate high-dense deployments, and *iii*) include the operation of sequential learning algorithms within the simulation of the network.

To characterize the operation of WLANs in a realistic manner, Komondor reproduces actual transmissions on a per-packet basis. To that purpose, Komondor is based on the COST library,

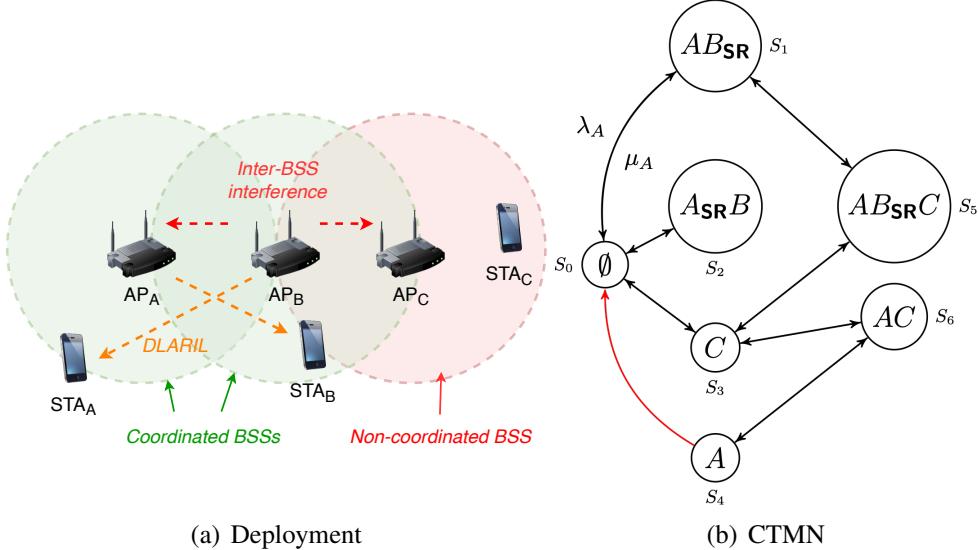


Figure 4.2: Example of the CTMN model for IEEE 802.11be C-SR.

which allows building interactions between components (e.g., wireless nodes, buffers, packets) through synchronous and asynchronous events. The implementation of DCF in Komondor has been validated against ns-3, and the CTMN [166] and Bianchi models [156].

For 11ax OBSS/PD-based SR, the behavior of Komondor is as follows (also illustrated in Figure 4.3):

- Devices implementing SR must announce that they support the operation so that devices in a BSS are setup.
- On initiating an SR-enabled transmission, a device must indicate its BSS Color and SRG.
- Overlapping devices implementing SR, can take advantage of SR-enabled transmissions and transmit concurrently. The following conditions must hold for any detected transmission (i.e., the received power is above the minimum CCA/CS threshold):
 - Detected transmissions must indicate support for SR.
 - Detected transmissions belong to different BSS Color sets or SRGs than the device detecting SR-based transmission opportunities.
 - The power detected from the transmission must not exceed the assigned OBSS/PD threshold.
 - The transmit power to be used in the concurrent transmission must be limited according to the SR rules (see Chapter 2).
- In case of not meeting the abovementioned requirements, it is not possible to transmit during the SR TXOP, so the device activates a NAV timer.

The IEEE 802.11ax SR operation has also been developed for ns-3, but it has been published very recently¹. In particular, we provided the implementation in Komondor as a complementary way to characterize and understand SR. Notice that ns-3 includes an exhaustive model of

¹<https://gitlab.com/nsnam/ns-3-dev/-/tags/ns-3.30>

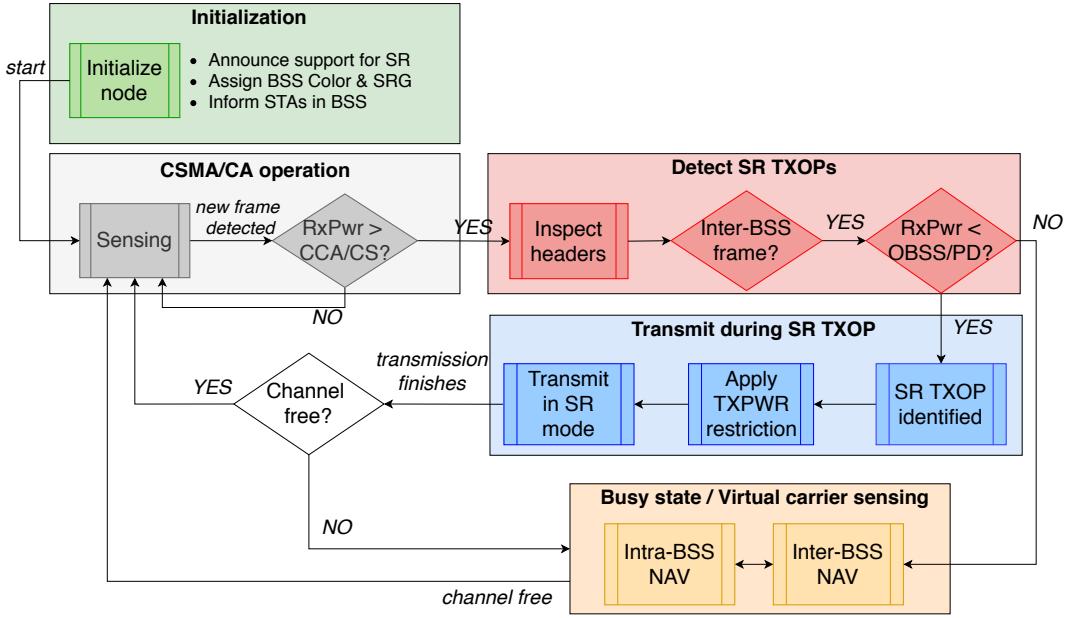


Figure 4.3: Implementation of IEEE 802.11ax OBSS/PD-based SR in Komondor.

the physical (PHY) layer, whereas the primary goal of this thesis lies in the MAC interactions occurring when using SR. Related to this, the lower complexity of Komondor allows us to simulate high-density deployments with reasonable simulation times. Apart from that, Komondor continues providing extended features of 11ax SR. The ns-3 implementation includes baseline OBSS/PD-based SR with constant thresholds. In contrast, Komondor allows us to easily introduce algorithms to adjust the OBSS/PD threshold dynamically. Besides, the implementation of SRGs is provided to explore its potential utilization in future IEEE 802.11 amendments.

4.2.2 Agents-based Implementation

Regarding the integration of ML mechanisms into network simulation, we find the integration of AI Gym with ns-3 [168], which provides a discrete interaction between the simulated network components and AI libraries. When it comes to Komondor, a fully integrated implementation of agents was provided. These agents interact with simulated nodes in a network for providing monitoring, processing, and decision-making functionalities. In particular, different communication-based approaches are considered to enable the application of decentralized, distributed, centralized, and hybrid ML mechanisms (see Figure 4.4).

On designing sequential learning mechanisms in Komondor, the following considerations should be taken into account:

- Agents acquire information from BSSs on a periodical-basis. The information retrieval interval depends on the established monitoring phase, which entails a trade-off between the delay in making decisions and the quality of the sample extracted from monitoring.
- The information acquired by agents can be later shared among other BSSs or with a centralized entity, thus allowing to build cooperative, distributed, or centralized mechanisms. Agents with different purposes can coexist.

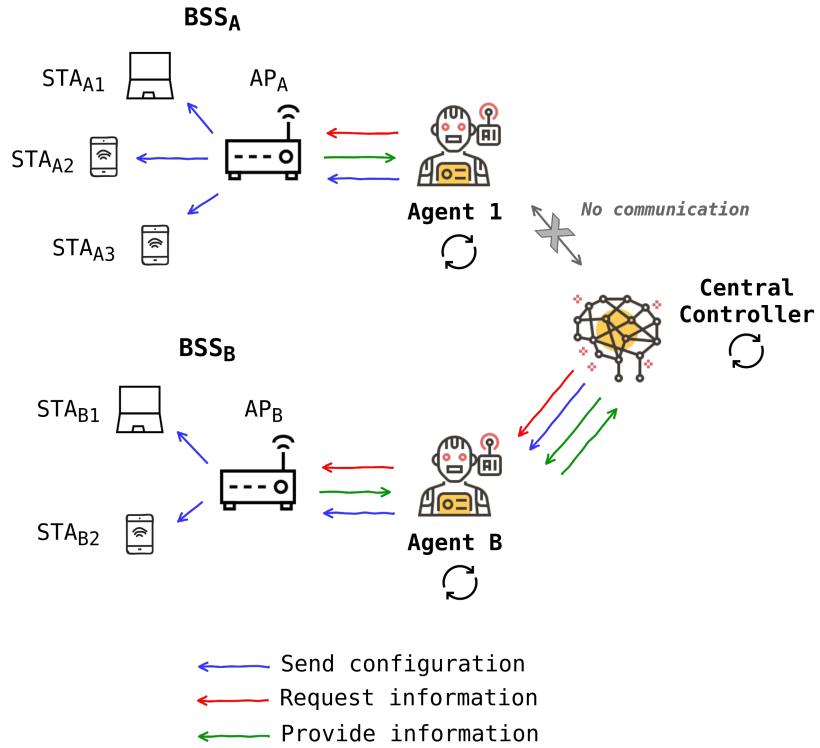


Figure 4.4: Agents implementation in Komondor.

- The communication among agents and APs can be assumed to be virtual or entail certain costs (e.g., delay, overhead). Besides, packet losses can be included when agents exchange information, which can impact on the learning procedure followed by agents.

Chapter 5

MAIN FINDINGS

In this Chapter, we present the main findings of this dissertation, which result from the analysis of the 11ax SR operation and the application of ML to address SR. **Paper #5** and **Paper #6** focus on the model, characterization, and analysis of IEEE 802.11 SR, thus aiming to show the potential achievable gains of this technology. Then, **Paper #1**, **Paper #2** and **Paper #3** study the applicability of sequential learning for SR through different proposed mechanisms. Specifically, **Paper #1** proposes a multi-agent setting in which players implement a stateless version of Q-learning [169, 170] to optimize a selfish reward based on the local throughput. Similarly, **Paper #2** and **Paper #3** propose the utilization of MABs for addressing the multi-agent SR problem. While **Paper #2** analyzes the impact of different algorithms (ϵ -greedy, EXP3, UCB, and Thompson sampling) in selfish settings, **Paper #3** extends the analysis to collaborative rewards. Finally, **Paper #6**, **Paper #7**, and **Paper #8** delve into the necessary infrastructure and tools to carry out the main object of study of this thesis.

Finding #1: SR is a fair mechanism that allows increasing the number of simultaneous transmissions in dense OBSSs, thus enhancing the throughput in high-interference scenarios.

The analysis of 11ax SR has been conducted in **Paper #5** and **Paper #6** for residential and enterprise-like Wi-Fi deployments (see Figure 5.1). Our results compare the legacy setting (CCA/CS is used for all the transmissions) to SR with the best OBSS/PD threshold. In the first place, Figure 5.2 shows the throughput experienced by the BSSs in an overlapping deployment for different network densities and traffic load values. Solid bars refer to the performance achieved by the BSS implementing SR, while dashed bars are for the rest of overlapping BSSs. The first important conclusion is that SR is a fair mechanism that protects the detected legacy transmissions to transmit in SR mode. Notice that the performance of BSSs that are not applying SR (dashed bars) is barely affected when SR is applied by other BSSs (solid bars). Besides, BSSs applying SR can overcome high-interference settings that lead to suffering starvation and other performance anomalies (e.g., collisions). Accordingly, the throughput of dense deployments is improved.

Finding #2: SR allows to significantly improve the delay in dense deployments.

This finding is illustrated in Figure 5.3, which shows the CDF of the delay for different network densities and traffic load values. As for the throughput evaluation, solid lines refer to

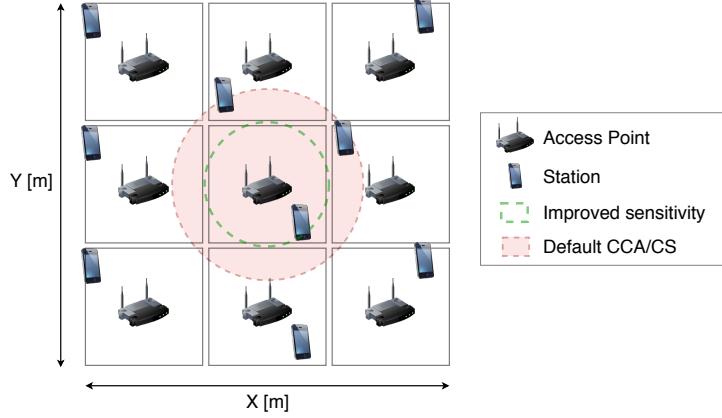


Figure 5.1: Residential and enterprise-like deployments to evaluate the performance gains of IEEE 802.11ax Spatial Reuse.

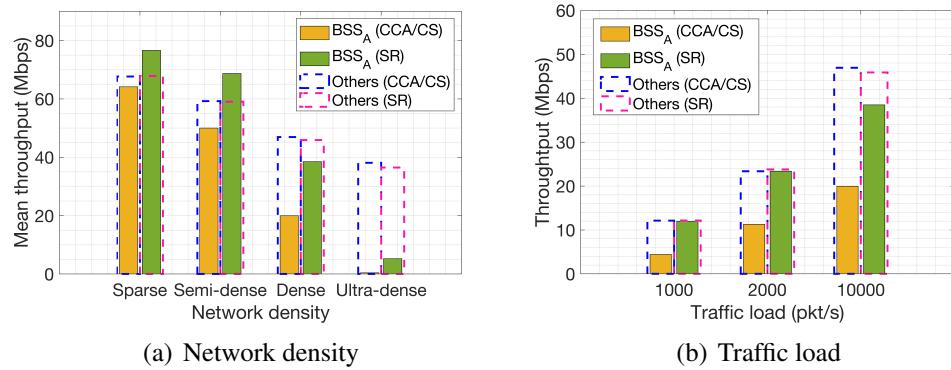


Figure 5.2: Throughput gains obtained by IEEE 802.11ax Spatial Reuse in comparison to default carrier sensing approaches. Different network densities and traffic load values have been considered.

the performance achieved by the BSS implementing SR, while dashed lines are for the rest of overlapping BSSs. As illustrated, the probability of experiencing a high delay increases rapidly with the traffic load and the deployment's density. Nevertheless, the SR operation provides a substantial improvement in the delay, thus outperforming the legacy carrier sensing approach. The leading cause of this improvement lies in the number of simultaneous transmissions that are allowed through SR. **Paper #5** and **Paper #6** fully describe this finding.

Finding #3: The lack of coordination of the IEEE 802.11ax SR mechanism prevents providing further performance gains in dense deployments.

Although SR has been shown to enhance current IEEE 802.11ax deployments, its performance is bounded by the following characteristics:

- The current transmit power limitation is too conservative, and further gains are expected if it is properly adjusted by BSSs participating in SR-based transmissions.
- The lack of coordination among BSSs implementing SR prevents to find the optimal

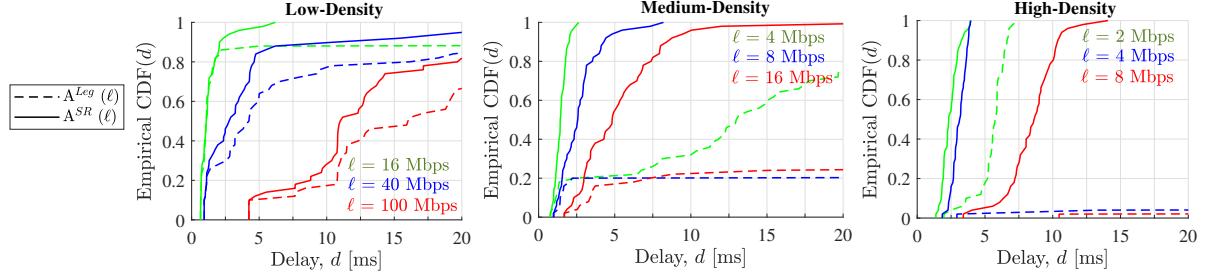


Figure 5.3: CDF of the delay gains obtained by IEEE 802.11ax Spatial Reuse in comparison to default carrier sensing approaches. Different network densities and traffic load values have been considered.

scheduling allocations. Through coordinated transmissions, BSSs can determine the optimal set of transmitter-receiver nodes in each SR-based transmission.

- The rigidity of the current approach, which is applied homogeneously in a BSS instead of considering per-STA behavior.
- SR is not combined with other technologies such as OFDMA or beamforming, which is expected to provide further performance gains.

Paper #5 and **Paper #6** give some insights regarding this finding. Apart from that, in Figure 5.4, we devise the potential gains provided by CSR in comparison to the default carrier sensing approach and the 11ax OBSS/PD-based SR operation. Notice that the obtained results are preliminary since the CSR is currently under study and is part of the future work. In particular, we study the effect of coordination in a simple deployment with two BSSs, namely BSS_A and BSS_B. As illustrated, the fact of coordinating the transmissions held between the devices in BSS_A and BSS_B allows improving the transmit power used in each case, thus granting a higher throughput.

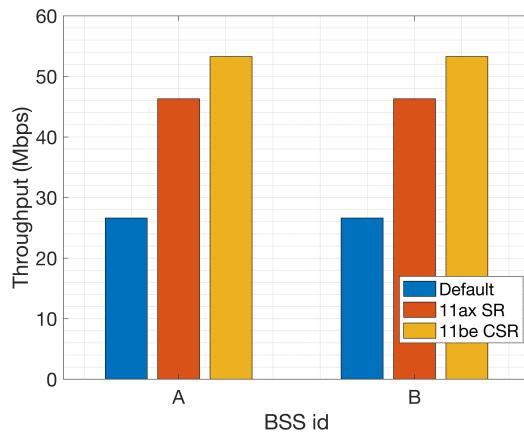


Figure 5.4: Further SR gains provided by CSR in comparison to default carrier sensing and OBSS/PD-based SR.

Finding #4: Sequential learning mechanisms for multi-agent SR allow improving the performance of WLANs but may fail at finding a global optimum due to the competition among BSSs.

To address the SR problem in decentralized WLANs, we proposed the application of different multi-player sequential learning approaches, which vary according to the available information that agents have about the surrounding environment (i.e., other players).

In **Papers #1** and **Paper #2**, we studied the effects of using selfish rewards (local information only) in competitive settings where multiple agents coexist. We showed that the concurrent learning operation may lead to a competitive setting, which has an impact on the learning procedure carried out by each agent. In particular, the joint learning operation was shown to prevent finding the optimal global configuration and potentially leading to a high variability on the rewards obtained by each agent (intermittent good/poor performance depending on the neighbor decisions). These effects were shown to be motivated by the throughput demands of individual agents, making the global function highly non-convex. In turn, the selfish setting was also shown to boost fairness in some situations. This is the case where the global optimum solution, which depends on each agent (e.g., location, set of available configurations), favors collaboration among BSSs, even if individual rewards are selfish.

The effects of learning SR selfishly in a multi-player setting are summarized in Figure 5.5, which shows the throughput evolution of four BSSs that apply sequential learning for tuning both the sensitivity and the transmit power. As shown, the concurrent learning operation prevents reaching the optimal individual performance (depicted by the red dashed line). Regarding stability aspects, we noticed that the provided algorithms converged slowly to a stable configuration. Nevertheless, the learning procedure is characterized by a high throughput variability. The main cause lies in the competitive setting, whereby agents intermittently select good/bad performing actions. In practice, agents end up picking a protective configuration that prevents to reach the global optimum. The variability issues that arise from this kind of multi-agent deployments can be mitigated from an implementation point of view (e.g., define stop goals, enforce mechanisms for convergence, reduce the learning rate, etc.).

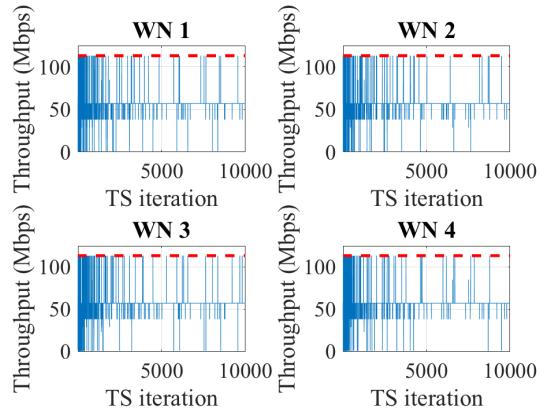


Figure 5.5: Implications of learning SR selfishly in a multi-player setting.

Finding #5: Collaborative rewards enhance fairness in decentralized settings, but do not ensure reaching an optimal global solution.

In **Paper #2**, we also studied the performance of collaborative settings, whereby the reward of an individual agent takes into consideration the performance of the other players. In this regard, we observed that collaborative rewards are useful to prevent unfairness, but face the same challenges as selfish approaches for converging to an optimal global solution. In the first place, each agent is responsible for its own action set but has no control on the joint action profile. Therefore, the action-selection procedure is kept decentralized. Otherwise, due to the combinatorial nature of the problem, the complexity would grow very fast. Apart from that, the global optimization function has multiple maximums, even if it is based on a collaborative metric.

To illustrate the effect of using cooperative rewards in a multi-player setting, Figure 5.6 shows the average throughput and the max-min throughput obtained in random enterprise-like deployments, for different network densities. The default (static) configuration is compared to learning approaches based on non-cooperative (selfish) and cooperative (environment-aware) rewards. As illustrated, the cooperative reward allows improving fairness (shown in Figure 5.6(b)), but leads to a lower overall performance if compared to the selfish mechanism (shown in Figure 5.6(a)).

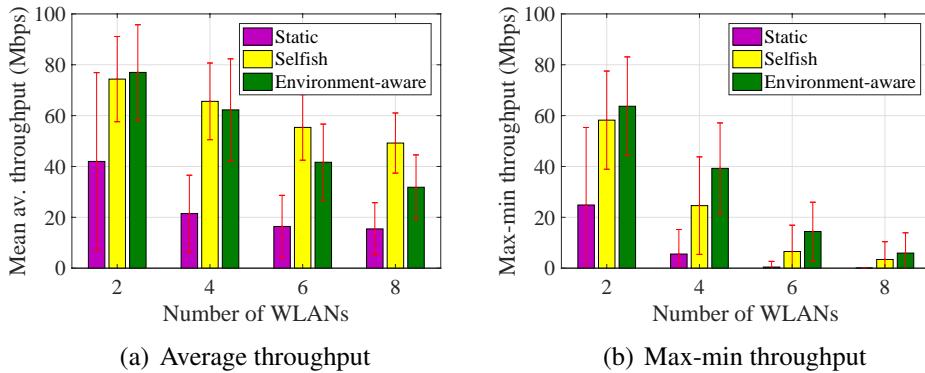


Figure 5.6: Effect of cooperative vs non-cooperative rewards in a multi-player setting.

Finding #6: Bayesian-based exploration methods have the potential to address the non-stochasticity behind decentralized SR.

Regarding the learning procedure in decentralized SR (both selfish and collaborative settings), in **Paper #2**, we showed that Bayesian-based exploration methods (e.g., Thompson sampling) perform well and grant better results than other types of algorithms that are indeed oriented to adversarial settings (e.g., EXP3). Notice that, as a result of the competition among multi-player SR agents, rewards distributions are highly non-stationary. Nevertheless, Bayesian exploration allows quick tracking of the best/worst performing actions but fails at identifying more subtle interactions when learning decentralized SR. This is illustrated in Figure 5.7, which shows the temporal evolution of throughput obtained in an OBSS, for different exploration strategies. We considered the popular algorithms ϵ -greedy, EXP3, UCB, and

Thompson sampling (TS). In the proposed setting, all the strategies allow finding the optimal global performance. Nevertheless, Thompson sampling does it more efficiently, thus reducing the variability on the throughput that results from exploration.

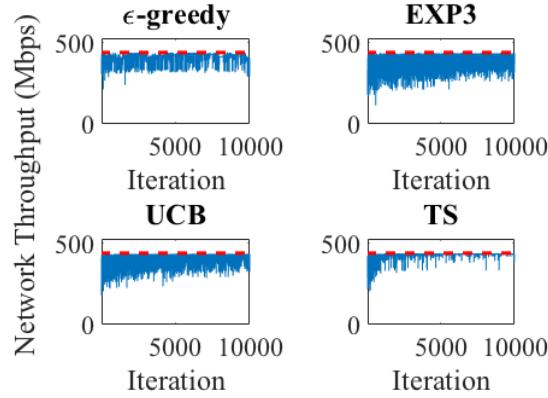


Figure 5.7: Temporal throughput achieved in an OBSS when applying different exploration methods in a multi-player setting.

Finding #7: Network dynamics and lack of information in local agents can severely affect the performance of sequential learning algorithms for decentralized SR.

In **Paper #3**, we delved into practical aspects on the application of sequential learning to the decentralized SR problem (results are summarized in Figure 5.8):

1. Networks are dynamic in different aspects (on/off devices, mobility, varying traffic requirements, channel fluctuations, etc.). This requires ML mechanisms to adapt to changes in the network, which entails the trade-off between old and new knowledge (how to assess data are becoming obsolete?). In this regard, we studied the effect of maintaining past information in sequential learning when significant changes in the network occur. In particular, we consider an OBSS whose topology changes radically at a given point (iteration 500). Figure 5.8(a) shows the temporal evolution of the overall performance as a result of applying sequential learning. The fact is that multi-player MAB algorithms are shown to adapt to sudden changes in the network topology, which modify the global optimum completely. However, finding a joint configuration that improves the new situation's performance requires several training iterations, which may degrade network throughput if changes occur too fast.
2. Sequential learning algorithms typically require rewards to be normalized. This normalization procedure is evident when the maximum achievable performance is known. However, this bound is unknown to the devices implementing SR and should, therefore, be approximated (e.g., by the maximum theoretical capacity, by the maximum data rate given the selected MCS, etc.). The fact of not knowing the upper bound reward of a given agent has implications on how learning is carried out (e.g., high variability, getting stuck in suboptimal actions, etc.). To study the effect of approximating the upper bound reward, we have considered a multi-agent setting composed of two BSSs. The temporal

throughput of each BSS is shown in Figure 5.8(b), which compares the learning procedure that results from a known (in blue) and an approximated (in red) upper bound reward. As illustrated, the approximated metric leads to a higher throughput variability since agents cannot precisely differentiate optimal actions from others leading to lower performance.

3. Sharing a reward in a collaborative setting can be challenging for the SR problem. The fact is that inter-agent interactions appear/disappear according to the sensitivity and the transmit power, which is a key impediment for agents to learn the hidden reward distributions of each action. Sharing the reward with the appropriate neighbors is therefore critical. To show the possible effects of not sharing the reward properly, we have considered a two BSS deployment. We compare long and short-range reward sharing approaches (i.e., the radius used to determine the set of agents that share a reward). Figure 5.8(c) shows the temporal throughput evolution of the two BSSs for each of the reward sharing approach. In this case, the short-range approach is the best since it allows considering only the agents that are implied in the spatial interactions (in red). In turn, the long-range sharing approach (in blue) leads to suboptimal performance because the throughput of independent agents is considered for building a collaborative reward, hence introducing noise to the learning procedure.

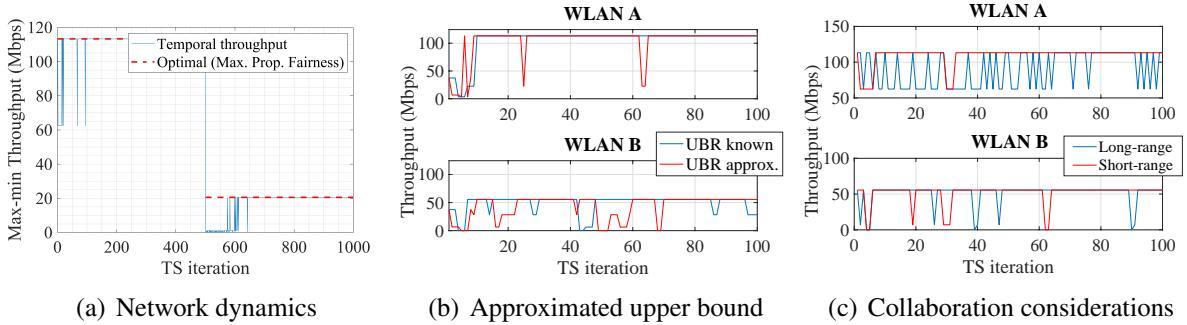


Figure 5.8: Practical considerations for the application of sequential learning to decentralized Spatial Reuse.

Chapter 6

CONCLUDING REMARKS

In this thesis, we investigated the viability of addressing SR in future IEEE 802.11 WLANs from a decentralized perspective. To that purpose, we first provided an in-depth study of the current 11ax SR operation and analyzed its ways forward through the 11be amendment. Then, we proposed the use of ML for addressing the uncertainty and non-stochasticity resulting from the concurrent adaptation of sensitivity and transmit power by devices in an OBSS. This approach entails a set of challenges related to multi-player agent settings, which are also studied in this thesis. In particular, the competition among nodes leads to a game-theoretic setting, where notions on equilibria gain significance. In this regard, we showed the main challenges that WLANs may face when applying sequential learning mechanisms for SR, including efficiency and convergence aspects. In order to conduct our research, we analytically modeled the SR operation and analyzed the new kind of interactions among BSSs. Besides, we implemented the 11ax SR operation in an ML-enabling network simulator, which allowed us to study the performance gains of SR and the proposed solutions based on sequential learning.

Our main findings confirm the potential of SR for improving dense wireless networks' capacity, especially on enhancing the average delay in an OBSS. Apart from that, we have shown that the decentralized SR mechanisms proposed in this thesis improve the default carrier sense approach and to enhance the performance in enterprise and residential-like scenarios effectively. However, the application of ML in a multi-player setting leads to a set of implications that may severely affect the fairness and the global performance of an OBSS. We based part of our research on analyzing the effects of selfish and collaborative methods for multi-agent SR.

We left as future work the potential of coordinated and centralized mechanisms to further enhance SR in next-generation WLANs. In this regard, an interesting topic lies in the trade-off between the potential achievable improvements and the corresponding overhead, which may include data acquisition, data exchange, or synchronization. Future research work also encompasses the convergence of SR with other technologies such as beamforming/null steering [85], OFDMA [171, 172], multiple antenna systems [173], or scheduled transmissions [174], and whether their joint operation can improve separate optimization. In this regard, AI may contribute to addressing the inherent complexity of end-to-end communications systems, thus revealing the potential of moving beyond block-based optimization. Finally, regarding the multi-agent setting, important research questions remain open concerning convergence aspects or training under dynamic topologies.

Bibliography

- [1] Boris Bellalta. IEEE 802.11 ax: High-efficiency wlans. *IEEE Wireless Communications*, 23(1):38–46, 2016.
- [2] Der-Jiunn Deng, Kwang-Cheng Chen, and Rung-Shiang Cheng. IEEE 802.11 ax: Next generation wireless local area networks. In *10th international conference on heterogeneous networking for quality, reliability, security and robustness*, pages 77–82. IEEE, 2014.
- [3] Evgeny Khorov, Anton Kiryanov, Andrey Lyakhov, and Giuseppe Bianchi. A tutorial on ieee 802.11 ax high efficiency wlans. *IEEE Communications Surveys & Tutorials*, 21(1):197–216, 2018.
- [4] S Merlin, G Barriac, H Sampath, L Cariou, T Derham, JP Le Rouzic, R Stacey, M Park, R Porat, N Jindal, et al. TGax simulation scenarios. doc.: IEEE, pages 802–11, 2015.
- [5] Emilio Calvanese Strinati, Sergio Barbarossa, Jose Luis Gonzalez-Jimenez, Dimitri Ktenas, Nicolas Cassiau, Luc Maret, and Cedric Dehos. 6g: The next frontier: From holographic messaging to artificial intelligence using subterahertz and visible light communication. *IEEE Vehicular Technology Magazine*, 14(3):42–50, 2019.
- [6] Rubayet Shafin, Lingjia Liu, Vikram Chandrasekhar, Hao Chen, Jeffrey Reed, and Jianzhong Charlie Zhang. Artificial intelligence-enabled cellular networks: A critical path to beyond-5g and 6g. *IEEE Wireless Communications*, 2020.
- [7] Marco Giordani, Michele Polese, Marco Mezzavilla, Sundeep Rangan, and Michele Zorzi. Toward 6g networks: Use cases and technologies. *IEEE Communications Magazine*, 58(3):55–61, 2020.
- [8] Tae-Suk Kim, Hyuk Lim, and Jennifer C Hou. Improving spatial reuse through tuning transmit power, carrier sense threshold, and data rate in multihop wireless networks. In *Proceedings of the 12th annual international conference on Mobile computing and networking*, pages 366–377. ACM, 2006.
- [9] Basel Alawieh, Yongning Zhang, Chadi Assi, and Hussein Mouftah. Improving spatial reuse in multihop wireless networks-a survey. *IEEE Communications Surveys & Tutorials*, 11(3), 2009.
- [10] Lin Zhang, Ming Xiao, Gang Wu, Muhammad Alam, Ying-Chang Liang, and Shaoqian Li. A survey of advanced techniques for spectrum sharing in 5g networks. *IEEE Wireless Communications*, 24(5):44–51, 2017.

- [11] Christina Thorpe and Liam Murphy. A survey of adaptive carrier sensing mechanisms for ieee 802.11 wireless networks. *IEEE Communications Surveys & Tutorials*, 16(3):1266–1293, 2014.
- [12] Robert Vilzmann and Christian Bettstetter. A survey on mac protocols for ad hoc networks with directional antennas. In *EUNICE 2005: Networks and Applications Towards a Ubiquitously Connected World*, pages 187–200. Springer, 2006.
- [13] Elias Yaacoub and Zaher Dawy. A survey on uplink resource allocation in ofdma wireless networks. *IEEE Communications Surveys & Tutorials*, 14(2):322–337, 2011.
- [14] Ruizhi Liao, Boris Bellalta, Miquel Oliver, and Zhisheng Niu. Mu-mimo mac protocols for wireless local area networks: A survey. *IEEE Communications Surveys & Tutorials*, 18(1):162–183, 2014.
- [15] Yongjun Xu, Xiaohui Zhao, and Ying-Chang Liang. Robust power control and beam-forming in cognitive radio networks: A survey. *IEEE Communications Surveys & Tutorials*, 17(4):1834–1857, 2015.
- [16] Jun Fang, Xingjian Li, Wen Cheng, Zhi Chen, and Hongbin Li. Intelligent power control for spectrum sharing: A deep reinforcement learning approach. *CoRR*, 2017.
- [17] Kodai Murakami, Tatsuya Ito, and Susumu Ishihara. Improving the spatial reuse of ieee 802.11 wlan by adaptive carrier sense threshold of access points based on node positions. In *2015 Eighth International Conference on Mobile Computing and Ubiquitous Networking (ICMU)*, pages 132–137. IEEE, 2015.
- [18] Fei Liang, Cong Shen, Wei Yu, and Feng Wu. Towards optimal power control via ensembling deep neural networks. *IEEE Transactions on Communications*, 2019.
- [19] Yanfeng Zhu, Qian Zhang, Zhisheng Niu, and Jing Zhu. On optimal QoS-aware physical carrier sensing for IEEE 802.11 based WLANs: Theoretical analysis and protocol design. *IEEE transactions on wireless communications*, 7(4):1369–1378, 2008.
- [20] Hui Ma, Rajiv Vijayakumar, Sumit Roy, and Jing Zhu. Optimizing 802.11 wireless mesh networks based on physical carrier sensing. *IEEE/ACM Transactions on Networking*, 17(5):1550–1563, 2009.
- [21] Jing Deng, Ben Liang, and Pramod K Varshney. Tuning the carrier sensing range of ieee 802.11 mac. In *IEEE Global Telecommunications Conference, 2004. GLOBECOM’04.*, volume 5, pages 2987–2991. IEEE, 2004.
- [22] Xue Yang and Nitin Vaidya. On physical carrier sensing in wireless ad hoc networks. In *Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies.*, volume 4, pages 2525–2535. IEEE, 2005.
- [23] Hui Ma, Hamed MK Alazemi, and Sumit Roy. A stochastic model for optimizing physical carrier sensing and spatial reuse in wireless ad hoc networks. In *IEEE International Conference on Mobile Adhoc and Sensor Systems Conference, 2005.*, pages 8–pp. IEEE, 2005.

- [24] Soma Tayamon, Gustav Wikström, Kevin Perez Moreno, Johan Söder, Yu Wang, and Filip Mestanov. Analysis of the potential for increased spectral reuse in wireless lan. In *2015 IEEE 26th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pages 1963–1967. IEEE, 2015.
- [25] Imad Jamil, Laurent Cariou, and Jean-Francois Helard. Improving the capacity of future ieee 802.11 high efficiency wlans. In *2014 21st International Conference on Telecommunications (ICT)*, pages 303–307. IEEE, 2014.
- [26] Jing Zhu, Benjamin Metzler, Xingang Guo, and York Liu. Adaptive csma for scalable network capacity in high-density wlan: A hardware prototyping approach. In *Infocom*, 2006.
- [27] Ehsan Haghani, Michael N Krishnan, and Avideh Zakhor. Adaptive carrier-sensing for throughput improvement in ieee 802.11 networks. In *2010 IEEE Global Telecommunications Conference GLOBECOM 2010*, pages 1–6. IEEE, 2010.
- [28] Christina Thorpe, Sean Murphy, and Liam Murphy. Ieee802. 11k enabled adaptive carrier sense management mechanism (kapcs2). In *12th IFIP/IEEE International Symposium on Integrated Network Management (IM 2011) and Workshops*, pages 509–515. IEEE, 2011.
- [29] Liqun Fu, Soung Chang Liew, and Jianwei Huang. Effective carrier sensing in csma networks under cumulative interference. *IEEE Transactions on Mobile Computing*, 12(4):748–760, 2012.
- [30] Dong Min Kim and Seong-Lyun Kim. An iterative algorithm for optimal carrier sensing threshold in random csma/ca wireless networks. *IEEE communications letters*, 17(11):2076–2079, 2013.
- [31] Bo Yin, Koji Yamamoto, Takayuki Nishio, Masahiro Morikura, and Hirantha Abeysekera. Learning-based spatial reuse for wlans with early identification of interfering transmitters. *IEEE Transactions on Cognitive Communications and Networking*, 6(1):151–164, 2019.
- [32] Robert K Schmidt, Achim Brakemeier, Tim Leinmüller, Frank Kargl, and Günter Schäfer. Advanced carrier sensing to resolve local channel congestion. In *Proceedings of the Eighth ACM international workshop on Vehicular inter-networking*, pages 11–20, 2011.
- [33] Kyung-Joon Park, Jennifer C Hou, Tamer Basar, and Hwangnam Kim. Noncooperative carrier sense game in wireless networks. *IEEE Transactions on Wireless Communications*, 8(10):5280–5289, 2009.
- [34] G Smith. Dynamic sensitivity control-v2. *IEEE*, 802:802–11, 2015.
- [35] Tanguy Ropitault and Nada Golmie. Etp algorithm: Increasing spatial reuse in wireless lans dense environment using etx. In *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pages 1–7. IEEE, 2017.

- [36] Ioannis Selinis, Konstantinos Katsaros, Seiamak Vahid, and Rahim Tafazolli. Control OBSS/PD Sensitivity Threshold for IEEE 802.11 ax BSS Color. In *2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pages 1–7. IEEE, 2018.
- [37] Tanguy Ropitault. Evaluation of rtot algorithm: A first implementation of obss_pd-based sr method for ieee 802.11 ax. In *2018 15th IEEE Annual Consumer Communications & Networking Conference (CCNC)*, pages 1–7. IEEE, 2018.
- [38] Anastasios Valkanis, Athanasios Iossifides, Periklis Chatzimisios, Marios Angelopoulos, and Vasilis Katos. Ieee 802.11 ax spatial reuse improvement: An interference-based channel-access algorithm. *IEEE Vehicular Technology Magazine*, 14(2):78–84, 2019.
- [39] Phillip B Oni and Steven D Blostein. Ap association optimization and cca threshold adjustment in dense wlans. In *2015 IEEE Globecom Workshops (GC Wkshps)*, pages 1–6. IEEE, 2015.
- [40] Nakahira, Toshiro and Ishihara, Koichi and Asai, Yusuke and Takatori, Yasushi and Kudo, Riichi and Mizoguchi, Masato. Centralized control of carrier sense threshold and channel bandwidth in high-density WLANs. In *Microwave Conference (APMC), 2014 Asia-Pacific*, pages 570–572. IEEE, 2014.
- [41] Wessam Afifi, Enrico-Henrik Rantala, Esa Tuomaala, Sayantan Choudhury, and Marwan Krunz. Throughput-fairness tradeoff evaluation for next-generation wlans with adaptive clear channel assessment. In *2016 IEEE International Conference on Communications (ICC)*, pages 1–6. IEEE, 2016.
- [42] Yuan Li, Ke Li, Wenwen Li, Yan Zhang, Min Sheng, and Jianxiang Chu. An energy-efficient power control approach for ieee 802.11n wireless lans. In *2014 IEEE International Conference on Computer and Information Technology*, pages 49–53. IEEE, 2014.
- [43] Chevillat, Pierre and Jelitto, Jens and Truong, Hong Linh. Dynamic data rate and transmit power adjustment in IEEE 802.11 wireless LANs. *International Journal of Wireless Information Networks*, 12(3):123–145, 2005.
- [44] Md Manowarul Islam, Nobuo Funabiki, Rahardhita Widyatratna Sudibyo, Kwenga Ismael Munene, and Wen-Chung Kao. A dynamic access-point transmission power minimization method using pi feedback control in elastic wlan system for iot applications. *Internet of Things*, 8:100089, 2019.
- [45] Fabiano S Chaves, André M Cavalcante, Erika PL Almeida, Fuad M Abinader, Robson D Vieira, Sayantan Choudhury, and Klaus Doppler. Adaptive transmit power for wifi dense deployments. In *2014 IEEE 80th vehicular technology conference (VTC2014-Fall)*, pages 1–6. IEEE, 2014.
- [46] Carlos Gandarillas, Carlos Martín-Engeños, Héctor López Pombo, and Antonio G Marques. Dynamic transmit-power control for wifi access points based on wireless link occupancy. In *2014 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1093–1098. IEEE, 2014.

- [47] Chih-Yung Chang and Hsu-Ruey Chang. Power control and fairness mac mechanisms for 802.11 wlans. *Computer communications*, 30(7):1527–1537, 2007.
- [48] Suhua Tang, Akio Hasegawa, Riichiro Nagareda, Akito Kitaura, Tatsuo Shibata, and Sadao Obama. Improving throughput of wireless LANs with transmit power control and slotted channel access. In *Personal Indoor and Mobile Radio Communications (PIMRC), 2011 IEEE 22nd International Symposium on*, pages 834–838. IEEE, 2011.
- [49] Minseok Kim, Sungjin Shin, and Jong-Moon Chung. Distributed power control for enhanced spatial reuse in csma/ca based wireless networks. *IEEE Transactions on Wireless Communications*, 13(9):5015–5027, 2014.
- [50] Hiroyasu Shimizu, Bo Yin, Koji Yamamoto, Motoki Iwata, Takayuki Nishio, Masahiro Morikura, and Hirantha Abeysekera. Joint channel selection and spatial reuse for starvation mitigation in ieee 802.11 ax wlans. In *2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall)*, pages 1–6. IEEE, 2019.
- [51] Jean-Pierre Ebert, Björn Stremmel, Eckhardt Wiederhold, and Adam Wolisz. An energy-efficient power control approach for wlans. *Journal of Communications and Networks*, 2(3):197–206, 2000.
- [52] Xiaoying Lei and Seung Hyong Rhee. Performance enhancement of overlapping bsss via dynamic transmit power control. *EURASIP Journal on Wireless Communications and Networking*, 2015(1):8, 2015.
- [53] Wei Li, Yong Cui, Xiuzhen Cheng, Mznah A Al-Rodhaan, and Abdullah Al-Dhelaan. Achieving proportional fairness via AP power control in multi-rate WLANs. *IEEE Transactions on Wireless Communications*, 10(11):3784–3792, 2011.
- [54] Oghenekome Oteri, Pengfei Xia, Frank LaSita, and Robert Olesen. Advanced power control techniques for interference mitigation in dense 802.11 networks. In *2013 16th International symposium on wireless personal multimedia communications (WPMC)*, pages 1–7. IEEE, 2013.
- [55] Suhua Tang, Hiroyuki Yomo, Akio Hasegawa, Tatsuo Shibata, and Masayoshi Ohashi. Joint transmit power control and rate adaptation for wireless lans. *Wireless personal communications*, 74(2):469–486, 2014.
- [56] Roohollah Amiri, Mojtaba Ahmadi Almasi, Jeffrey G Andrews, and Hani Mehrpouyan. Reinforcement learning for self organization and power control of two-tier heterogeneous networks. *IEEE Transactions on Wireless Communications*, 18(8):3933–3947, 2019.
- [57] Koji Yamamoto, Xuedan Yang, Takayuki Nishio, Masahiro Morikura, and Hirantha Abeysekera. Analysis of inversely proportional carrier sense threshold and transmission power setting. In *2017 14th IEEE Annual Consumer Communications & Networking Conference (CCNC)*, pages 13–18. IEEE, 2017.

- [58] Motoki Iwata, Koji Yamamoto, Bo Yin, Takayuki Nishio, Masahiro Morikura, and Hirantha Abeysekera. Analysis of inversely proportional carrier sense threshold and transmission power setting based on received power for ieee 802.11 ax. In *2019 16th IEEE Annual Consumer Communications & Networking Conference (CCNC)*, pages 1–6. IEEE, 2019.
- [59] Jason A Fuemmeler, Nitin H Vaidya, and Venugopal V Veeravalli. Selecting transmit powers and carrier sense thresholds in csma protocols for wireless ad hoc networks. In *Proceedings of the 2nd annual international workshop on Wireless internet*, pages 15–es, 2006.
- [60] Imad Jamil, Laurent Cariou, and Jean-François Hélard. Preserving fairness in super dense wlans. In *2015 IEEE International Conference on Communication Workshop (ICCW)*, pages 2276–2281. IEEE, 2015.
- [61] Vivek P Mhatre, Konstantina Papagiannaki, and Francois Baccelli. Interference mitigation through power control in high density 802.11 wlans. In *IEEE INFOCOM 2007-26th IEEE International Conference on Computer Communications*, pages 535–543. IEEE, 2007.
- [62] Koki Iwai, Takanobu Ohnuma, Hiroshi Shigeno, and Yusuke Tanaka. Improving of fairness by dynamic sensitivity control and transmission power control with access point cooperation in dense wlan. In *2019 16th IEEE Annual Consumer Communications & Networking Conference (CCNC)*, pages 1–4. IEEE, 2019.
- [63] Imad Jamil, Laurent Cariou, and Jean-François Hélard. Efficient mac protocols optimization for future high density wlans. In *2015 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1054–1059. IEEE, 2015.
- [64] Imad Jamil, Laurent Cariou, and Jean-Fran Hélard. Novel learning-based spatial reuse optimization in dense WLAN deployments.
- [65] G Smith. Dsc and obss_pd. *Presentation doc. IEEE*, pages 802–11, 2017.
- [66] Masahito Mori et al. Performance analysis of bss color and dsc. *Nov*, 3:11–14, 2014.
- [67] Oghenekome Oteri, Frank La Sita, Rui Yang, Monisha Ghosh, and Robert Olesen. Improved spatial reuse for dense 802.11 wlans. In *2015 IEEE Globecom Workshops (GC Wkshps)*, pages 1–6. IEEE, 2015.
- [68] Jin Liu, Masahide Hatanaka, and Takao Onoye. A collision mitigation method on spatial reuse for wlan in a dense residential environment.
- [69] M Shahwaiz Afiaqui, Eduard Garcia-Villegas, Elena Lopez-Aguilera, Graham Smith, and Daniel Camps. Evaluation of dynamic sensitivity control algorithm for IEEE 802.11 ax. In *Wireless Communications and Networking Conference (WCNC), 2015 IEEE*, pages 1060–1065. IEEE, 2015.
- [70] M Shahwaiz Afiaqui, Eduard Garcia-Villegas, Elena Lopez-Aguilera, and Daniel Camps-Mur. Dynamic sensitivity control of access points for IEEE 802.11 ax. In *Communications (ICC), 2016 IEEE International Conference on*, pages 1–7. IEEE, 2016.

- [71] Parag Kulkarni and Fengming Cao. Taming the densification challenge in next generation wireless LANs: An investigation into the use of dynamic sensitivity control. In *Wireless and Mobile Computing, Networking and Communications (WiMob), 2015 IEEE 11th International Conference on*, pages 860–867. IEEE, 2015.
- [72] Zhenzhe Zhong, Fengming Cao, Parag Kulkarni, and Zhong Fan. Promise and perils of dynamic sensitivity control in ieee 802.11 ax wlans. In *2016 International Symposium on Wireless Communication Systems (ISWCS)*, pages 439–444. IEEE, 2016.
- [73] Kiryanov A. Krotov A. Gallo P. Garlisi D. Khorov, E. and I. Tinnirello. Joint usage of dynamic sensitivity control and time division multiple access in dense 802.11 ax networks. In *In International Workshop on Multiple Access Communications*, pages 57–71. Springer, Cham., 2016.
- [74] Ioannis Selinis, Marcin Filo, Seiamak Vahid, Jonathan Rodriguez, and Rahim Tafazolli. Evaluation of the DSC algorithm and the BSS color scheme in dense cellular-like IEEE 802.11 ax deployments. In *Personal, Indoor, and Mobile Radio Communications (PIMRC), 2016 IEEE 27th Annual International Symposium on*, pages 1–7. IEEE, 2016.
- [75] Ioannis Selinis, Konstantinos Katsaros, Seiamak Vahid, and Rahim Tafazolli. Exploiting the Capture Effect on DSC and BSS Color in Dense IEEE 802.11 ax Deployments. In *Proceedings of the Workshop on ns-3*, pages 47–54. ACM, 2017.
- [76] Yun Wen, Hiroshi Fujita, and Dai Kimura. Throughput-aware dynamic sensitivity control algorithm for next generation wlan system. In *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pages 1–7. IEEE, 2017.
- [77] Jaha Mvulla and Eun-Chan Park. Enhanced dual carrier sensing with transmission time control for fair spatial reuse in heterogeneous and dense wlans. *IEEE Access*, 6:22140–22155, 2018.
- [78] Qiao Qu, Bo Li, Mao Yang, Zhongjiang Yan, Annan Yang, Jian Yu, Ming Gan, Yunbo Li, Xun Yang, Osama Aboul-Magd, et al. Survey and Performance Evaluation of the Upcoming Next Generation WLAN Standard-IEEE 802.11 ax. *arXiv preprint arXiv:1806.05908*, 2018.
- [79] Zhao Shen, Bo Li, Mao Yang, Zhongjiang Yan, Xiaobo Li, and Yi Jin. Research and Performance Evaluation of Spatial Reuse Technology for Next Generation WLAN. In *International Wireless Internet Conference*, pages 41–51. Springer, 2018.
- [80] Arjun Malhotra, Mukulika Maity, and Avik Dutta. How much can we reuse? an empirical analysis of the performance benefits achieved by spatial-reuse of ieee 802.11 ax. In *2019 11th International Conference on Communication Systems & Networks (COM-SNETS)*, pages 432–435. IEEE, 2019.
- [81] Lorenzo Galati Giordano Eloise de Carvalho Rodrigues, Adrian Garcia-Rodriguez and Giovanni Geraci. On the latency of ieee 802.11ax wlans with parameterized spatial reuse. 2020.

- [82] Qiao Qu, Bo Li, Mao Yang, Zhongjiang Yan, Annan Yang, Der-Jiunn Deng, and Kwang-Cheng Chen. Survey and performance evaluation of the upcoming next generation wlans standard-ieee 802.11 ax. *Mobile Networks and Applications*, 24(5):1461–1474, 2019.
- [83] Coordinated spatial reuse operation. <https://mentor.ieee.org/802.11/dcn/20/11-20-0033-01-00be-coordinated-spatial-reuse-operation.pptx>. Accessed: 2020-06-12.
- [84] Terminology for AP Coordination. <https://mentor.ieee.org/802.11/dcn/18/11-18-1926-02-0eht-terminology-for-ap-coordination.pptx>. Accessed: 2020-06-12.
- [85] Downlink spatial reuse parameter framework with coordinated beamforming and null steering for 802.11be. <https://mentor.ieee.org/802.11/dcn/19/11-19-1779-05-00be-downlink-spatial-reuse-parameter-framework-with-o.pptx>. Accessed: 2020-06-12.
- [86] AP coordination in EHT. <https://mentor.ieee.org/802.11/dcn/19/11-19-0801-00-00be-ap-coordination-in-eht.pptx>. Accessed: 2020-06-16.
- [87] Coordinated Beamforming for 802.11be. <https://mentor.ieee.org/802.11/dcn/20/11-20-0099-01-00be-coordinated-beamforming-for-802-11be.pptx>. Accessed: 2020-06-16.
- [88] Coordinated OFDMA Operation. <https://mentor.ieee.org/802.11/dcn/19/11-19-1788-00-00be-coordinated-ofdma-operation.pptx>. Accessed: 2020-06-16.
- [89] Mohammad Abu Alsheikh, Shaowei Lin, Dusit Niyato, and Hwee-Pink Tan. Machine learning in wireless sensor networks: Algorithms, strategies, and applications. *IEEE Communications Surveys & Tutorials*, 16(4):1996–2018, 2014.
- [90] Mario Bkassiny, Yang Li, and Sudharman K Jayaweera. A survey on machine-learning techniques in cognitive radios. *IEEE Communications Surveys & Tutorials*, 15(3):1136–1159, 2013.
- [91] Wenbo Wang, Andres Kwasinski, Dusit Niyato, and Zhu Han. A survey on applications of model-free strategy learning in cognitive wireless networks. *IEEE Communications Surveys & Tutorials*, 18(3):1717–1757, 2016.
- [92] Chunxiao Jiang *et al.* Machine learning paradigms for next-generation wireless networks. *IEEE Wireless Comm.*, 24(2):98–105, 2016.
- [93] Chaoyun Zhang, Paul Patras, and Hamed Haddadi. “Deep learning in mobile and wireless networking: A survey”. *IEEE Comm. Surveys & Tutorials*, 2019.
- [94] Muhammad Usama *et al.* “Unsupervised machine learning for networking: Techniques, applications and research challenges”. *IEEE Access*, 7:65579–65615, 2019.

- [95] Paulo Valente Klaine, Muhammad Ali Imran, Oluwakayode Onireti, and Richard Demo Souza. A survey of machine learning techniques applied to self-organizing cellular networks. *IEEE Communications Surveys & Tutorials*, 19(4):2392–2431, 2017.
- [96] Manuel Eugenio Morocho Cayamcela and Wansu Lim. Artificial intelligence in 5g technology: A survey. In *2018 International Conference on Information and Communication Technology Convergence (ICTC)*, pages 860–865. IEEE, 2018.
- [97] Jessica Moysen and Lorenza Giupponi. From 4g to 5g: Self-organized network management meets machine learning. *Computer Communications*, 129:248–268, 2018.
- [98] Suzhi Bi *et al.* “Wireless communications in the era of big data”. *IEEE Comm. Magazine*, 53(10):190–199, 2015.
- [99] I Chih-Lin *et al.* “The Big-Data-Driven Intelligent Wireless Network: Architecture, Use Cases, Solutions, and Future Trends”. *IEEE Vehic. Tech. Magazine*, 12(4):20–29, 2017.
- [100] Mowei Wang *et al.* “Machine learning for networking: Workflow, advances and opportunities”. *IEEE Network*, 32(2):92–99, 2018.
- [101] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.
- [102] Jakub Konečný, H Brendan McMahan, Felix X Yu, Peter Richtárik, Ananda Theertha Suresh, and Dave Bacon. Federated learning: Strategies for improving communication efficiency. *arXiv preprint arXiv:1610.05492*, 2016.
- [103] Virginia Smith, Chao-Kai Chiang, Maziar Sanjabi, and Ameet S Talwalkar. Federated multi-task learning. In *Advances in Neural Information Processing Systems*, pages 4424–4434, 2017.
- [104] Chaoyun Zhang and Paul Patras. Long-term mobile traffic forecasting using deep spatio-temporal neural networks. In *Proceedings of the Eighteenth ACM International Symposium on Mobile Ad Hoc Networking and Computing*, pages 231–240, 2018.
- [105] Sebastian Troia, Rodolfo Alvizu, Youduo Zhou, Guido Maier, and Achille Pattavina. Deep learning-based traffic prediction for network optimization. In *2018 20th International Conference on Transparent Optical Networks (ICTON)*, pages 1–4. IEEE, 2018.
- [106] Udita Paul, Jiamo Liu, Sebastian Troia, Olabisi Falowo, and Guido Maier. Traffic-profile and machine learning based regional data center design and operation for 5g network. *Journal of Communications and Networks*, 21(6):569–583, 2019.
- [107] Bilong Shen, Xiaodan Liang, Yufeng Ouyang, Miaofeng Liu, Weimin Zheng, and Kathleen M Carley. Stepdeep: a novel spatial-temporal mobility event prediction framework based on deep neural network. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 724–733, 2018.
- [108] Honggang Zhang, Yuxiu Hua, Chujie Wang, Rongpeng Li, and Zhifeng Zhao. Deep learning based traffic and mobility prediction. *Machine Learning for Future Wireless Communications*, pages 119–136, 2020.

- [109] Lixin Li, Yang Xu, Jiaying Yin, Wei Liang, Xu Li, Wei Chen, and Zhu Han. Deep reinforcement learning approaches for content caching in cache-enabled d2d networks. *IEEE Internet of Things Journal*, 2019.
- [110] Zhiyuan Xu, Yanzhi Wang, Jian Tang, Jing Wang, and Mustafa Cenk Gursoy. A deep reinforcement learning based framework for power-efficient resource allocation in cloud trans. In *2017 IEEE International Conference on Communications (ICC)*, pages 1–6. IEEE, 2017.
- [111] Ji Li, Hui Gao, Tiejun Lv, and Yueming Lu. Deep reinforcement learning based computation offloading and resource allocation for mec. In *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, pages 1–6. IEEE, 2018.
- [112] Ying He, F Richard Yu, Nan Zhao, and Hongxi Yin. Secure social networks in 5g systems with mobile edge computing, caching, and device-to-device communications. *IEEE Wireless Communications*, 25(3):103–109, 2018.
- [113] Neelakantan Nurani Krishnan, Eric Torkildson, Narayan B Mandayam, Dipankar Raychaudhuri, Enrico-Henrik Rantala, and Klaus Doppler. Optimizing throughput performance in distributed mimo wi-fi networks using deep reinforcement learning. *IEEE Transactions on Cognitive Communications and Networking*, 2019.
- [114] 3GPP TR 23.791 V16.2.0 (2019-06). “Study of Enablers for Network Automation for 5G”. 2019.
- [115] ETSI GS ZSM 002 V0.13.5 (2019-07). Draft “Zero-touch network and Service Management (ZSM); Reference Architecture”. 2019.
- [116] ITU-T Rec. Y.3172, “Architectural framework for machine learning in future networks including IMT-2020”, 2019.
- [117] ITU-T Rec. Y.3174, “Framework for data handling to enable machine learning in future networks including IMT-2020”, 2019.
- [118] William R Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294, 1933.
- [119] Robert R Bush and Frederick Mosteller. A stochastic model with applications to learning. *The Annals of Mathematical Statistics*, pages 559–585, 1953.
- [120] Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multi-armed bandit problem. *Machine learning*, 47(2-3):235–256, 2002.
- [121] Nicolo Cesa-Bianchi and Gabor Lugosi. *Prediction, learning, and games*. Cambridge university press, 2006.
- [122] John Gittins, Kevin Glazebrook, and Richard Weber. *Multi-armed bandit allocation indices*. John Wiley & Sons, 2011.

- [123] Sébastien Bubeck, Nicolo Cesa-Bianchi, et al. Regret analysis of stochastic and non-stochastic multi-armed bandit problems. *Foundations and Trends® in Machine Learning*, 5(1):1–122, 2012.
- [124] Tor Lattimore and Csaba Szepesvári. Bandit algorithms. *preprint*, 2018.
- [125] Aleksandrs Slivkins. Introduction to multi-armed bandits. *arXiv preprint arXiv:1904.07272*, 2019.
- [126] Yi Gai, Bhaskar Krishnamachari, and Rahul Jain. Learning multiuser channel allocations in cognitive radio networks: A combinatorial multi-armed bandit formulation. In *New Frontiers in Dynamic Spectrum, 2010 IEEE Symposium on*, pages 1–9. IEEE, 2010.
- [127] Cem Tekin and Mingyan Liu. Online learning in opportunistic spectrum access: A restless bandit approach. In *2011 Proceedings IEEE INFOCOM*, pages 2462–2470. IEEE, 2011.
- [128] Yunxia Chen, Qing Zhao, Vikram Krishnamurthy, and Dejan Djonin. Transmission scheduling for sensor network lifetime maximization: A shortest path bandit formulation. In *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, volume 4, pages IV–IV. IEEE, 2006.
- [129] Marc Carrascosa and Boris Bellalta. Decentralized ap selection using multi-armed bandits: Opportunistic $\{\backslash\text{epsilon}\}$ -greedy with stickiness. *arXiv preprint arXiv:1903.00281*, 2019.
- [130] Keqin Liu and Qing Zhao. A restless bandit formulation of opportunistic access: Indexability and index policy. In *2008 5th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks Workshops*, pages 1–5. IEEE, 2008.
- [131] Animashree Anandkumar, Nithin Michael, Ao Kevin Tang, and Ananthram Swami. Distributed algorithms for learning and cognitive medium access with logarithmic regret. *IEEE Journal on Selected Areas in Communications*, 29(4):731–745, 2011.
- [132] Jonathan Rosenski, Ohad Shamir, and Liran Szlak. Multi-player bandits—a musical chairs approach. In *International Conference on Machine Learning*, pages 155–163, 2016.
- [133] Setareh Maghsudi and Sławomir Stańczak. Channel selection for network-assisted d2d communication via no-regret bandit learning with calibrated forecasting. *IEEE Transactions on Wireless Communications*, 14(3):1309–1322, 2015.
- [134] Setareh Maghsudi and Sławomir Stańczak. Joint channel selection and power control in infrastructureless wireless networks: A multiplayer multiarmed bandit framework. *IEEE Transactions on Vehicular Technology*, 64(10):4565–4578, 2014.
- [135] Setareh Maghsudi and Ekram Hossain. Distributed user association in energy harvesting dense small cell networks: A mean-field multi-armed bandit approach. *IEEE Access*, 5:3513–3523, 2017.

- [136] Marc Carrascosa and Boris Bellalta. Multi-armed bandits for decentralized ap selection in enterprise wlans. *arXiv preprint arXiv:2001.00392*, 2020.
- [137] Pierre Coucheney, Kinda Khawam, and Johanne Cohen. Multi-armed bandit for distributed inter-cell interference coordination. In *2015 IEEE International Conference on Communications (ICC)*, pages 3323–3328. IEEE, 2015.
- [138] Afef Feki and Veronique Capdevielle. Autonomous resource allocation for dense lte networks: A multi armed bandit formulation. In *2011 IEEE 22nd International Symposium on Personal, Indoor and Mobile Radio Communications*, pages 66–70. IEEE, 2011.
- [139] Jose A Ayala-Romero, Juan J Alcaraz, Andrea Zanella, and Michele Zorzi. Contextual bandit approach for energy saving and interference coordination in hetnets. In *2018 IEEE International Conference on Communications (ICC)*, pages 1–6. IEEE, 2018.
- [140] Tony Daher, Sana Ben Jemaa, and Laurent Decreusefond. Cognitive management of self-organized radio networks based on multi armed bandit. In *2017 IEEE 28th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, pages 1–5. IEEE, 2017.
- [141] Richard Combes and Alexandre Proutiere. Dynamic rate and channel selection in cognitive radio systems. *IEEE Journal on Selected Areas in Communications*, 33(5):910–921, 2014.
- [142] Yapeng Zhao, Hua Qian, Kai Kang, and Yanliang Jin. A non-stationary bandit strategy for rate adaptation with delayed feedback. *IEEE Access*, 2020.
- [143] Cristina Cano and Gergely Neu. Wireless optimisation via convex bandits: Unlicensed lte/wifi coexistence. In *Proceedings of the 2018 Workshop on Network Meets AI & ML*, pages 41–47, 2018.
- [144] Peter Auer, Nicolo Cesa-Bianchi, Yoav Freund, and Robert E Schapire. The nonstochastic multiarmed bandit problem. *SIAM journal on computing*, 32(1):48–77, 2002.
- [145] Neha Gupta, Ole-Christoffer Granmo, and Ashok Agrawala. Thompson sampling for dynamic multi-armed bandits. In *2011 10th International Conference on Machine Learning and Applications and Workshops*, volume 1, pages 484–489. IEEE, 2011.
- [146] Marco Di Felice, Kaushik Roy Chowdhury, and Luciano Bononi. Learning with the bandit: A cooperative spectrum selection scheme for cognitive radio networks. In *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*, pages 1–6. IEEE, 2011.
- [147] Kobi Cohen, Qing Zhao, and Anna Scaglione. Restless multi-armed bandits under time-varying activation constraints for dynamic spectrum access. In *Signals, Systems and Computers, 2014 48th Asilomar Conference on*, pages 1575–1578. IEEE, 2014.
- [148] Nadine Abbas, Youssef Nasser, and Karim El Ahmad. Recent advances on artificial intelligence and learning techniques in cognitive radio networks. *EURASIP Journal on Wireless Communications and Networking*, 2015(1):174, 2015.

- [149] Setareh Maghsudi and Sławomir Stańczak. Joint channel selection and power control in infrastructureless wireless networks: A multiplayer multiarmed bandit framework. *IEEE Transactions on Vehicular Technology*, 64(10):4565–4578, 2015.
- [150] Orly Avner and Shie Mannor. Concurrent bandits and cognitive radio networks. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 66–81. Springer, 2014.
- [151] Tze Leung Lai and Herbert Robbins. Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics*, 6(1):4–22, 1985.
- [152] Rajeev Agrawal. Sample mean based index policies with $O(\log n)$ regret for the multi-armed bandit problem. *Advances in Applied Probability*, pages 1054–1078, 1995.
- [153] Peter Auer, Nicolo Cesa-Bianchi, Yoav Freund, and Robert E Schapire. Gambling in a rigged casino: The adversarial multi-armed bandit problem. In *Proceedings of IEEE 36th Annual Foundations of Computer Science*, pages 322–331. IEEE, 1995.
- [154] Keqin Liu and Qing Zhao. Distributed learning in multi-armed bandit with multiple players. *IEEE Transactions on Signal Processing*, 58(11):5667–5681, 2010.
- [155] Dov Monderer and Lloyd S Shapley. Potential games. *Games and economic behavior*, 14(1):124–143, 1996.
- [156] Giuseppe Bianchi. Performance analysis of the ieee 802.11 distributed coordination function. *IEEE Journal on selected areas in communications*, 18(3):535–547, 2000.
- [157] Piyush Gupta and Panganmala R Kumar. The capacity of wireless networks. *IEEE Transactions on information theory*, 46(2):388–404, 2000.
- [158] Xingang Guo, Sumit Roy, and W Steven Conner. Spatial reuse in wireless ad-hoc networks. In *2003 IEEE 58th Vehicular Technology Conference. VTC 2003-Fall (IEEE Cat. No. 03CH37484)*, volume 3, pages 1437–1442. IEEE, 2003.
- [159] Ramin Hekmat and Piet Van Mieghem. Interference in wireless multi-hop ad-hoc networks and its effect on network capacity. *Wireless Networks*, 10(4):389–399, 2004.
- [160] Yong Yang, Jennifer C Hou, and L-C Kung. Modeling the effect of transmit power and physical carrier sense in multi-hop wireless networks. In *IEEE INFOCOM 2007-26th IEEE International Conference on Computer Communications*, pages 2331–2335. IEEE, 2007.
- [161] Hesham ElSawy, Ahmed Sultan-Salem, Mohamed-Slim Alouini, and Moe Z Win. Modeling and analysis of cellular networks using stochastic geometry: A tutorial. *IEEE Communications Surveys & Tutorials*, 19(1):167–203, 2016.
- [162] Xiaoguang Zhao, Xiangming Wen, Tao Lei, Zhaoming Lu, and Biao Zhang. On stochastic geometry analysis of dense wlan with dynamic carrier sense threshold and rate control. In *2016 19th International Symposium on Wireless Personal Multimedia Communications (WPMC)*, pages 211–216. IEEE, 2016.

- [163] Zhiwei Zhang, Yunzhou Li, Kaizhi Huang, and Chen Liang. On stochastic geometry modeling of wlan capacity with dynamic sensitive control. In *2015 13th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, pages 78–83. IEEE, 2015.
- [164] Zhenzhe Zhong and Fengming Cao. Stochastic analysis of 802.11 uplink with dynamic sensitivity control. In *2016 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6. IEEE, 2016.
- [165] Motoki Iwata, Koji Yamamoto, Bo Yin, Takayuki Nishio, Masahiro Morikura, and Hirantha Abeysekera. Stochastic geometry analysis of individual carrier sense threshold adaptation in ieee 802.11 ax wlans. *IEEE Access*, 7:161916–161927, 2019.
- [166] Boris Bellalta, Alessandro Zocca, Cristina Cano, Alessandro Checco, Jaume Barcelo, and Alexey Vinel. Throughput analysis in CSMA/CA networks using continuous time Markov networks: a tutorial. In *Wireless Networking for Moving Objects*, pages 115–133. Springer, 2014.
- [167] Boris Bellalta. Throughput Analysis in High Density WLANs. *IEEE Communications Letters*, 21(3):592–595, 2017.
- [168] Piotr Gawłowicz and Anatolij Zubow. ns-3 meets openai gym: The playground for machine learning in networking research. In *Proceedings of the 22nd International ACM Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, pages 113–120, 2019.
- [169] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.
- [170] Christopher JCH Watkins and Peter Dayan. Q-learning. *Machine learning*, 8(3-4):279–292, 1992.
- [171] Dmitry Bankov, Andre Didenko, Evgeny Khorov, and Andrey Lyakhov. OFDMA Uplink Scheduling in IEEE 802.11 ax Networks. In *2018 IEEE International Conference on Communications (ICC)*, pages 1–6. IEEE, 2018.
- [172] Konstantinos Dovelos and Boris Bellalta. Optimal Resource Allocation in IEEE 802.11ax Uplink OFDMA with Scheduled Access. *arXiv preprint arXiv:1811.00957*, 2019.
- [173] Ruizhi Liao, Boris Bellalta, Miquel Oliver, and Zhisheng Niu. MU-MIMO MAC protocols for wireless local area networks: A survey. *IEEE Communications Surveys & Tutorials*, 18(1):162–183, 2016.
- [174] Maddalena Nurchis and Boris Bellalta. Target wake time: scheduled access in IEEE 802.11 ax WLANs. *IEEE Wireless Communications*, 26(2):142–150, 2019.

Implications of Decentralized Q-learning Resource Allocation in Wireless Networks

Francesc Wilhelmi, Boris Bellalta, Cristina Cano, and Anders Jonsson

Abstract

Reinforcement Learning is gaining attention by the wireless networking community due to its potential to learn good-performing configurations only from the observed results. In this work we propose a stateless variation of Q-learning, which we apply to exploit spatial reuse in a wireless network. In particular, we allow networks to modify both their transmission power and the channel used solely based on the experienced throughput. We concentrate in a completely decentralized scenario in which no information about neighbouring nodes is available to the learners. Our results show that although the algorithm is able to find the best-performing actions to enhance aggregate throughput, there is high variability in the throughput experienced by the individual networks. We identify the cause of this variability as the adversarial setting of our setup, in which the most played actions provide intermittent good/poor performance depending on the neighbouring decisions. We also evaluate the effect of the intrinsic learning parameters of the algorithm on this variability.

1 Introduction

Reinforcement Learning (RL) has recently spread use in the wireless communications field to solve many kinds of problems such as Access Point (AP) association [1], channel selection [2] or transmit power adjustment [3], as it allows learning good-performing configurations only from the observed results. Among these, Q-learning has been applied to dynamic channel assignment in mobile networks in [4] and to automatic channel selection in Femto Cell networks in [5]. However, to the best of our knowledge, the case of a fully decentralized scenario where nodes do not have knowledge from each other, has not yet been considered.

In this work we propose a stateless variation of Q-learning in which nodes select the transmission power and channel to use solely based on their resulting throughput. We concentrate on a fully decentralized scenario where no information about the actions and resulting performance of the other nodes is available to the learners. Note that inferring the throughput of neighbouring nodes allocated to different channels is costly as periodic sensing in the other channels would then be needed. We aim to characterize the performance of Q-learning in

such scenarios, obtaining insight on the most played actions (i.e., channel and transmit power selected) and the resulting performance. We observe that when no information about the neighbours is available to the learners, these will tend to apply selfish strategies that result in alternating good/poor performance depending on the actions of the others. In such scenarios, we show that the use of Q-learning allows each network to find the best-performing actions, though without reaching a steady solution. Note that achieving a steady solution in a decentralized environment relies in finding a Nash Equilibrium, a concept used in Game Theory to define a set of individual strategies that maximize the profits of each player in a non-cooperative game, regardless of the others' strategy. Formally, a set of best player actions $a^* = (a_1^*, \dots, a_n^*) \in A$ leads to a Nash Equilibrium if $a_i^* \in B_i(a_{-i}^*), \forall i \in N$, where $B_i(a_{-i})$ is the best response to the others actions (a_{-i}) . Thus, the consequences of not reaching a Nash Equilibrium can have an impact on performance variability.

In addition, we look at the resulting performance in terms of throughput when varying several parameters intrinsic to the learning algorithm, which helps in understanding the interactions between the degree of exploration and learning rate, and the variability of the resulting performance.

The remaining of this document is structured as follows: Section 2 introduces the simulation scenario and considerations. Then, Section 3 presents our Stateless variation of Q-learning and its practical implementation for the resource allocation problem in Wireless Networks (WNs). Simulation results are later discussed in Section 4. Finally, some final remarks are provided in Section 5.

2 System model

For the remainder of this work, we consider a scenario in which several WNs are placed in a 3D-map (with parameters described later in Section 4.1), each one formed by an Access Point (AP) transmitting to a single Station (STA) in downlink manner.

2.1 Channel modelling

Path-loss and shadowing effects are modelled using the log-distance model for indoor communications. The path-loss between WN i and j is given by

$$\begin{aligned} PL_{i,j} &= P_{tx,i} - P_{rx,j} = \\ &= PL_0 + 10\alpha_{PL} \log_{10}(d_{i,j}) + G_s + \frac{d_{i,j}}{d_{obs}} G_o, \end{aligned}$$

where $P_{tx,i}$ is the transmitted power in dBm by WN i , $P_{rx,j}$ is the power in dBm received in WN j , PL_0 is the path-loss at one meter in dB, α_{PL} is the path-loss exponent, $d_{i,j}$ is the distance between the transmitter and the receiver in meters, G_s is the shadowing loss in dB, and G_o is the obstacles loss in dB. Note

that we include the factor d_{obs} , which is the distance between two obstacles in meters.

2.2 Throughput calculation

By using the power received and the interference, we calculate the maximum theoretical throughput of each WN i at time $t \in \{1, 2, \dots\}$ by using the Shannon Capacity.

$$\Gamma_{i,t} = B \log_2(1 + \text{SINR}_{i,t}),$$

where B is the channel bandwidth and the experienced Signal to Interference plus Noise Ratio (SINR) is given by:

$$\text{SINR}_{i,t} = \frac{P_{i,t}}{I_{i,t} + N},$$

where $P_{i,t}$ and $I_{i,t}$ are the received power and the sum of the interference at WN i at time t , respectively, and N is the floor noise power. For each STA in a WN, the interference is considered to be the total power received from all the APs of the other coexisting WNs as if they were continuously transmitting. Adjacent channel interference is also considered in $I_{i,t}$, $i \in \{1, \dots, W\}$, where W is the number of neighbouring WNs. We consider that the transmitted power leaked to adjacent channels is 20 dBm lower for each channel separation.

3 Decentralized Stateless Q-learning for enhancing Spatial Reuse in WNs

Q-learning [6] is an RL technique that enables an agent to learn the optimal policy to follow in a given environment. A set of possible states describing the environment and actions are defined in this model. In particular, an agent maintains an estimate of the expected long-term discounted reward for each state-action pair, and selects actions with the aim of maximizing it. The expected cumulative reward $V^\pi(s)$ is given by:

$$V^\pi(s) = \lim_{N \rightarrow \infty} \mathbb{E} \left(\sum_{t=1}^N r_t^\pi(s) \right),$$

where $r_t^\pi(s)$ is the reward obtained at iteration t after starting from state s and by following policy π . Since the reward may easily get unbounded, a discount factor parameter ($\gamma < 1$) is used. The optimal policy π^* that maximizes the total expected reward is given by the Bellman's Optimality Equation [?]:

$$Q^*(s, a) = \mathbb{E} \left\{ r_{t+1} + \gamma \max_{a'} Q^*(s_{t+1}, a') | s_t = s, a_t = a \right\}.$$

Henceforth, Q-learning receives information about the current state-action tuple (s_t, a_t) , the generated reward r_t and the next state s_{t+1} , in order to update the

Q-table:

$$\hat{Q}(s_t, a_t) \leftarrow (1 - \alpha_t)\hat{Q}(s_t, a_t) + \alpha_t \left(r_t + \gamma \left(\max_{a'} \hat{Q}(s_{t+1}, a') \right) \right),$$

where α_t is the learning rate at time t , and $\max_{a'} \hat{Q}(s_{t+1}, a')$ is the best estimated value for the next state s_{t+1} . The optimal solution is theoretically achieved with probability 1 if $\sum_{t=0}^{\infty} \alpha_t = \infty$, and $\sum_{t=0}^{\infty} \alpha_t^2 < \infty$, which satisfies that $\lim_{t \rightarrow \infty} \hat{Q}(s, a) = Q^*(s, a)$. Since we focus on a completely decentralized scenario where no information about the other nodes is available, the system can then be fully described by the set of actions and rewards.¹ Thus, we propose using a stateless variation of the original Q-learning algorithm. To implement decentralized learning to the resource allocation problem, we consider each WN to be an agent running Stateless Q-learning through an ε -greedy action-selection strategy, so that actions $a \in \mathcal{A}$ correspond to all the possible configurations that can be chosen with respect to the channel and transmit power. During the learning process we assume that WNs select actions sequentially, so that at each learning iteration, every agent takes an action in an ordered way. The order at which WNs choose an action at each iteration is randomly selected at the beginning of it. The reward after choosing an action is set as:

$$r_{i,t} = \frac{\Gamma_{i,t}}{\Gamma_i^*},$$

where $\Gamma_{i,t}$ is the experienced throughput at time t by WN $i \in \{1, \dots, n\}$, being n the number of WNs in the scenario, and $\Gamma_i^* = B \log_2(1 + \text{SNR}_i)$ is WN i maximum achievable throughput (i.e., when it uses the maximum transmission power and there is no interference). Each WN applies the Stateless Q-learning as follows:

- Initially, it sets the estimates of its actions $k \in \{1, \dots, K\}$ to 0: $\hat{Q}(a_k) = 0$.
- At each iteration, it applies an action by following the ε -greedy strategy, i.e., it selects the best-rewarding action with probability $1 - \varepsilon_t$, and a random one (uniformly distributed) the rest of the times.
- After choosing action a_k , it observes the generated reward (the relative experienced throughput), and updates the estimated value $\hat{Q}(a_k)$.
- Finally, ε_t is updated to follow a decreasing sequence: $\varepsilon_t = \frac{\varepsilon_0}{\sqrt{t}}$.

Note, as well, that the optimal policy cannot be derived for the presented scenario, but it can be approximated to enhance spatial reuse. This is due to the nature of the presented environment, as well as WNs decisions affect the others performance. Formally, the implementation details of Stateless Q-learning are described in Algorithm 1. The presented learning approach is intended

¹Local information like the observed instantaneous channel quality could be incorporated in the state definition. However, such a description of the system entails increased complexity.

to operate at the PHY level, allowing the operation of the current MAC-layer communication standards (e.g., in IEEE 802.11 WLANs, the channel access is governed by the CSMA/CA operation, so that Stateless Q-learning may contribute to improve spatial reuse at the PHY level).

Algorithm 1: Stateless Q-learning

```

1 Function Stateless Q-learning (SINR,  $\mathcal{A}$ );
  Input : SINR: Signal-to-Interference-plus-Noise Ratio sensed at the
           STA
            $\mathcal{A}$ : set of possible actions in {1, ..., K}
  Output:  $\bar{\Gamma}$ : Mean throughput experienced in the WN
2 initialize:  $t = 0$ ,  $\hat{Q}(a_k) = 0, \forall a_k \in \mathcal{A}$ 
3 while active do
4   Select  $a_k$   $\begin{cases} \underset{k=1,\dots,K}{\operatorname{argmax}} \hat{Q}(a_k), & \text{with prob } 1 - \varepsilon \\ i \sim \mathcal{U}(1, K), & \text{otherwise} \end{cases}$ 
5   Observe reward  $r_{a_k} = \frac{\Gamma_{a_k,t}}{\Gamma^*}$ 
6    $\hat{Q}(a_k) \leftarrow \hat{Q}(a_k) + \alpha \cdot (r_{a_k} + \gamma \cdot \max \hat{Q} - \hat{Q}(a_k))$ 
7    $\varepsilon_t \leftarrow \varepsilon_0 / \sqrt{t}$ 
8    $t \leftarrow t + 1$ 
9 end

```

4 Performance Evaluation

In this section we introduce the simulation parameters, describe the experiments.² and show the main results.

4.1 Simulation Parameters

According to [7], a typical high-density scenario for residential buildings contains 0.0033APs/m³. We then consider a map scenario with dimensions $10 \times 5 \times 10$ m containing 4 WNs that form a grid topology in which STAs are placed at the maximum possible distance from the other networks. This toy scenario allows us to study the performance of Stateless Q-learning in a controlled environment , which is useful to check the applicability of RL in WNs by only using local information ³. We consider that the number of channels is equal to half the number of coexisting WNs, so that we can study a challenging situation regarding the spatial reuse. Table 1 details the parameters used.

²The code used for simulations can be found at https://github.com/wn-upf/Decentralized_Qlearning_Resource_Allocation_in_WNs.git (Commit: eb4042a1830c8ea30b7eae3d72a51afe765a8d86).

³The analysis of the presented learning mechanisms in more congested scenarios is left as future work.

Parameter	Value
Map size (m)	$10 \times 5 \times 10$
Number of coexistent WNs	4
APs/STAs per WN	1 / 1
Distance AP-STA (m)	$\sqrt{2}$
Number of Channels	2
Channel Bandwidth (MHz)	20
Initial channel selection model	Uniformly distributed
Transmit power values (dBm)	{5, 10, 15, 20}
PL_0 (dB)	5
α_{PL}	4.4
G_s (dB)	Normally distributed with mean 9.5
G_o (dB)	Uniformly distributed with mean 30
d_{obs} (meters between two obstacles)	5
Noise level (dBm)	-100
Traffic model	Full buffer (downlink)

Table 1: Simulation parameters

4.2 Optimal solution

We first identify the optimal solutions that maximize: *i*) the aggregate throughput, and *ii*) the proportional fairness, which is computed as the logarithmic sum of the throughput experienced by each WN, i.e., $PF = \max_{k \in \mathcal{A}} \sum_i \log(\Gamma_{i,k})$. The optimal solutions are listed in Table 2. Note that, since the considered scenario is symmetric, there are two equivalent solutions. Note, as well, that in order to maximize the aggregate network throughput two of the WNs sacrifice themselves by choosing a lower transmit power. This result is then not likely to occur in an adversarial selfish setting.

WN id	Action that maximizes the Aggregate Throughput	Action that maximizes the Proportional Fairness
1	1 (2)	7 (8)
2	1 (2)	8 (7)
3	7 (8)	7 (8)
4	8 (7)	8 (7)

Table 2: Optimal configurations (action indexes) to achieve the maximum network throughput and prop. fairness, resulting in 1124 Mbps and 891 Mbps, respectively. In parenthesis the analogous solution is shown. Actions indexes range from 1 to 8 are mapped to {channel number, transmit power (dBm)}: {1,5}, {2,5}, {1,10}, {2,10}, {1,15}, {2,15}, {1,20} and {2,20}, respectively.

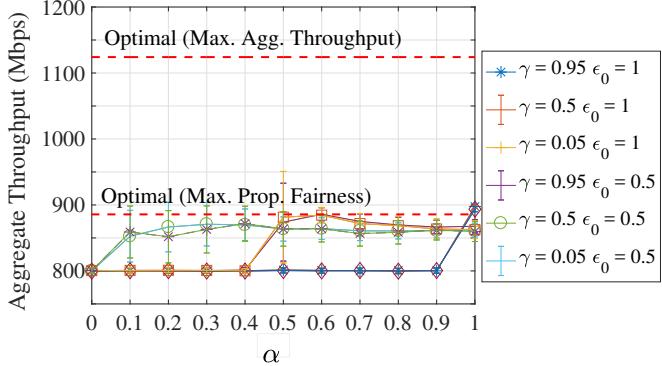


Figure 1: Effect of α , γ and ϵ_0 in the average aggregate throughput (100 simulation runs per sample).

4.3 Input Parameters Analysis

We first analyse the effects of modifying α (learning rate), γ (discount factor) and ϵ_0 (initial exploration coefficient) with respect to the achieved network throughput. We run simulations of 10000 iterations and capture the results of the last 5000 iterations to ensure that the initial transitory phase has ended. Each simulation is repeated 100 times for averaging purposes.

Figure 1 shows the average aggregate throughput achieved for each of the proposed combinations. It can be observed that the best results with respect to the aggregate throughput, regarding both average and variance, are achieved when $\alpha = 1$, $\gamma = 0.95$ and $\epsilon_0 = 1$. This means that for achieving the best results (i.e., high average aggregate throughput and low variance), the immediate reward of a given action must be considered rather than any previous information ($\alpha = 1$). We see that the difference between the pay-off offered by the best action and the current one must also be high ($\gamma = 0.95$). In addition, exploration must be highly boosted at the beginning ($\epsilon_0 = 1$). For this setting, the resulting throughput (902.739 Mbps) represents 80.29% of the one provided by the optimal configuration that maximizes the aggregate throughput (shown in Table 2). Regarding proportional fairness, the algorithm's resulting throughput is only 1.32% higher than the optimal.

We also evaluate the relationship between different values of α and γ in the average aggregate throughput and standard deviation (shown in Figure 2). We observe a remarkably higher aggregate throughput when $\alpha > \gamma$. We also see that the variability between different simulation runs is much lower when the average throughput is higher. Additionally, we note a peak in the standard deviation when $\gamma \approx \alpha$ and $\gamma > \alpha$.

To further understand the effects of modifying each of the aforementioned parameters, we show for different ϵ_0 , α and γ : *i*) the individual throughput experienced by each WN during the total 10000 iterations of a single simulation

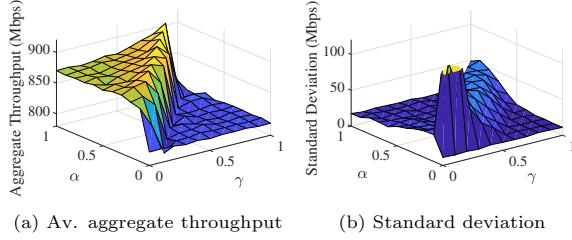


Figure 2: Evaluation of α and γ .

run (Figure 3), *ii*) the average throughput experienced by each WN for the last 5000 iterations, also for a single simulation run (Figure 4), and *iii*) the probability of choosing each action at each WN (Figure 5). We observe the following aspects:

- In Figure 3 a high variability of the throughput experienced by each WN can be observed, specially if ϵ_0 is high (as in Figures 3(a), 3(c)). A high degree of exploration allows WNs to discover changes in the resulting performance of their actions due to the activity of the other nodes, which at the same time generates more variability (WN adapt to changes in the environment).
- Despite the variability generated, we obtain fairer results for high ϵ_0 (Figure 4). Henceforth, there is a relationship between the variability generated and the average throughput fairness.
- Finally, in Figures 5(a) and 5(c) we observe that for the former, there are two favourite actions that are being played the most, but for the latter there is only one preferred action. The lower the learning rate (α), and consequently the discount factor (γ), the higher the probability of choosing a unique action, which results to be the one that provided the best performance in the past. The opposite occurs for higher α and γ values, since giving more importance to the immediate reward allows for a reaction only to the recently-played actions of the neighbouring nodes: the algorithm is short-sighted.

5 Conclusions

Decentralized Q-learning can be used to improve spatial reuse in dense wireless networks, enhancing performance as a result of exploiting the most rewarding actions. We have shown in this article, by means of a toy scenario, that Stateless Q-learning in particular allows finding good-performing configurations that achieve close-to-optimal (in terms of throughput maximization and proportional fairness) solutions.

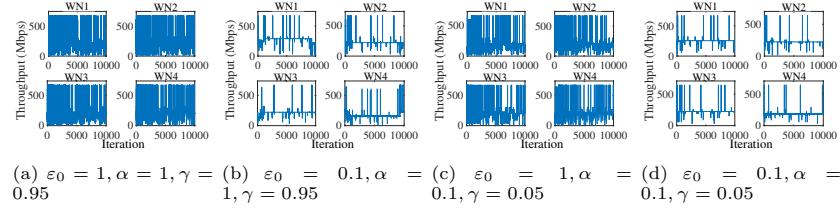


Figure 3: Individual throughput experienced by each WN during a single simulation run for different ε_0 , α and γ .

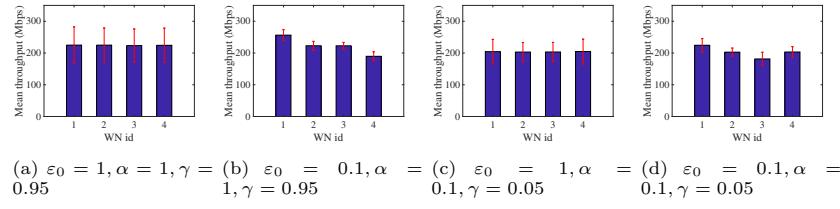


Figure 4: Average throughput experienced per WN during the last 5000 iterations of a total of 10000 iterations (in a single simulation run) and for different ε_0 , α and γ .

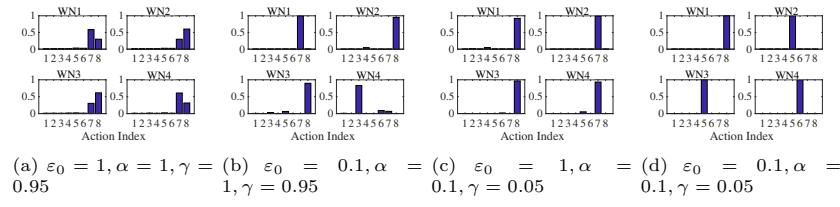


Figure 5: Probability of choosing the different actions at each WN for a single (10000 iterations) simulation run and different ε_0 , α and γ values

However, the competitiveness of the presented fully-decentralized environment involves the non-existence of a Nash Equilibrium. Thus, we have also identified high variability in the experienced individual throughput due to the constant changes of the played actions, motivated by the fact that the reward generated by each action changes according to the opponents' ones. We have evaluated the impact of the parameters intrinsic to the learning algorithm on this variability showing that it can be reduced by decreasing the exploration degree and learning rate. The individual reduction on the throughput variability occurs at the expense of losing aggregate performance. This variability can potentially result in negative effects on the overall WN's performance. Throughput fluctuation in higher layers of the protocol stack can have severe consequences depending on the time scale at which they occur (e.g., noticing high fluctuations may trigger congestion recovery procedures in TCP, which may harm the performance).

We left for future work to further extend the decentralized approach to find collaborative algorithms that allow the neighbouring WNs to reach an equilibrium that grants acceptable individual performance. Acquiring any kind of knowledge about the neighbouring WNs (by direct exchange or inferred from observation) is assumed to solve the decentralization variability issues. Furthermore, other learning approaches are intended to be analysed in the future for performance comparison in the resource allocation problem.

Acknowledgment

This work has been partially supported by the Spanish Ministry of Economy and Competitiveness under the Maria de Maeztu Units of Excellence Programme (MDM-2015-0502), and by the European Regional Development Fund under grant TEC2015-71303-R (MINECO/FEDER).

References

- [1] Chen, L. (2010, May). A distributed access point selection algorithm based on no-regret learning for wireless access networks. In Vehicular Technology Conference (VTC 2010-Spring), 2010 IEEE 71st (pp. 1-5). IEEE.
- [2] Maghsudi, S., & Stańczak, S. (2015). Channel selection for network-assisted D2D communication via no-regret bandit learning with calibrated forecasting. *IEEE Transactions on Wireless Communications*, 14(3), 1309-1322.
- [3] Maghsudi, S., & Stańczak, S. (2015). Joint channel selection and power control in infrastructureless wireless networks: A multiplayer multiarmed bandit framework. *IEEE Transactions on Vehicular Technology*, 64(10), 4565-4578.

- [4] Nie, J., & Haykin, S. (1999). A Q-learning-based dynamic channel assignment technique for mobile communication systems. *IEEE Transactions on Vehicular Technology*, 48(5), 1676-1687.
- [5] Bennis, M., & Niyato, D. (2010, December). A Q-learning based approach to interference avoidance in self-organized femtocell networks. In *GLOBECOM Workshops (GC Wkshps)*, 2010 IEEE (pp. 706-710). IEEE.
- [6] Watkins, C. J., & Dayan, P. (1992). Q-learning. *Machine learning*, 8 (3-4), 279-292.
- [7] Bellalta, B. "IEEE 802.11 ax: High-efficiency WLANs." *IEEE Wireless Communications* 23.1 (2016): 38-46.

Collaborative Spatial Reuse in Wireless Networks via Selfish Multi-Armed Bandits

Francesc Wilhelmi, Cristina Cano, Gergely Neu, Boris Bellalta,
Anders Jonsson, and Sergio Barrachina-Muñoz

Abstract

Next-generation wireless deployments are characterized by being dense and uncoordinated, which often leads to inefficient use of resources and poor performance. To solve this, we envision the utilization of completely decentralized mechanisms to enable Spatial Reuse (SR). In particular, we focus on dynamic channel selection and Transmission Power Control (TPC). We rely on Reinforcement Learning (RL), and more specifically on Multi-Armed Bandits (MABs), to allow networks to learn their best configuration. In this work, we study the exploration-exploitation trade-off by means of the ε -greedy, EXP3, UCB and Thompson sampling action-selection, and compare their performance. In addition, we study the implications of selecting actions simultaneously in an adversarial setting (i.e., concurrently), and compare it with a sequential approach. Our results show that optimal proportional fairness can be achieved, even when no information about neighboring networks is available to the learners and Wireless Networks (WNs) operate selfishly. However, there is high temporal variability in the throughput experienced by the individual networks, especially for ε -greedy and EXP3. These strategies, contrary to UCB and Thompson sampling, base their operation on the absolute experienced reward, rather than on its distribution. We identify the cause of this variability to be the adversarial setting of our setup in which the set of most played actions provide intermittent good/poor performance depending on the neighboring decisions. We also show that learning sequentially, even if using a selfish strategy, contributes to minimize this variability. The sequential approach is therefore shown to effectively deal with the challenges posed by the adversarial settings that are typically found in decentralized WNs.

1 Introduction

Due to the growing popularity of wireless deployments, especially the ones based on the IEEE 802.11 standard (i.e., Wi-Fi), it is very common to find independent overlapping Wireless Networks (WNs) sharing the same channel resources. The decentralized nature of such kind of deployments leads to a significant lack of organization and/or agreement on sharing policies. As a result, resources are typically used inefficiently. An illustrative example of this can be found in [1], where the authors show that the power level used by wireless devices is typically set, by default, to the maximum, regardless of the distance between communicating nodes, and the channel occupancy. Consequently, increasing the capacity of such networks has become very challenging.

Wireless networks operate in three main domains: time, frequency and space. While the first two have been largely exploited, the spatial domain still shows plenty of room for improvement. According to [2], Spatial Reuse (SR) can be addressed by means of Transmission Power Control (TPC), Carrier Sense Threshold (CST) adjustment, rate adaptation (related to power control), and directional transmissions. In addition, interference cancellation can play a key role on spectral efficiency optimization [3]. On one side, TPC and CST adjustment aim at increasing spectral efficiency omnidirectionally. On the other hand, beamforming is meant for directional transmissions. Both beamforming and interference cancellation can be categorized as multiple antenna strategies. While the former allows to reduce the interference levels, the second one is useful to perform multiple simultaneous transmissions.

In this work, we focus on Dynamic Channel Allocation (DCA) and TPC to address the decentralized SR problem. A proper frequency planning allows to reduce the interference between wireless devices, and tuning the transmit power adds an extra level of SR that can result in improved throughput and fairness. The application of TPC and DCA is particularly challenging by itself. The interactions among devices depend on many features (such as position, environment or transmit power) and are hard to derive. Including beamforming and/or interference cancellation techniques [4], on the other hand, requires first a clear understanding of TCP and DCA performance alone, and is therefore left as future work.

Motivated by these challenges, we focus attention on Reinforcement Learning (RL), which has recently emerged as a very popular method to solve many well-known problems in wireless communications. RL allows to reduce the complexity generated in wireless environments by finding practical solutions. By applying RL, optimal (or near-to-optimal) solutions can be obtained without having a full understanding on the problem in advance. So, one of the main goals of this paper is to show its feasibility for the decentralized SR problem. Some RL-based applications can be found for packet routing [5], Access Point (AP) selection [6, 7], optimal rate sampling [8], or energy harvesting in heterogeneous networks [9]. All these applications make use of online learning, where a learner (or agent) obtains data periodically and uses it to predict future good-performing actions. Online learning is particularly useful to cope with complex and dynamic environments. This background encourages us to approach a solution for the decentralized SR problem in WNs through online learning techniques.

From the family of online algorithms, we are interested on analyzing the performance of Multi-Armed Bandits (MABs) [10] when applied to WNs. The MAB model is well-known in the online learning literature for solving resource allocation problems. In MABs, a given agent seeks to learn a hidden reward distribution while maximizing the gains. This is known as the exploration-exploitation trade-off. Exploitation is meant to maximize the long-term reward given the current estimate, and exploration aims to improve the estimate. Unlike classical RL, MABs do not consider states¹ in general, which can be hard to define for the decentralized SR problem presented in this work. On the one hand, spatial interference cannot be binary treated, thus leading to complex interactions among nodes. On the other hand, the adversarial setting unleashed by decentralized deployments increases the system complexity. Therefore, the obtained reward does not only depends on the actions taken by a given node, but also on the adversaries behavior.

This article extends our previous results presented in [11]. Here we generalize the contributions done by implementing several action-selection strategies to find the best combination of frequency channel and transmit power in WNs. These strategies are applied to the decentralized SR problem, where independent WNs learn selfishly, based on their own experienced performance. On the one hand, we evaluate the impact of varying parameters intrinsic to the proposed algorithms on the resulting throughput and fairness. In addition, we analyze the effects of learning selfishly, and shed light on the future of decentralized approaches. Notably, we observe that even though players act selfishly, some of the algorithms learn to play actions that enhance the overall performance, some times at the cost of high temporal variability. Considering selfish WNs and still obtaining collaborative behaviors is appealing to typical chaotic and dynamic deployments. Finally, the adversarial setting in WNs is studied under two learning implementations: namely *concurrent* and *sequential*. Both procedures rule the operation followed by learners (based on the proposed action-selection strategies). In particular, WNs select an action at the same time for the concurrent approach. In contrast, an ordered action-selection procedure is followed for the sequential case. We study the performance of the aforementioned techniques in terms of convergence speed, average throughput and variability. The main contributions of this work are summarized below:

- We devise the feasibility of applying MAB algorithms as defined in the online learning literature to solve the resource allocation problem in WNs.
- We study the impact of different parameters intrinsic to the action-selection strategies considered (e.g., exploration coefficients, learning rates) on network performance. In addition, we analyze the

¹A state refers to a particular situation experienced by a given agent, which is defined by a set of conditions. By having an accurate knowledge of its current situation, an agent can define state-specific strategies that maximize its profits.

implications derived from the application of different learning procedures, referred to as *concurrent* and *sequential*, which rule the moment at which WNs act.

- We show the impact of learning concurrently and sequentially. In particular, the former leads to a high throughput variability experienced by WNs, which is significantly reduced by the sequential approach. Accordingly, we envision the utilization of sequential approaches to achieve decentralized learning in adversarial wireless networks.
- Finally, we show that there are algorithms that learn to play collaborative actions even though the WNs act selfishly, which is appealing to practical application in chaotic and dynamic environments. In addition, we shed light on the root causes of this phenomena.

The remaining of this document is structured as follows: Section 2 outlines relevant related work. Section 3 introduces the proposed learning algorithms and their practical implementation for the resource allocation problem in WNs. Then, Section 4 presents the simulation scenarios and the considerations taken into account. The simulation results are later presented in Section 5. Finally, Section 6 provides the final remarks.

2 Related Work

Decentralized SR has been considerably studied by the wireless research community. The authors in [12] propose using relay nodes to re-transmit packets lost as a result of a collision. The relay node is able to decode different signals from the environment and to detect if a collision took place. Then, with the aim of improving the re-transmissions operation, it benefits from the current transmission to forward the decoded packets to their original destinations. Although this method shows performance improvements in dense scenarios where collisions are very likely to occur, its effectiveness is subject to the network topology. Regarding directional transmissions, the authors in [13] propose two novel access schemes to allow multiple simultaneous transmissions. In particular, nodes' activity information is sensed, which, together with antenna's directionality information, allows to build a new set of channel access rules.

Despite approaches based on directional transmissions and interference cancellation are very powerful and allow to significantly increase SR, they strongly rely on having multiple antennas. Such a requirement is not mandatory for the SR operation based on TPC and CST adjustment. In this work, we focus on the former because tuning the transmit power has a direct impact on the generated interference. This allows to purely study the interactions that occur among nodes implementing decentralized SR. Moreover, we consider DCA to be combined with TPC, so that further potential gains can be achieved.

DCA has been extensively studied from the centralized perspective, especially through techniques based on graph coloring [14, 15]. Despite these kind of approaches allow to effectively reduce the interference between WNs, a certain degree of communication is required. Regarding decentralized methods, the authors in [16] propose a very simple approach in which each AP maintains an interference map of their neighbors, so that channel assignment is done through interference minimization. Unfortunately, the interactions among APs in the decentralized setting are not studied. Separately, [17] proposes two decentralized approaches that rely on the interference measured at both APs and stations (STAs) to calculate the best frequency channels for dynamic channel allocation. To do so, a WN, in addition to the interference sensed by its associated devices, considers other metrics such as the amount of traffic, so that some coordination is required at the neighbor level (e.g., periodic reporting). The authors in [18] show that the decentralized DCA problem is NP-hard. In addition, they propose a distributed algorithm whereby APs select the best channel according to the observed traffic information (i.e., channel sensing is considered).

In this work we aim to extend the approach in [18] in two ways. First, we aim to provide a flexible solution based on the performance achieved by a given WN. Second, we aim to tackle the spatial domain through TPC, which has been shown to provide large improvements in wireless networks [19]. However, dealing with the spatial dimension leads to unpredictable interactions in terms of interference. Such a complexity is illustrated in [20], which performs power control and rate adaptation in subgroups of Wireless Local Area Networks (WLANs). The creation of clusters allows defining independent power levels between devices in the same

group, which are useful to avoid asymmetric links. However, to represent all the possible combinations, graphs can become very large, especially in high-density deployments. When it comes to decentralized mechanisms, we find the work in [21], which applies TPC based on real-time channel measurements [21]. The proposed mechanism (so called Dynamic Transmission Power Control) is based on a set of triggered thresholds that increase/decrease the transmit power according to the state of the system. The main problem is that thresholds are set empirically (based on simulations), which limits the potential of the mechanisms in front of multiple scenarios.

As shown by previous research, the optimal decentralized SR in WNs through TPC and DCA is very hard to be derived analytically, mostly because of the adversarial setting and the lack of information at nodes. The existing decentralized solutions barely provide flexibility with respect to the scenario, so that potential use cases are disregarded. For that, we focus on online learning, and more precisely Multi-Armed Bandits (MABs). The MAB framework allows to reduce the complexity of the SR problem, since detailed information about the scenario is not considered. In contrast, learners gain knowledge on all the adversaries as a whole, thus facing a single environment. To the best of our knowledge, there is very little related work on applying MAB techniques to the problem of resource allocation in WNs. In [22], the authors propose modeling a resource allocation problem in Long Term Evolution (LTE) networks through MABs. In particular, a set of Base Stations (BS) learn the best configuration of Resource Blocks (RBs) in a decentralized way. For that purpose, a variation of EXP3 (so-called Q-EXP3) is proposed, which is shown to reduce the strategy set. Despite a regret bound is provided, it is subject to the fact that an optimal resource allocation exists, i.e., every BS obtains the necessary resources. In addition, a large number of iterations is required to find the optimal solution in a relatively small scenario, thus revealing the difficulties shown by decentralized settings.

More related to the problem proposed here, the authors in [23] show a channel selection and power control approach in infrastructureless networks, which is modeled through bandits. In particular, two different strategies are provided to improve the performance of two Device to Device (D2D) users (each one composed by a transmitter and a receiver), which must learn the best channel and transmit power to be selected. Similarly to our problem, users do not have any knowledge on the channel or the other's configuration, so they rely on the experienced performance in order to find the best configuration. An extension of [23] is provided by the same authors in [24], which includes a calibrated predictor (referred in the work as *forecaster*) to infer the behavior of the other devices in order to counteract their actions. In each agent, the information of the forecaster is used to choose the highest-rewarding action with a certain probability, while the rest of actions are randomly selected. Henceforth, assuming that all the networks use a given strategy \mathcal{X} , fast convergence is ensured. Results show that channel resources are optimally distributed in a very short time frame through a fully decentralized algorithm that does not require any kind of coordination. Both aforementioned works rely on the existence of a unique Nash Equilibrium, which favors convergence. In contrast, in this article we aim to extend Bandits utilization to denser deployments, and, what is more important, to scenarios with limited available resources in which there is not a unique Nash Equilibrium (NE) that allows fast-convergence. Thus, we aim to capture the effects of applying selfish strategies in a decentralized way (i.e., agent i follows a strategy \mathcal{X}_i that does not consider the strategies of the others) and we also provide insight about the importance of past information for learning in dense WNs, which has not been studied before.

3 Multi-Armed Bandits for Improving Spatial Reuse in WNs

In this work, we address the decentralized SR problem through online learning because of the uncertainty generated in an adversarial setting. The practical application of MABs in WNs is detailed next:

3.1 The Multi-Armed Bandits Framework

In the online learning literature, several MAB settings have been considered such as stochastic bandits [25, 26, 27], adversarial bandits [28, 29], restless bandits [30], contextual bandits [31] and linear bandits [32, 33], and numerous exploration-exploitation strategies have been proposed such as ε -greedy [34, 27],

upper confidence bound (UCB) [26, 35, 36, 27], *exponential weight algorithm for exploration and exploitation* (EXP3) [28, 27] and *Thompson sampling* [25]. The classical multi-armed bandit problem models a sequential interaction scheme between a learner and an environment. The learner sequentially selects one out of K actions (often called *arms* in this context) and earns some rewards determined by the chosen action and also influenced by the environment. Formally, the problem is defined as a repeated game where the following steps are repeated in each round $t = 1, 2, \dots, T$:

1. The environment fixes an assignment of rewards $r_{a,t}$ for each action $a \in [K] \stackrel{\text{def}}{=} \{1, 2, \dots, K\}$,
2. the learner chooses action $a_t \in [K]$,
3. the learner obtains and observes reward $r_{a_t,t}$.

The bandit literature largely focuses on the perspective of the learner with the objective of coming up with learning algorithms that attempt to maximize the sum of the rewards gathered during the whole procedure (either with finite or infinite horizon). As noted above, this problem has been studied under various assumptions made on the environment and the structure of the arms. The most important basic cases are the *stochastic* bandit problem where, for each particular arm a , the rewards are i.i.d. realizations of random variables from a fixed (but unknown) distribution ν_a , and the *non-stochastic* (or *adversarial*) bandit problem where the rewards are chosen arbitrarily by the environment. In both cases, the main challenge for the learner is the *partial observability* of the rewards: the learner only gets to observe the reward associated with the chosen action a_t , but never observes the rewards realized for the other actions.

Let $r_{a^*,t}$ and $r_{a,t}$ be the rewards obtained at time t from choosing actions a^* (optimal) and a , respectively. Then, the performance of learning algorithms is typically measured by the *total expected regret* defined as

$$R_T = \sum_{t=0}^T \mathbb{E}[(r_{a^*,t} - r_{a,t})].$$

An algorithm is said to *learn* if it guarantees that the regret grows sublinearly in T , that is, if $R_T = o(T)$ is guaranteed as T grows large, or, equivalently, that the average regret R_T/T converges to zero. Intuitively, sublinear regret means that the learner eventually identifies the action with the highest long-term payoff. Note, as well, that the optimal action a^* is the same across all the rounds. Most bandit algorithms come with some sort of a guaranteed upper bound on R_T which allows for a principled comparison between various methods.

3.2 Multi-Armed Bandits Formulation for Decentralized Spatial Reuse

We model the decentralized SR problem through adversarial bandits. In such a model, the reward experienced by a given agent (WN) is influenced by the whole action profile, i.e., the configurations used by other competing WNs. From a decentralized perspective, the adversarial setting poses several challenges with respect to the existence of a NE. Ideally, the problem is solved if all the competitors implement a pure strategy² that allows maximizing a certain performance metric. However, finding such a strategy may not be possible in unplanned deployments, due to the competition among nodes and the scarcity of the available resources. Understanding the implications derived from such an adversarial setting in the absence of a NE is one the main goals of this paper, which, to the best of our knowledge, has been barely considered in the previous literature.

In particular, we model this adversarial problem as follows. Let arm $a \in \mathcal{A}$ (we denote the size of \mathcal{A} with K) be a configuration in terms of channel and transmit power (e.g., $a_1 = \{\text{Channel: 1, TPC: -15 dBm}\}$). Let $\Gamma_{i,t}$ be the throughput experienced by WN_i at time t , and Γ_i^* the optimal throughput.³ We then define the reward $r_{i,t}$ experienced by WN_i at time t as:

²A pure strategy NE is conformed by a set of strategies and payoffs, so that no player can obtain further benefits by deviating from its strategy.

³The optimal throughput is achieved in case of isolation (i.e., when no interference is experienced in the selected channel).

$$r_{i,t} = \frac{\Gamma_{i,t}}{\Gamma_i^*} \leq 1,$$

In order to attempt to maximize the reward, we have considered the ε -greedy, EXP3, UCB and Thompson sampling action-selection strategies, which are described next in this section. While ε -greedy and EXP3 explicitly include the concepts of *exploration coefficient* and *learning rate*, respectively, UCB and Thompson sampling are parameter-free policies that extend the concept of exploration (actions are explored according to their estimated value and not by commitment). The aforementioned policies are widely spread and considered of remarkable importance in the MAB literature.

3.2.1 ε -greedy

The ε -greedy policy [34, 27] is arguably the simplest learning algorithm attempting to deal with exploration-exploitation trade-offs. In each round t , the ε -greedy algorithm explicitly decides whether to explore or exploit: with probability ε , the algorithm picks an arm uniformly at random (exploration), and otherwise it plays the arm with the highest empirical return $\hat{r}_{k,t}$ (exploitation).

In case ε is fixed for the entire process, the expected regret is obviously going to grow linearly as $\Omega(\varepsilon T)$ in general. Therefore, in order to obtain a sublinear regret guarantee (and thus an asymptotically optimal growth rate for the total rewards), it is critical to properly adjust the exploration coefficient. Thus, in our ε -greedy implementation, we use a time-dependent exploration rate of $\varepsilon_t = \varepsilon_0/\sqrt{t}$, as suggested in the literature [27]. The adaptation of this policy to our setting is shown as Algorithm 1.

Algorithm 1: Implementation of Multi-Armed Bandits (ε -greedy) in a WN. $\mathcal{U}(1, K)$ is a uniform distribution that randomly chooses from 1 to K .

Input: SNR: information about the Signal-to-Noise Ratio received at the STA, \mathcal{A} : set of possible actions in $\{a_1, \dots, a_K\}$

1 **Initialize:** $t = 0$, $\varepsilon_t = \varepsilon_0$, $r_k = 0$, $\forall a_k \in \mathcal{A}$

2 **while** active **do**

3 Select a_k $\begin{cases} \text{argmax } r_{k,t}, & \text{with prob. } 1 - \varepsilon \\ k \sim \mathcal{U}(1, K), & \text{otherwise} \end{cases}$

4 Observe the throughput experienced Γ_t

5 Compute the reward $r_{k,t} = \frac{\Gamma_t}{\Gamma^*}$, where $\Gamma^* = B \log_2(1 + \text{SNR})$

6 $\varepsilon_t \leftarrow \varepsilon_0/\sqrt{t}$

7 $t \leftarrow t + 1$

8 **end**

3.2.2 EXP3

The EXP3 algorithm [28, 29] is an adaptation of the weighted majority algorithm of [37, 38] to the non-stochastic bandit problem. EXP3 maintains a set of non-negative weights assigned to each arm and picks the actions randomly with a probability proportional to their respective weights (initialized to 1 for all arms). The aim of EXP3 is to provide higher weights to the best actions as the learning procedure proceeds.

More formally, letting $w_{k,t}$ be the weight of arm k at time $t \in \{1, 2, \dots\}$, EXP3 computes the probability $p_{k,t}$ of choosing arm k in round t as

$$p_{k,t} = (1 - \gamma) \frac{w_{k,t}}{\sum_{i=1}^K w_{i,t}} + \frac{\gamma}{K},$$

where $\gamma \in [0, 1]$ is a parameter controlling the rate of exploration. Having selected arm a_t , the learner observes the generated pay-off $r_{a_t,t}$ and computes the importance-weighted reward estimates for all $k \in [K]$

$$\hat{r}_{k,t} = \frac{\mathbb{I}_{\{I_t=k\}} r_{k,t}}{p_{k,t}},$$

where $\mathbb{I}_{\{A\}}$ denoting the indicator function of the event A taking a value of 1 if A is true and 0 otherwise. Finally, the weight of arm k is updated as a function of the estimated reward:

$$w_{k,t+1} = w_{k,t} e^{\frac{\eta \cdot \hat{r}_{k,t}}{K}},$$

where $\eta > 0$ is a parameter of the algorithm often called the *learning rate*. Intuitively, η regulates the rate in which the algorithm incorporates new observations. Large values of η correspond to more confident updates and small values lead to more conservative behaviors. As we did for the exploration coefficient in ε -greedy, we use a time-dependent learning rate of $\eta_t = \eta_0/\sqrt{t}$ [27]. Our implementation of EXP3 is detailed in Algorithm 2.

Algorithm 2: Implementation of Multi-Armed Bandits (EXP3) in a WN

Input: SNR: information about the Signal-to-Noise Ratio received at the STA, \mathcal{A} : set of possible actions in $\{a_1, \dots, a_K\}$

1 **Initialize:** $t = 0$, $\eta_t = \eta_0$, $w_{k,t} = 1, \forall k \in \mathcal{A}$

2 **while** active **do**

3 $p_{k,t} \leftarrow (1 - \gamma) \frac{w_{k,t}}{\sum_{i=1}^K w_{i,t}} + \frac{\gamma}{K}$

4 Draw $a_k \sim p_{k,t} = (p_{1,t}, p_{2,t}, \dots, p_{K,t})$

5 Observe the throughput experienced Γ_t

6 Compute the reward $r_{k,t} = \frac{\Gamma_t}{\Gamma^*}$, where $\Gamma^* = B \log_2(1 + \text{SNR})$

7 $\hat{r}_{k,t} \leftarrow \frac{r_{k,t}}{p_{k,t}}$

8 $w_{k,t} \leftarrow w_{k,t-1}^{\frac{\eta_t}{\eta_{t-1}}} \cdot e^{\eta_t \cdot \hat{r}_{k,t}}$

9 $w_{k',t} \leftarrow w_{k',t-1}^{\eta_t / \eta_{t-1}}, \forall k' \neq k$

10 $\eta_t \leftarrow \frac{\eta_0}{\sqrt{t}}$

11 $t \leftarrow t + 1$

12 **end**

3.2.3 UCB

The *upper confidence bound* (UCB) action-selection strategy [35, 36, 27] is based on the principle of *optimism in face of uncertainty*: in each round, UCB selects the arm with the highest statistically feasible mean reward given the past observations. Statistical feasibility here is represented by an upper confidence bound on the mean rewards which shrinks around the empirical rewards as the number of observations increases. Intuitively, UCB trades off exploration and exploitation very effectively, as upon every time a suboptimal arm is chosen, the corresponding confidence bound will shrink significantly, thus quickly decreasing the probability of drawing this arm in the future. The width of the confidence intervals is chosen carefully so that the true best arm never gets discarded accidentally by the algorithm, yet suboptimal arms are drawn as few times as possible. To obtain the first estimates, each arm is played once at the initialization.

Formally, let n_k be the number of times that arm k has been played, and $\Gamma_{k,t}$ the throughput obtained by playing arm k at time t . The average reward $\bar{r}_{k,t}$ of arm k at time t is therefore given by:

$$\bar{r}_{k,t} = \frac{1}{n_k} \sum_{s=1}^{n_k} r_{k,s}$$

Based on these average rewards, UCB selects the action that maximizes $\bar{r}_{k,t} + \sqrt{\frac{2 \ln(t)}{n_k}}$. By doing so, UCB implicitly balances exploration and exploitation, as it focuses efforts on the arms that are *i*) the most promising (with large estimated rewards) or *ii*) not explored enough (with small n_k). Our implementation of UCB is detailed in Algorithm 3.

Algorithm 3: Implementation of Multi-Armed Bandits (UCB) in a WN

Input: SNR: information about the Signal-to-Noise Ratio received at the STA, \mathcal{A} : set of possible actions in $\{a_1, \dots, a_K\}$

1 **Initialize:** $t = 0$, play each arm $a_k \in \mathcal{A}$ once

2 **while** active **do**

3 Draw $a_k = \operatorname{argmax}_{k=1,\dots,K} \bar{r}_k + \sqrt{\frac{2\ln(t)}{n_k}}$

4 Observe the throughput experienced Γ_t

5 Compute the reward $r_{k,t} = \frac{\Gamma_t}{\Gamma^*}$, where $\Gamma^* = B \log_2(1 + \text{SNR})$

6 $n_k \leftarrow n_k + 1$

7 $\bar{r}_k \leftarrow \frac{1}{n_k} \sum_{s=1}^{n_k} r_{k,s}$

8 $t \leftarrow t + 1$

9 **end**

3.2.4 Thompson sampling

Thompson sampling [25] is a well-studied action-selection technique that had been known for its excellent empirical performance [39] and was recently proven to achieve strong performance guarantees, often better than those warranted by UCB [40, 41, 42]. Thompson sampling is a Bayesian algorithm: it constructs a probabilistic model of the rewards and assumes a prior distribution of the parameters of said model. Given the data collected during the learning procedure, this policy keeps track of the posterior distribution of the rewards, and pulls arms randomly in a way that the drawing probability of each arm matches the probability of the particular arm being optimal. In practice, this is implemented by sampling the parameter corresponding to each arm from the posterior distribution, and pulling the arm yielding the maximal expected reward under the sampled parameter value.

For the sake of practicality, we assume that rewards follow a Gaussian distribution with a standard Gaussian prior, as suggested in [43]. By standard calculations, it can be verified that the posterior distribution of the rewards under this model is Gaussian with mean and variance

$$\hat{r}_k(t) = \frac{\sum_{w=1:k}^{t-1} r_k(w)}{n_k(t) + 1} / \sigma_k^2(t) = \frac{1}{n_k + 1},$$

where n_k is the number of times that arm k was drawn until the beginning of round t . Thus, implementing Thompson sampling in this model amounts to sampling a parameter θ_k from the Gaussian distribution $\mathcal{N}(\hat{r}_k(t), \sigma_k^2(t))$ and choosing the action with the maximal parameter. Our implementation of Thompson sampling to the WN problem is detailed in Algorithm 4.

4 System model

For the remainder of this work, we study the interactions among several WNs placed in a 3-D scenario that occur when applying MABs in a decentralized manner (with parameters described later in Section 4.4). For simplicity, we consider WNs to be composed by an AP transmitting to a single Station (STA) in a downlink manner. Note that in typical uncoordinated wireless deployments (e.g., residential buildings), STAs are typically close to the AP to which they are associated. Thus, having several STAs associated to the same AP does not significantly impact the inter-WNs interference studied in this work.

Algorithm 4: Implementation of Multi-Armed Bandits (Thompson s.) in a WN

Input: SNR: information about the Signal-to-Noise Ratio received at the STA, \mathcal{A} : set of possible actions in $\{a_1, \dots, a_K\}$

- 1 **Initialize:** $t = 0$, for each arm $a_k \in \mathcal{A}$, set $\hat{r}_k = 0$ and $n_k = 0$
- 2 **while** active **do**
- 3 For each arm $a_k \in \mathcal{A}$, sample $\theta_k(t)$ from normal distribution $\mathcal{N}(\hat{r}_k, \frac{1}{n_k+1})$
- 4 Play arm $a_k = \operatorname{argmax}_{k=1,\dots,K} \theta_k(t)$
- 5 Observe the throughput experienced Γ_t
- 6 Compute the reward $r_{k,t} = \frac{\Gamma_t}{\Gamma^*}$, where $\Gamma^* = B \log_2(1 + \text{SNR})$
- 7 $\hat{r}_{k,t} \leftarrow \frac{\hat{r}_{k,t} n_{k,t} + r_{k,t}}{n_{k,t} + 2}$
- 8 $n_{k,t} \leftarrow n_{k,t} + 1$
- 9 $t \leftarrow t + 1$
- 10 **end**

4.1 Channel modeling

Path-loss and shadowing effects are modeled using the log-distance model for indoor communications. The path-loss between WN i and WN j is given by:

$$\text{PL}_{i,j} = P_{\text{tx},i} - P_{\text{rx},j} = \text{PL}_0 + 10\alpha \log_{10}(d_{i,j}) + G_s + \frac{d_{i,j}}{d_{\text{obs}}} G_o,$$

where $P_{\text{tx},i}$ is the transmitted power in dBm by the AP in WN $_i$, α is the path-loss exponent, $P_{\text{rx},j}$ is the power in dBm received at the STA in WN $_j$, PL_0 is the path-loss at one meter in dB, $d_{i,j}$ is the distance between the transmitter and the receiver in meters, G_s is the log-normal shadowing loss in dB, and G_o is the obstacles loss in dB. Note that we include the factor d_{obs} , which is the average distance between two obstacles in meters.

4.2 Throughput calculation

The throughput experienced by WN i at time t is given by $\Gamma_{i,t} = B \log_2(1 + \text{SINR}_{i,t})$, where B is the channel width and SINR is the experienced Signal to Interference plus Noise Ratio. The latter is computed as $\text{SINR}_{i,t} = \frac{P_{i,t}}{I_{i,t} + N}$, where $P_{i,t}$ and $I_{i,t}$ are the received power and the sum of the interference at WN i at time t , respectively, and N is the floor noise power. Adjacent channel interference is also considered in $I_{i,t}$, so that the transmitted power leaked to adjacent channels is 20 dBm lower for each extra channel separation. Similarly, the optimal throughput is computed as $\Gamma_i^* = B \log_2(1 + \text{SNR}_i)$, which frames the operation of a given WN in isolation.

4.3 Learning procedure

We frame the decentralized learning procedure in two different ways, namely *concurrent* and *sequential*. Figure 1 illustrates the procedure followed by agents to carry out decentralized SR learning. As shown, in each iteration⁴ there is a monitoring phase (shown in grey), where the current selected action is analyzed by each agent to quantify the hidden reward (which depends on the adversarial setting). Such a reward is the same for all the policies presented in this work, so that a fair comparison can be provided. After the monitoring phase is completed, agents update their knowledge (shown in purple) and choose a new action

⁴The time between iterations (T) must be large enough to provide an accurate estimation of the throughput experienced for a given action profile.

(shown in yellow). Note, as well, that both approaches rely on a synchronization phase (shown in green), which can be achieved through message passing [44, 45] and/or environment sensing.⁵

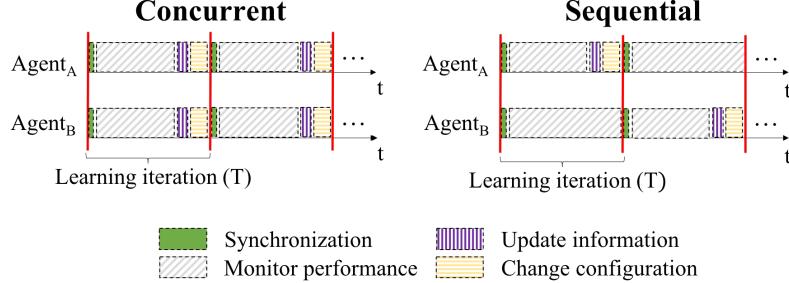


Figure 1: Concurrent and sequential procedures.

In the concurrent approach, agents (or WNs) make decisions simultaneously, thus leading to a more variable (and chaotic) environment. In practice the fully decentralized learning process will most probably not be synchronized but we leave the study of any possible effects of that desynchronization to future work. In contrast, for the sequential approach, WNs need to wait for their turn in order to pick a new action. As a result, the performance of their last selected action (arm) is measured for several iterations (equal to the number of overlapping networks). In particular, during the update phase, a WN computes the reward of its last selected arm according to the throughput experienced in average. Accordingly, the performance of a given action is measured against different adversarial settings, since the environment changes gradually. Despite agents still learn selfishly, they can better assess how robust an action is against the joint actions profile, in comparison to the concurrent approach.

4.4 Simulation Parameters

According to [46], which provides an overview of the IEEE 802.11ax-2019 standard, a typical high-density scenario for residential buildings contains 0.0033 APs/m^3 (i.e., 100 APs in a $100 \times 20 \times 15 \text{ m}$ area). Accordingly, for simulation purposes, we define a map scenario with dimensions $10 \times 5 \times 10 \text{ m}$, containing from 2 to 8 APs. In addition, for the first part of the simulations, we consider a setting containing 4 WNs that form a grid topology. In it, STAs are placed at the maximum possible distance from the other networks. Table 1 details the parameters used.

5 Performance Evaluation

In this Section, we evaluate the performance of each action-selection strategy presented in Section 3 when applied to the decentralized SR problem in WNs.⁶ For that purpose, we first evaluate in Section 5.1 the ϵ -greedy, EXP3, UCB and Thompson sampling policies in a fixed adversarial environment. This allows us to provide insights on the decentralized learning problem in a competitive scenario. Accordingly, we are able to analyze in detail the effect of applying each learning policy on the network's performance. Without loss of generality, we consider a symmetric configuration and analyze the competition effects when WNs have the same opportunities for accessing the channel. Finally, Section 5.2 provides a performance comparison of the aforementioned scenarios with different densities and with randomly located WNs.

⁵The IEEE 802.11k amendment, which is devoted to measurement reporting, may enable the environment sensing operation.

⁶The source code used in this work is open [47], encouraging sharing of knowledge with potential contributors under the GNU General Public License v3.0.

Parameter	Value
Map size (m)	$10 \times 5 \times 10$
Number of coexistent WNs	{2, 4, 6, 8}
APs/STAs per WN	1 / 1
Distance AP-STA (m)	$\sqrt{2}$
Number of orthogonal channels	3
Channel bandwidth (MHz)	20
Initial channel selection model	Uniformly distributed
Transmit power values (dBm)	{-15, 0, 15, 30}
PL_0 (dB)	5
G_s (dB)	Normally distributed with mean 9.5
G_o (dB)	Uniformly distributed with mean 30
d_{obs} (meters between two obstacles)	5
Noise level (dBm)	-100
Traffic model	Full buffer (downlink)
Number of learning iterations	10,000

Table 1: Simulation parameters

5.1 Toy Grid Scenario

The toy grid scenario contains 4 WNs and is illustrated in Figure 2. This scenario has the particularity of being symmetric, so that adversarial WNs have the same opportunities to compete for the channel resources. The optimal solution in terms of proportional fairness⁷ is achieved when channel reuse is maximized and WNs sharing the channel moderate their transmit power. The PF solution provides an aggregate performance of 440.83 Mbps (i.e., 106.212 Mbps per WN on average). The optimal solution is computed by brute force (i.e., trying all the combinations), and it is used as a baseline.

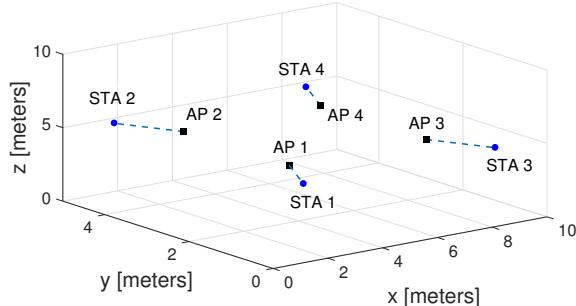


Figure 2: Grid scenario containing 4 WNs, each one composed by an AP and a STA.

5.1.1 Configuration of the Learning Parameters

Before comparing the performance of each algorithm, we first analyze the effect of modifying each one's internal parameters. Since the versions of UCB and Thompson sampling analyzed in this work are parameterless, in this section we focus only on ε -greedy and EXP3 methods.

Firstly, ε -greedy allows to regulate the explicit exploration rate at which the agent operates, which is referred to as ε . In this paper, ε is dynamically adjusted as $\varepsilon_t = \frac{\varepsilon_0}{\sqrt{t}}$, with the aim of exploring more efficiently. Accordingly, we study the impact of modifying the initial exploration coefficient in the experienced

⁷The proportional fairness (PF) result accomplishes that the logarithmic sum of each individual throughput is maximized: $\max \sum_{i \in \text{WN}} \log(\Gamma_i)$.

performance by a WN. Secondly, when it comes to EXP3, we find two parameters, namely η and γ . While η controls how fast old beliefs are replaced by newer ones, γ regulates explicit exploration by tuning the importance of weights in the action-selection procedure. Setting $\gamma = 1$ results in completely neglecting weights (actions have the same probability to be chosen). On the other side, by setting $\gamma = 0$, the effect of weights are at its highest importance. Thus, in order to clearly analyze the effects of the EXP3 weights, which directly depend on η , we fix γ to 0. As we did for ε -greedy, we analyze the impact of modifying the parameter η_0 in EXP3 on the WN's performance.

Figure 3 shows the aggregate throughput obtained in the grid scenario when applying both ε -greedy and EXP3 during 10,000 iterations, and for each ε_0 and η_0 values, respectively. The results are presented for values ε_0 and η_0 between 0 and 1 in 0.1 steps. The average and standard deviation of the throughput from 100 simulation runs are also shown, and compared with the proportional fair solution.

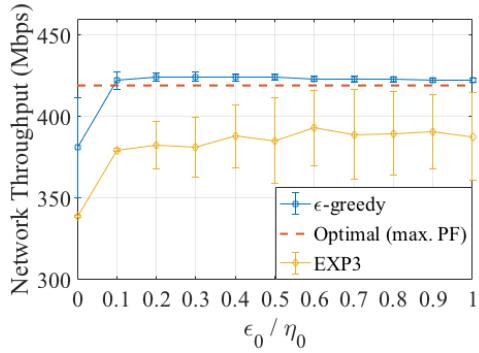


Figure 3: Average network throughput and standard deviation obtained for each ε_0 and η_0 value in ε -greedy and EXP3, respectively. Results are from 100 simulations lasting 10,000 iterations each. The proportional fair solution is also shown (red dashed line).

As shown, the aggregate throughput obtained on average is quite similar for all ε_0 and η_0 values, except for the complete random case where no exploration is done (i.e., when ε_0 and η_0 are equal to 0). For ε -greedy, the lower the ε_0 parameter, the less exploration is performed. Consequently, for low ε_0 , the average throughput is highly dependent on how good/bad were the actions taken at the beginning of the learning process, which results in a higher standard deviation as ε_0 goes to 0. As for EXP3, the lower η_0 , the more slowly weights are updated. For $\eta_0 = 0$, weights are never updated, so that arms have always the same probability to be chosen. To conclude, we choose $\varepsilon_0 = 1$ and $\eta_0 = 0.1$, respectively, for the rest of simulations, which provide the highest ratio between the aggregate throughput and the variability among different runs.

5.1.2 Performance of the MAB-based Policies

Once we established the initial parameters to be used by both ε -greedy and EXP3, we now compare the performance of all the studied action-selection strategies when applied to decentralized WNs. First, we focus on the average throughput achieved by each WN in the toy grid scenario, for each of the methods (Figure 4(a)). As shown, the proportional fair solution is almost achieved by all the learning methods. However, Thompson sampling is shown to be much more stable than the other mechanisms, since its variability in the aggregate throughput is much lower (depicted in Figure 4(b)).

In order to dig deeper into the behavior of agents for each policy, Figure 5 shows the probability of each WN to choose each action. Regarding ε -greedy, EXP3 and UCB, a large set of actions is chosen with similar probabilities. Note that there are only three frequency channels, so that two WNs need to share one of them, thus leading to a lower performance with respect to the other two. Therefore, WNs are constantly changing

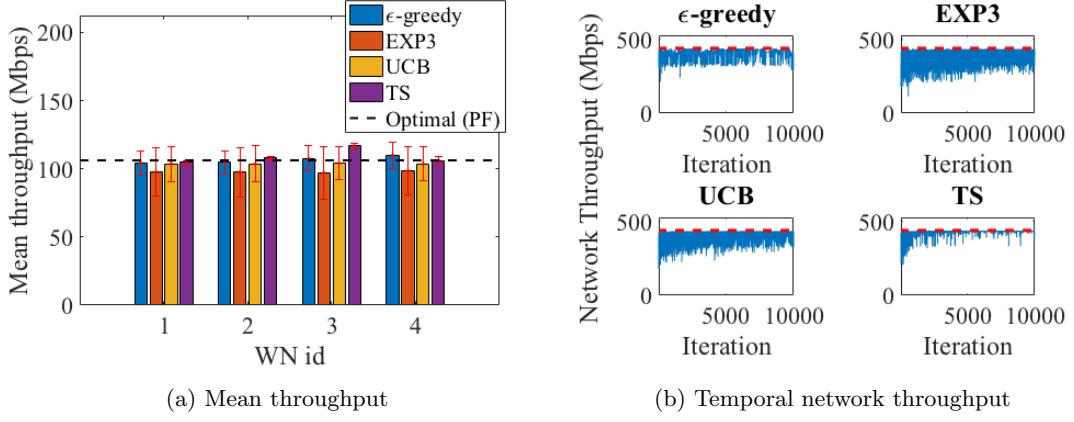


Figure 4: Mean throughput achieved per WN, for each action-selection strategy (the standard deviation is shown in red). The black dashed line indicates the PF result.

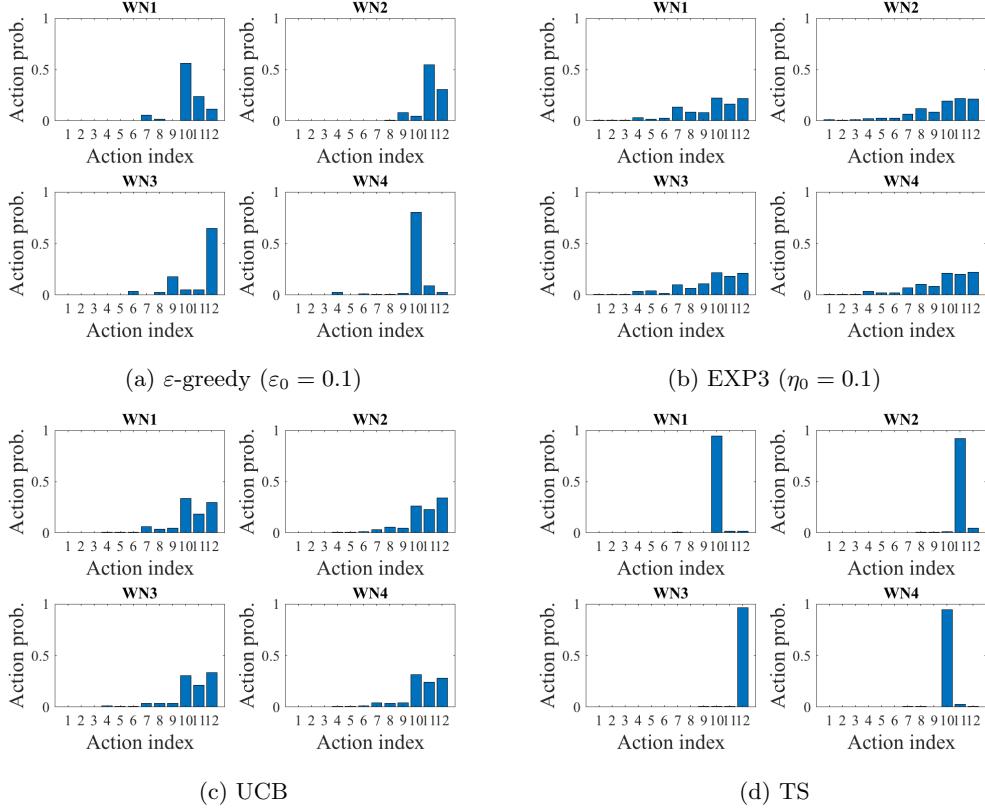


Figure 5: Probability of selecting each given action for a simulation of 10,000 iterations.

their channel and experiencing intermittent good/poor performance. Thus, the degree of exploration is kept very high, resulting in high temporal variability. In contrast, Thompson sampling shows a clearer preference

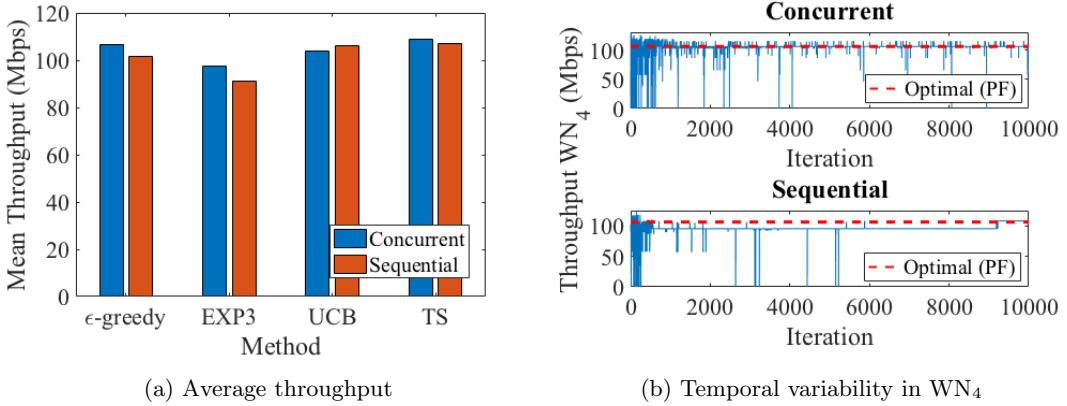


Figure 6: Concurrent vs sequential approaches performance. (a) Mean average throughput achieved for each learning procedure. (b) Temporal variability experienced by WN₄ for the Thompson sampling action-selection strategy and for each learning procedure.

for selecting a single action, which allows reducing the aforementioned variability.

5.1.3 Learning Sequentially

In order to alleviate the strong throughput variability experienced when applying decentralized learning, we now focus on the sequential approach introduced in Section 4.3. Now, only one WN is able to select an action at a time. With that, we aim to reduce the adversarial effect on the estimated rewards. Therefore, by having a more stable environment (not all the WNs learn simultaneously), the actual reward of a given selected action can be estimated more accurately. Figure 6 shows the differences between learning through concurrent and sequential mechanisms. Firstly, the throughput experienced on average along the entire simulation is depicted in Figure 6(a). Secondly, without loss of generality, Figure 6(b) shows the temporal variability experienced by WN₄ when applying Thompson sampling. Note that showing the performance of a single WN is representative enough for the entire set of WNs (the scenario is symmetric), and allows us to analyze in detail the behavior of the algorithms.

On the one hand, a lower throughput is experienced on average when learning in a sequential way, but the differences are very small. In such a situation, WNs spend more time observing sub-optimal actions, since they need to wait for their turn. Note, as well, that the time between iterations (T) depends on the implementation. In this particular case, we assume that T is the same for both sequential and concurrent approaches.

On the other hand, the temporal variability shown by the sequential approach is much lower than for the concurrent one (Figure 6(b)). The high temporal variability may negatively impact on the user's experience and the operation of upper layer protocols (e.g., TCP) may be severely affected. Notice that a similar effect is achieved for the rest of algorithms.

5.1.4 Learning in a Dynamic Environment

Finally, we show the performance of the proposed learning mechanisms in a dynamic scenario. For that, we propose the following situation. Firstly, WN₁ and WN₂ are active for the whole simulation. Secondly, WN₃ turns on at iteration 2,500, when WN₁ and WN₂ are supposed to have acquired enough knowledge to maximize SR. Finally, WN₄ turns on at iteration 5,000, similarly than for WN₃.

Through this simulation, we aim to show how each learning algorithm adapts to changes in the environment, which highly impact on the rewards distributions. Figure 7 shows the temporal aggregate throughput achieved by each action-selection strategy. As done in Subsection 5.1, we only plot the results of the best-performing algorithm, i.e., Thompson sampling, both for the concurrent and the sequential procedures.

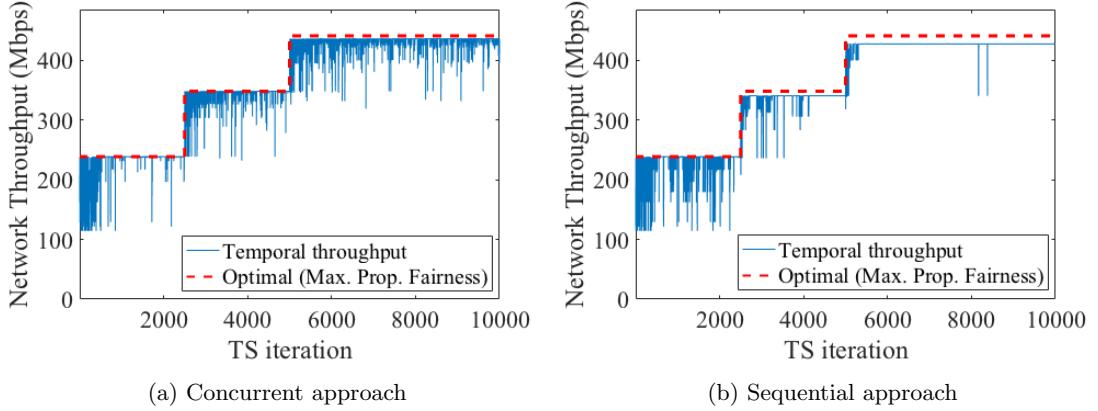


Figure 7: Temporal aggregate throughput experienced for a 10,000-iteration Thompson sampling simulation.

As shown, WNs are able to adapt to the changes in the environment. In particular, for the concurrent case (see Figure 7(a)), changes are harder to be captured as the network size increases. In contrast, learning in an ordered way (see Figure 7(b)) allows reducing the temporal variability, even if new WNs turn on. However, there is a little loss in the aggregate performance with respect to the concurrent approach. The difference between the maximum network performance is mostly provoked by the reduced exploration shown by the sequential approach.

5.2 Random Scenarios

We now evaluate whether the previous conclusions generalize to random scenarios with an arbitrary number of WNs. To this aim, we use the same $10 \times 5 \times 10$ m scenario and randomly allocate $N = \{2, 4, 6, 8\}$ WNs. Figures 8 and 9 show the mean throughput and variability experienced for each learning strategy, and for each number of coexistent WNs, respectively. The variability is measured as the standard deviation that a given WN experiences along an entire simulation. We consider the average results of 100 different random scenarios for each number of networks. In particular, we are interested on analyzing the gains achieved by each algorithm, even if convergence cannot be provided due to the competition between networks. For that, we display the average performance for the following learning intervals: [1-100, 101-500, 501-1000, 1001-2500, 2501-10000]. Note that the first intervals represent few iterations. This allows us to observe the performance achieved during the transitory phase in more detail. In addition, the performance achieved in a static situation (i.e., when no learning is performed) is shown in Figure 8. With that, we aim to compare the gains obtained by each learning strategy with respect to the current IEEE 802.11 operation in unplanned and chaotic deployments.

First of all, let us focus on the throughput improvements achieved with respect to the static situation. As shown in Figure 8, each learning strategy easily outperforms the static scenario for low densities (i.e., 2 and 4 overlapping WNs). However, as density increases, improving the average throughput becomes more challenging. This is clearly evidenced for $N = \{6, 8\}$ WNs, where EXP3 performs worse than the static situation.

Secondly, we concentrate on the concurrent learning procedure. As shown in Figure 8, Thompson sampling outperforms the other action-selection strategies for all the scenarios, provided that enough exploration is done (up to 500 iterations). On the other hand, ε -greedy allows to increase the average performance very quickly, but its growth stalls from iteration 200. Note that ε -greedy is based on the absolute throughput value, thus preventing to find a collaborative behavior in which the scarce resources are optimally shared. Finally, EXP3 and UCB are shown to improve the average throughput linearly, but offering poor performance.

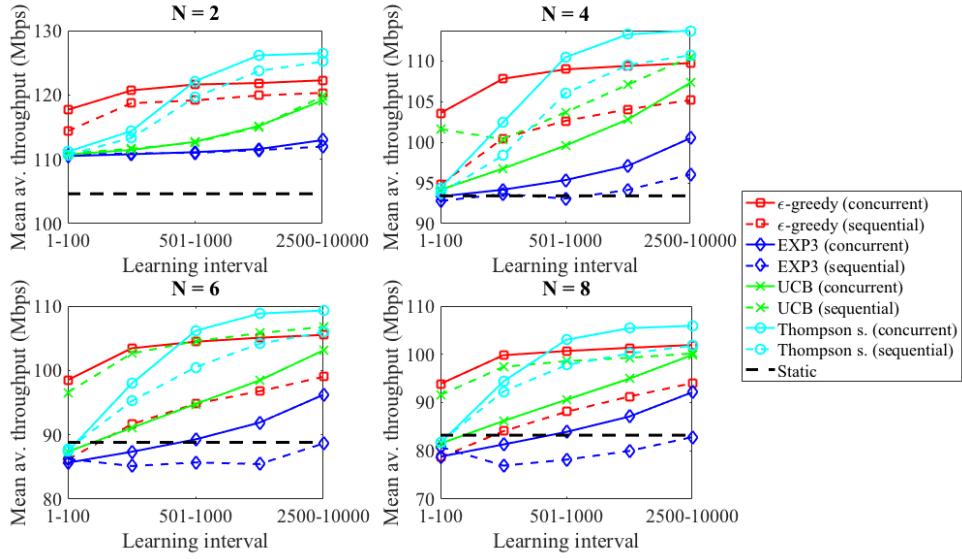


Figure 8: Average throughput experienced in each learning interval for each action-selection strategy. Results from 100 repetitions are considered for each different number of overlapping WNs ($N = \{2, 4, 6, 8\}$). The black dashed line indicates the default IEEE 802.11 performance (static situation).

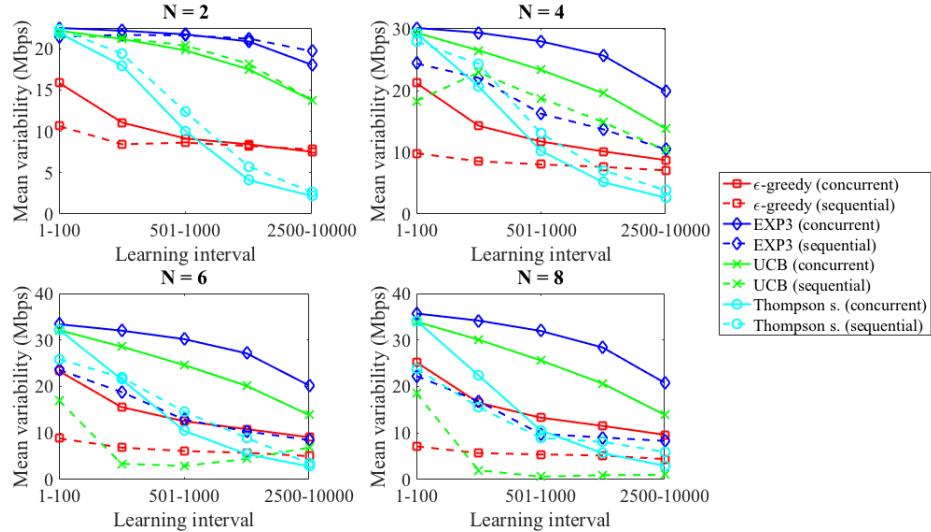


Figure 9: Average variability experienced in each learning interval for each action-selection strategy. Results from 100 repetitions are considered for each different number of overlapping WNs ($N = \{2, 4, 6, 8\}$).

When it comes to the sequential approach, we find the following:

- On the one hand, the average throughput is reduced in almost all the cases in comparison with the concurrent approach (see Figure 8). We find that this is due to the larger phases in which agents exploit sub-optimal actions. As previously pointed out, the time between iterations is considered to be the same for both sequential and concurrent learning approaches.

- On the other hand, the sequential procedure is shown to substantially improve the variability experienced by ε -greedy, EXP3 and UCB (see Figure 9). The performance of the latter is particularly shown to be improved when the learning procedure is ordered. The sequential approach, therefore, allows UCB to produce more accurate estimates on the actions rewards. In contrast, learning sequentially does not improve the concurrent version of Thompson sampling in any performance metric. We attribute this suboptimal behavior to the way in which Thompson sampling performs estimates of the played actions, which depends on the number of times each one is selected. In particular, suboptimal actions can eventually provide good enough performance according to the adversarial setting. The same issue can lead to underestimate optimal actions, so that their actual potential is not observed. Since Thompson sampling bases its estimates on the number of times each action is selected, the aforementioned effect may lead to increase the exploitation on suboptimal actions.

6 Conclusions

In this paper, we provided an implementation of MABs to address the decentralized SR problem in dense WNs. Unlike previous literature, we have focused on a situation in which few resources are available, thus bringing out the competition issues raised from the adversarial setting. Our results show that decentralized learning allows improving SR in dense WN, so that collaborative results in symmetric scenarios, sometimes close to optimal proportional fairness, can be achieved. This result is achieved even though WNs act selfishly, aiming to maximize their own throughput. In addition, this behavior is observed for random scenarios, where the effects of asymmetries cannot be controlled. These collaborative actions are, at times, accompanied by high temporal throughput variability, which can be understood as a consequence of the rate at which networks change their configuration in response of the opponents behavior. A high temporal variability may provoke negative issues in a node's performance, as its effects may be propagated to higher layers of the protocol stack. For instance, a high throughput fluctuation may entail behavioral anomalies in protocols such as Transmission Control Protocol (TCP).

We have studied this trade-off between fair resource allocation and high temporal throughput variability in ε -greedy, EXP3, UCB and Thompson sampling action-selection strategies. Our results show that while this trade-off is hard to regulate via the learning parameters in ε -greedy and EXP3, UCB and, especially, Thompson sampling are able to achieve fairness at a reduced temporal variability. We identify the root cause of this phenomena to the fact that both UCB and Thompson sampling consider the probability distribution of the rewards, and not only their magnitude.

Furthermore, for the sake of alleviating the temporal variability, we studied the effects of learning concurrent and sequentially. We have shown that learning in an ordered way is very effective to reduce the throughput variability for almost all the proposed learning strategies, even if WNs maintain a selfish behavior. By learning sequentially, more knowledge is attained on a given action, thus allowing to differentiate quickly between good and bad performing actions. Apart from that, we found that Thompson sampling grants significantly better results than the other examined algorithms since it is able to capture meaningful information from chaotic environments.

We left as future work to further study the MABs application to WNs through distributed (with message passing) and centralized (with complete information) approaches with shared reward. In particular, we would like to extend this work to enhance both throughput and stability by inferring the actions of the opponents and acting in consequence, as well as further investigating dynamic scenarios. Defining the resource allocation problem as an adversarial game is one possibility to do so. In addition to this, the utilization of multiple antenna strategies (i.e., single and multi-user beamforming and interference cancellation) is expected to further improve the spectral efficiency in future WNs. Through these techniques, the SR problem can be relaxed in a similar way than using several non-overlapping frequency channels. However, its application would significantly increase the problem's complexity, and its analysis is also left as future work.

Acknowledgments

This work has been partially supported by the Spanish Ministry of Economy and Competitiveness under the Maria de Maeztu Units of Excellence Programme (MDM-2015-0502), by a Gift from CISCO University Research Program (CG#890107) & Silicon Valley Community Foundation, by the European Regional Development Fund under grant TEC2015-71303-R (MINECO/FEDER), and by the Catalan Government under grant SGR-2017-1188.

The authors would like to thank the anonymous reviewers. Their dedication and insightful comments were of great help to improve this paper. We would also like to show our gratitude to Andrés Burbano, who also contributed to improve the quality of this document through his thorough revision.

References

- [1] Aditya Akella, Glenn Judd, Srinivasan Seshan, and Peter Steenkiste. Self-management in chaotic wireless deployments. *Wireless Networks*, 13(6):737–755, 2007.
- [2] Basel Alawieh, Yongning Zhang, Chadi Assi, and Hussein Mouftah. Improving spatial reuse in multihop wireless networks-a survey. *IEEE Communications Surveys & Tutorials*, 11(3), 2009.
- [3] Nikolaos I Miridakis and Dimitrios D Vergados. A survey on the successive interference cancellation performance for single-antenna and multiple-antenna ofdm systems. *IEEE Communications Surveys & Tutorials*, 15(1):312–335, 2013.
- [4] Konstantinos Dovelos and Boris Bellalta. Breaking the interference barrier in dense wireless networks with interference alignment. In *2018 IEEE International Conference on Communications (ICC)*, pages 1–6. IEEE, 2018.
- [5] Michael Littman and Justin Boyan. A distributed reinforcement learning scheme for network routing. In *Proceedings of the international workshop on applications of neural networks to telecommunications*, pages 45–51. Psychology Press, 1993.
- [6] Biljana Bojovic, Nicola Baldo, Jaume Nin-Guerrero, and Paolo Dini. A supervised learning approach to cognitive access point selection. In *GLOBECOM Workshops (GC Wkshps), 2011 IEEE*, pages 1100–1105. IEEE, 2011.
- [7] Biljana Bojovic, Nicola Baldo, and Paolo Dini. A neural network based cognitive engine for ieee 802.11 WLAN access point selection. In *Consumer Communications and Networking Conference (CCNC), 2012 IEEE*, pages 864–868. IEEE, 2012.
- [8] Richard Combes, Alexandre Proutiere, Donggyu Yun, Jungseul Ok, and Yung Yi. Optimal rate sampling in 802.11 systems. In *INFOCOM, 2014 Proceedings IEEE*, pages 2760–2767. IEEE, 2014.
- [9] Marco Miozzo, Lorenza Giupponi, Michele Rossi, and Paolo Dini. Distributed q-learning for energy harvesting heterogeneous networks. In *Communication Workshop (ICCW), 2015 IEEE International Conference on*, pages 2006–2011. IEEE, 2015.
- [10] Sébastien Bubeck and Nicoló Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning*, 5(1):1–122, 2012.
- [11] Francesc Wilhelmi, Boris Bellalta, Cristina Cano, and Anders Jonsson. Implications of decentralized q-learning resource allocation in wireless networks. In *Personal, Indoor, and Mobile Radio Communications (PIMRC), 2017 IEEE 28th Annual International Symposium on*, pages 1–5. IEEE, 2017.
- [12] Antonios Argyriou and Ashish Pandharipande. Collision recovery in distributed wireless networks with opportunistic cooperation. *IEEE Communications Letters*, 14(4), 2010.
- [13] Fulvio Babich, Massimiliano Comisso, Alessandro Crismani, and Aljoša Dorni. On the design of mac protocols for multi-packet communication in ieee 802.11 heterogeneous networks using adaptive antenna arrays. *IEEE Transactions on Mobile Computing*, 14(11):2332–2348, 2015.
- [14] Janne Riihijarvi, Marina Petrova, and Petri Mahonen. Frequency allocation for WLANs using graph colouring techniques. In *Wireless On-demand Network Systems and Services, 2005. WONS 2005. Second Annual Conference on*, pages 216–222. IEEE, 2005.
- [15] Arunesh Mishra, Suman Banerjee, and William Arbaugh. Weighted coloring based channel assignment for WLANs. *ACM SIGMOBILE Mobile Computing and Communications Review*, 9(3):19–31, 2005.

- [16] Robert Akl and Anurag Arepally. Dynamic channel assignment in IEEE 802.11 networks. In *Portable Information Devices, 2007. PORTABLE07. IEEE International Conference on*, pages 1–5. IEEE, 2007.
- [17] Jeremy K Chen, Gustavo De Veciana, and Theodore S Rappaport. Improved measurement-based frequency allocation algorithms for wireless networks. In *Global Telecommunications Conference, 2007. GLOBECOM’07. IEEE*, pages 4790–4795. IEEE, 2007.
- [18] Xiaonan Yue, Chi-Fai Wong, and S-H Gary Chan. Cacao: Distributed client-assisted channel assignment optimization for uncoordinated wlans. *IEEE Transactions on Parallel and Distributed Systems*, 22(9):1433–1440, 2011.
- [19] Tamer A ElBatt, Srikanth V Krishnamurthy, Dennis Connors, and Son Dao. Power management for throughput enhancement in wireless ad-hoc networks. In *Communications, 2000. ICC 2000. 2000 IEEE International Conference on*, volume 3, pages 1506–1513. IEEE, 2000.
- [20] Suhua Tang, Hiroyuki Yomo, Akio Hasegawa, Tatsuo Shibata, and Masayoshi Ohashi. Joint transmit power control and rate adaptation for wireless LANs. *Wireless personal communications*, 74(2):469–486, 2014.
- [21] Carlos Gandarillas, Carlos Martín-Engeños, Héctor López Pombo, and Antonio G Marques. Dynamic transmit-power control for WiFi access points based on wireless link occupancy. In *Wireless Communications and Networking Conference (WCNC), 2014 IEEE*, pages 1093–1098. IEEE, 2014.
- [22] Pierre Coucheney, Kinda Khawam, and Johanne Cohen. Multi-armed bandit for distributed inter-cell interference coordination. In *ICC*, pages 3323–3328, 2015.
- [23] Setareh Maghsudi and Sławomir Stańczak. Joint channel selection and power control in infrastructure-less wireless networks: A multiplayer multiarmed bandit framework. *IEEE Transactions on Vehicular Technology*, 64(10):4565–4578, 2015.
- [24] Setareh Maghsudi and Sławomir Stańczak. Channel selection for network-assisted D2D communication via no-regret bandit learning with calibrated forecasting. *IEEE Transactions on Wireless Communications*, 14(3):1309–1322, 2015.
- [25] William R Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294, 1933.
- [26] Tze Leung Lai and Herbert Robbins. Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics*, 6(1):4–22, 1985.
- [27] Peter Auer, Nicoló Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2-3):235–256, 2002.
- [28] Peter Auer, Nicoló Cesa-Bianchi, Yoav Freund, and Robert E Schapire. Gambling in a rigged casino: The adversarial multi-armed bandit problem. In *Foundations of Computer Science, 1995. Proceedings., 36th Annual Symposium on*, pages 322–331. IEEE, 1995.
- [29] Peter Auer, Nicoló Cesa-Bianchi, Yoav Freund, and Robert E Schapire. The nonstochastic multiarmed bandit problem. *SIAM journal on computing*, 32(1):48–77, 2002.
- [30] Peter Whittle. Restless bandits: Activity allocation in a changing world. *Journal of applied probability*, 25(A):287–298, 1988.
- [31] Lihong Li, Wei Chu, John Langford, and Robert E Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*, pages 661–670. ACM, 2010.

- [32] Naoki Abe, Alan W Biermann, and Philip M Long. Reinforcement learning with immediate rewards and linear hypotheses. *Algorithmica*, 37(4):263–293, 2003.
- [33] Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. In *Advances in Neural Information Processing Systems*, pages 2312–2320, 2011.
- [34] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*, volume 1. MIT press Cambridge, 1998.
- [35] Rajeev Agrawal. Sample mean based index policies by $\mathcal{O}(\log n)$ regret for the multi-armed bandit problem. *Advances in Applied Probability*, 27(4):1054–1078, 1995.
- [36] Apostolos N Burnetas and Michael N Katehakis. Optimal adaptive policies for sequential allocation problems. *Advances in Applied Mathematics*, 17(2):122–142, 1996.
- [37] Nick Littlestone and Manfred K Warmuth. The weighted majority algorithm. *Information and computation*, 108(2):212–261, 1994.
- [38] Yoav Freund and Robert E Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of computer and system sciences*, 55(1):119–139, 1997.
- [39] Olivier Chapelle and Lihong Li. An empirical evaluation of thompson sampling. In J. Shawe-Taylor, R. S. Zemel, P. L. Bartlett, F. Pereira, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 24*, pages 2249–2257. Curran Associates, Inc., 2011.
- [40] Shipra Agrawal and Navin Goyal. Analysis of thompson sampling for the multi-armed bandit problem. In *Conference on Learning Theory*, pages 39–1, 2012.
- [41] Emilie Kaufmann, Nathaniel Korda, and Rémi Munos. Thompson sampling: An asymptotically optimal finite time analysis. In *Algorithmic Learning Theory*, pages 199–213. Springer, 2012.
- [42] Nathaniel Korda, Emilie Kaufmann, and Remi Munos. Thompson sampling for 1-dimensional exponential family bandits. In C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 1448–1456. Curran Associates, Inc., 2013.
- [43] Shipra Agrawal and Navin Goyal. Further optimal regret bounds for thompson sampling. In *Artificial Intelligence and Statistics*, pages 99–107, 2013.
- [44] Sujay Sanghavi, Devavrat Shah, and Alan S Willsky. Message passing for maximum weight independent set. *IEEE Transactions on Information Theory*, 55(11):4822–4834, 2009.
- [45] Kai Yang, Narayan Prasad, and Xiaodong Wang. A message-passing approach to distributed resource allocation in uplink dft-spread-ofdma systems. *IEEE transactions on Communications*, 59(4):1099–1113, 2011.
- [46] Boris Bellalta. IEEE 802.11 ax: High-efficiency WLANs. *IEEE Wireless Communications*, 23(1):38–46, 2016.
- [47] Francesc Wilhelmi. Collaborative spatial reuse in wireless networks via selfish mulit-armed bandits. https://github.com/fwilhelmi/collaborative_sr_in_wns_via_selfish_mabs. Commit: b9957e2, 2017.

Potential and Pitfalls of Multi-Armed Bandits for Decentralized Spatial Reuse in WLANs

Francesc Wilhelmi, Sergio Barrachina-Muñoz, Boris Bellalta,
Cristina Cano, Anders Jonsson, and Gergely Neu

Abstract

Spatial Reuse (SR) has recently gained attention to maximize the performance of IEEE 802.11 Wireless Local Area Networks (WLANs). Decentralized mechanisms are expected to be key in the development of SR solutions for next-generation WLANs, since many deployments are characterized by being uncoordinated by nature. However, the potential of decentralized mechanisms is limited by the significant lack of knowledge with respect to the overall wireless environment. To shed some light on this subject, we show the main considerations and possibilities of applying online learning to address the SR problem in uncoordinated WLANs. In particular, we provide a solution based on Multi-Armed Bandits (MABs) whereby independent WLANs dynamically adjust their frequency channel, transmit power and sensitivity threshold. To that purpose, we provide two different strategies, which refer to selfish and environment-aware learning. While the former stands for pure individual behavior, the second one considers the performance experienced by surrounding networks, thus taking into account the impact of individual actions on the environment. Through these two strategies we delve into practical issues of applying MABs in wireless networks, such as convergence guarantees or adversarial effects. Our simulation results illustrate the potential of the proposed solutions for enabling SR in future WLANs. We show that substantial improvements on network performance can be achieved regarding throughput and fairness.

1 Introduction

Wireless communications are rapidly evolving to satisfy the increasingly tighter requirements coming from the explosive growth of wireless devices. To solve that, future Wireless Networks (WNs) are foreseen to cover small areas in high-density scenarios, which evidences the need for novel mechanisms to maximize spectral efficiency. In particular, Spatial Reuse (SR) has gained attention in recent years as a potential solution to improve the use of the spectrum. One of the most prominent examples can be found in the IEEE 802.11ax-2019 (11ax) amendment (Bellalta, 2016), which defines High-Efficiency (HE) Wireless Local Area Networks (WLANs). The 11ax amendment aims to maximize spectral efficiency through the SR operation and other spectrum-efficient techniques like Orthogonal Frequency-Division Multiple Access (OFDMA), and Uplink/Downlink Multi-User Multiple-Input-Multiple-Output (MU-MIMO).

In this paper, we focus on IEEE 802.11 WLANs, which mostly represent uncoordinated deployments (e.g., residential buildings). These networks are limited in performance because of the scalability issues arising from the current decentralized channel access mechanisms, i.e., Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) (Ergin et al., 2007). To enable SR in WLANs, we consider the use of Transmit Power Control (TPC) and Carrier Sense Threshold (CST) adjustment. These mechanisms are of particular concern to the 11ax amendment. They facilitate and regulate the SR operation by providing a set of procedures and constraints for dynamically setting the transmit power and the sensitivity. Roughly, the idea of TPC lies in adjusting the transmit power for reducing the interference and/or saving energy. Similarly, CST adjustment seeks to increase the number of parallel transmissions by modifying the sensitivity on a per-device basis. In the context of IEEE 802.11ax, the Dynamic Sensitivity Control (DSC) has been proposed as a potential solution for enabling SR through sensitivity adjustment (Smith, 2015).

However, addressing the SR problem in WLANs through TPC and CST adjustment is not trivial for decentralized deployments. This is mostly motivated by *i*) the spatial interactions between nodes, and *ii*) the adversarial setting unleashed in decentralized wireless networks. For the former, tuning either the transmit power or the sensitivity entails dealing with the spatial dimension. Unlike for frequency and temporal approaches, spatial interference cannot be treated as a binary model. In the latter case, one can observe when devices transmit or not, thus obtaining full or null performance respectively. As shown later in Section 3, more complex interactions occur between WLANs tuning their transmit power and CST. As a result, modeling inter-WLANs interactions in next-generation deployments turns out to be extremely complex. Moreover, regarding the adversarial setting unleashed by decentralized deployments, strong competition between independent networks may occur.

In order to address the SR problem in high-density decentralized networks, we focus on the Multi-Armed Bandit (MAB) framework. The MAB approach frames the learning-by-interaction problem, and allows to properly approach the exploration-exploitation trade-off in face of uncertainty. In MABs, a learner (or agent) obtains information from the environment, which reacts according to the actions performed - in the SR problem, an action may refer to a certain configuration of transmit power and CST. By interacting with the environment, a given learner aims to maximize a numerical cumulative reward over time. Unlike classical Reinforcement Learning (RL), the MAB setting does not consider states in general.¹ A state may refer to a concrete temporal situation in which the learner is involved. Therefore, it allows the latter to construct a policy that determines the behavior to be followed in future situations. Accordingly, learning through states adds an extra layer of complexity and requires that a given agent learns additional contextual information.

The application of MABs into wireless communications problems has recently become very popular (Chen, 2010; Maghsudi and Stańczak, 2015a,b). To model the SR problem through MABs, we consider that WLANs are empowered with an agent that attempts to learn the best-performing action (i.e., a combination of the frequency channel, the transmit power and the sensitivity level). Its learning operation is based on the performance achieved in an unknown environment. In this way, MABs operate on the top of CSMA/CA, which operation is influenced by the spatial interactions generated by the taken actions. As a result, we expect WLANs to autonomously find their best configuration in an adversarial setting, given a performance maximization strategy.

The main goal of this paper, then, is to determine the feasibility of applying decentralized learning to improve spectral efficiency in next-generation wireless deployments. In particular, we apply online learning mechanisms to enable SR in dense and uncoordinated WLANs, and show the main derived implications and considerations. We highlight the impact on the aggregate performance and fairness experienced by WLANs, as well as on the guarantees for converging to the optimal solution. The implications of applying online learning to WLANs are studied through the utilization of selfish and environment-aware learning-based strategies. While a selfish strategy is based on the individual performance of a given learner, environment-aware considers a set of neighboring WLANs that share a reward. The SR problem presented in this work is non-convex, therefore prevents to provide any kind of convergence bound. However, our results suggest that the performance of WLANs can be maximized by using learning strategies that are based on probabilistic models of the reward. To summarize, the main contributions of this work are described next:

- We showcase the major inter-WLAN dependencies when modifying both the transmit power and the CST, and how they affect the network performance.
- In order to capture the CSMA/CA operation of IEEE 802.11 WLANs, we use Continuous Time Markov Networks (CTMNs) in spatially-distributed scenarios (Bellalta, 2017). We show that CTMNs models are able to capture the existing dependencies between overlapping WLANs.
- We model the SR problem in WLANs through MABs, where agents implementing Thompson sampling decide the configuration of a given network in terms of frequency channel, transmit power and CST. To the best of our knowledge, this is the first work applying MABs on a CSMA-based network.

¹There is a class of bandits problems that consider states (stateful bandits), which is not considered due to the characteristics of the problem addressed in this work. Essentially, in this class of bandits, states are usually represented by taking strong assumptions that hinder the accuracy of the analysis.

- We provide insights on the main considerations of using learning in decentralized and adversarial wireless networks. In particular, we showcase the implications of applying selfish and environment-aware learning in dense WLANs, thus emphasizing on the main potentials and pitfalls.
- We evaluate the performance of using self-configuring agents in dense WLANs, both in specific and random scenarios. The two learning approaches presented in this paper are shown to significantly improve the performance achieved by WLANs in terms of throughput and fairness, with respect to a default - and static - configuration.

The remaining of this document is structured as follows: Section 2 refers to the previous related work. In Section 3 we first provide details on the throughput model considered in this paper to fit the SR problem. Then, we characterize inter-WLAN interactions when tuning both the transmit power and the sensitivity threshold. A set of illustrative scenarios is used for that purpose. Section 4 formulates the SR problem through MABs and shows the main implications to be considered when applying learning to decentralized wireless networks. Then, Section 5 shows the main results of this paper with regard to selfish and environment-aware learning in WLANs. Finally, some remarks are given in Section 6.

2 Related Work

Machine Learning (ML), and more precisely RL, has received increasing interest from the wireless communications research community over the last years. One of the main reasons resides in the increased complexity of problems related to next-generation wireless systems. Such kind of environments are characterized by being particularly dense, so that the best configuration strategy may be difficult to foresee. Since the current preprogrammed approaches are likely to be suboptimal, RL is expected to improve the action selection from experience. In particular, RL-based methods are expected to provide close-to-optimal solutions to complex problems within an acceptable timescale, which is an indispensable requirement in wireless networks.

To the best of our knowledge, one of the first works to apply RL into a SR-related problem in wireless networks is (Nie and Haykin, 1999), in which the authors show a centralized Q-learning mechanism to dynamically select the channel in mobile networks. Other RL-based approaches for channel access can be found in (Li, 2009; Bennis and Niyato, 2010; Bennis et al., 2011; Sallent et al., 2015; Rupasinghe and Güvenç, 2015), covering cognitive radio, self-organizing cellular networks and coexistence problems. Despite Q-learning (or other Markovian-based methods) has been shown to properly fit to channel allocation problems, few applications have been provided to the SR problem. Note, as well, that dealing with the frequency domain allows to naturally define states,² which can be done according to the availability of channels (typically modeled through Bernoulli distributions). Therefore, an agent may observe the environment and define an accurate model where the state is defined by the set of channels that are available/occupied. Note that the contextual information provided to the learner is important for learning efficiently, since the agent is able to react to different situations. With a proper definition of states, a higher degree of control is conferred to the agent. Therefore, provided that the model of the states is accurate enough, the learning procedure carried out by a given learner can result into better performance than that of a stateless setting.

However, modeling states for the decentralized SR problem entails added complexity, thus hindering the learning procedure followed by a given agent. In the particular case of IEEE 802.11 WLANs, spatial interactions among nodes lead to complex scenarios. The performance achieved by a given WLAN depends on the additive interference coming from an unknown environment. Therefore, learning accurate enough states for the SR problem turns out to be challenging. Note that, if states do not reflect the actual situation of a given agent at a given moment, the learnings that can be generated become strongly limited, and can even be meaningless. To cope with the difficulties on modeling states for the decentralized SR problem, we focus on multi-player MABs (MP-MABs), which frame resource allocation problems where several agents compete against each other. MP-MABs have been recently broadly applied for opportunistic spectrum access in cognitive radio (Liu and Zhao, 2010; Anandkumar et al., 2011; Rosenski et al., 2016; Maghsudi and Stańczak, 2015b,a).

²An state describes a particular situation of a given agent at a specific time.

Firstly, in (Liu and Zhao, 2010), the authors provide a decentralized policy with logarithmic regret order, which is based on a time-division fair sharing of the best arms. However, such a policy requires coordination among agents and to know the exact number of adversarial nodes, which in addition must be constant and known in advance. Both requirements entail dedicated communication between nodes, which turns out to be unfeasible for decentralized problems such as the one presented in this paper. Another important contribution regarding multi-player learning for the opportunistic spectrum access problem is provided in (Anandkumar et al., 2011), where the authors provide a distributed learning algorithm that showcases order-optimal regret. However, the total number of secondary users is known by the system, which may not be feasible in real scenarios where no communication between nodes exists. In contrast, in this work we consider selfish and environment-aware learning approaches, none of which require explicit communication between independent learners. Furthermore, a less strict method is also provided in (Anandkumar et al., 2011), in which the number of secondary users (which is fixed) is estimated, so that nearly order-optimal regret is achieved. In both algorithms, it is assumed that all the users use the same policy. Regarding the work in (Rosenski et al., 2016), sublinear regret is achieved if all players implement the proposed algorithm. Some interesting thoughts are provided regarding varying environments, which, to the best of our knowledge, have been barely considered in the previous literature. For instance, the authors emphasize that, in the dynamic setting, the frequency at which players enter and leave the scenario must be limited in order to provide a sublinear regret. Unlike the SR presented in this paper, the work in (Rosenski et al., 2016) assumes that there exists an optimal solution whereby no collisions occur for any player. The concept of collision is inspired in the ALOHA channel access mechanism, and occurs if two or more players choose the same arm (or channel).

When it comes to the SR problem by means of TPC and/or sensitivity adjustment, we find MP-MABs application for joint channel selection and power control in (Maghsudi and Stańczak, 2015b,a). The work in (Maghsudi and Stańczak, 2015b) proposes a strategy based on the Signal-to-Interference-plus-Noise Ratio (SINR) to determine the channel and the transmit power to be used in cognitive radio networks. The authors prove that the MP-MAB game converges to a correlated equilibrium, which in addition maximizes the aggregate utility, if the two following assumptions hold: *i*) the problem is relaxed, so that the reward granted to a given agent depends only on the Signal-to-Noise Ratio (SNR), regardless of the overlapping interference, *ii*) user-specific penalties are provided to each agent. In contrast with the work in (Maghsudi and Stańczak, 2015b), here we aim to understand the potentials and limitations of applying MABs in a CSMA/CA-based setting, in which the previous assumptions do not hold. First of all, instead of relaxing the problem to only consider the SNR, we model the interactions at the MAC level. Secondly, defining user-specific penalties would require the use of either a centralized system or message exchanging.

Finally, in (Maghsudi and Stańczak, 2015a), the authors introduce the concept of calibrated forecaster, i.e., a predictor of the actions of the adversaries that improves with collected knowledge. By using such a predictor, if every learner is able to predict and respond to the others' actions, then the game converges to a correlated equilibrium. In other words, if a node can predict which channel will the adversary pick (and vice-versa), then it can select the other channel and experience the maximum performance. In contrast, in the SR problem we tackle here, developing an accurate forecaster in a decentralized way may be an extremely complex task. Refer to the non-linear relationships that occur in the spatial domain, which are hard to model, and thus to predict. Moreover, density and messy deployments may remove the existence of an equilibrium, hence invalidating the assumptions.

In summary, a lot of effort has been recently made to enable the evolution of wireless networks towards self-adjusting systems. In particular, the application of RL has been extensively studied for channel access problems. However, these type of methods do not properly suit to spatially-distributed problems such as the SR one. As a result, other stateless techniques, such as MABs, have been targeted, and have shown to effectively improve wireless networks performance, even in adversarial environments. Nevertheless, these mechanisms require that strong assumptions about the system model hold. Moreover, spatial interactions between WLANs have not yet been considered.

3 Interactions between WLANs when Spatial Reuse is Enabled

In this Section, for completeness, we first briefly introduce the CSMA/CA operation carried out by Wi-Fi networks for accessing the channel, as well as the CSMA/CA throughput model considered along this paper. Therewith, we aim to identify the main inter-WLAN interactions when modifying both the transmit power and the CST. Understanding these interactions is key to motivate the usage of MABs to the decentralized SR problem. As shown in Section 2, some of the previous work addressed similar problems and provided mechanisms that were proven to converge to an equilibrium. However, the novelty of this paper lies in the analysis of learning techniques in CSMA/CA-based networks. Unlike previous work, where the reward (i.e., the throughput) is mostly given by a linear function that only depends on the signal strength and the interference, here we deal with more complex interaction between networks. In Wi-Fi, due to the decentralized nature of CSMA/CA, an optimal solution in terms of SR is harder to derive than in cellular-based networks. In addition, there is a trade-off between performance maximization and fairness, which is not trivial to compute in a decentralized setting.

3.1 CSMA/CA

Channel access is performed in IEEE 802.11 WLANs by means of the Distributed Coordination Function (DCF), which is based on CSMA/CA. In DCF, before being able to transmit a packet, a transmitter must listen to the channel for a period of time called Distributed Inter Frame Space (DIFS). The channel is sensed to be free according to the Clear Channel Assessment (CCA) mechanism, i.e., if the power perceived is lower than a given threshold.³ The power received at a given node is the sum of all the interference generated by the other devices under the environment-constrained propagation effects. Furthermore, the access to the medium is randomized in order to reduce the number of potential collisions between other contenders. Specifically, each transmitter selects a random backoff value to start a countdown that is active as long as the channel remains free. In case the channel is sensed busy, the countdown is paused. It is resumed as soon as the ongoing transmission finishes and the channel is sensed free again. An example of the CSMA/CA operation is shown in Figure 1, where we show two overlapping WLANs, namely WLAN_A and WLAN_B, respectively. Station A (STA_A) is the first to gain access to the channel, so it starts a transmission to the Access Point A (AP_A). Meanwhile, AP_B senses the channel busy and freezes its backoff. After this first transmission, both AP_A and AP_B access to the channel simultaneously because they both randomly chose the same backoff counter. As a result, a collision is produced.

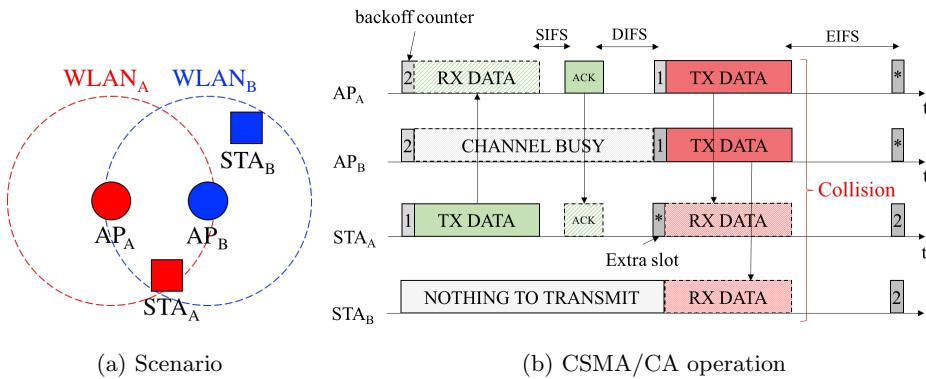


Figure 1: CSMA/CA operation in WLAN_A and WLAN_B. STA_A starts a transmission to AP_A, since its backoff counter reaches zero first. After that, a collision occurs due to simultaneous transmissions held by AP_A and AP_B.

³Throughout this paper, we refer to CCA and CST indistinctly.

3.2 CSMA/CA Throughput Model

For the rest of this paper we consider that WLANs are independent entities composed by an AP and a STA, in which saturation downlink traffic (i.e., from the AP to the STA) is assumed. Such an assumption is reasonable as long as we target home deployments, where STAs are often very close to the AP. Moreover, the main goal of this paper is to capture inter-WLAN interactions.

Regarding the throughput model, we rely on the CTMN-based analytical tool for spatially distributed WLANs presented in (Barrachina-Muñoz et al., 2018b), referred to as SFCTMN. This tool captures the interrelations given in scenarios where nodes operating in the same channel are not required to be within the carrier sense range of each other. Essentially, given a scenario (i.e., nodes' location, channels, transmission powers, CCA levels, path loss model, etc.), states and transitions are generated in accordance with the CSMA/CA mechanism. That is, WLANs are only allowed to decrement their backoff and start transmissions when the CCA condition is accomplished. In (Barrachina-Muñoz et al., 2018a), authors use SFCTMN to assess the performance of high density WLANs under different traffic loads.

A state in the CTMN is defined by the set of WLANs active and the channel in which they are transmitting.⁴ Accordingly, transitions between states occur if WLANs become active/disabled. For example in state $A_1B_2C_1$ there are three active WLANs: A, B and C transmit in channels 1, 2 and 1, respectively. Since states and transitions are generated according to the regular CSMA/CA mechanism, a CTMN may have both bidirectional and unidirectional transitions between states. It is the case of the toy scenario shown in Figure 2, where A uses a higher transmission power than B and C. While A is able to access channel 1 when C is transmitting, C is not able to do so when A is transmitting because of the high interference sensed in channel 1. Accordingly, only backward transitions are permitted from state $s_6 = A_1C_1$ to $s_2 = A_1$, and from state $s_8 = A_1B_2C_1$ to $s_6 = A_1B_2$. Essentially, given the channel and power configurations of this particular scenario, while A operates like in isolation, C's operation is subject to A's behavior. Note that B also operates like in isolation since it uses a different channel.

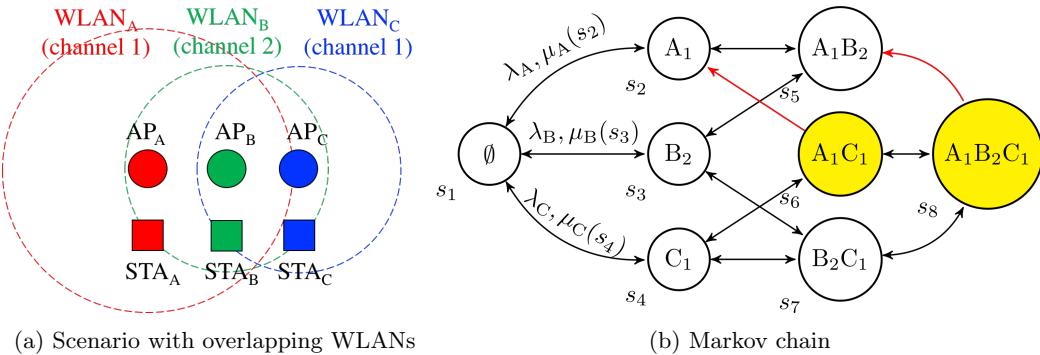


Figure 2: Toy scenario. WLANs A and C operate in channel 1 while B operates in channel 2. Note that C is in the carrier-sense range of A. Only the transition rate pairs (λ, μ) between states s_1, s_2, s_3 and s_4 are displayed for the sake of visualization. The states where C may lose packets because of the interference from A are displayed in yellow. Unidirectional backward transitions are shown in red.

Transitions between any two states s and s' in the CTMN have a corresponding transition rate $R_{s,s'}$. For *forward* transitions (i.e., when a WLAN starts a new transmission), the average packet transmission attempt rate is $\lambda = 1/E[B]$, being $E[B]$ the expected backoff duration. For *backward* transitions (i.e., when a WLAN finishes its transmission), the departure rate (μ) depends on the duration of a transmission. The latter is given by both the data rate (subject to the selected Modulation and Coding Scheme (MCS) and transmission channel width) and the average data packet length ($E[L]$). Thus, we simply say that the data rate of a WLAN w depends on the state s of the system, $\mu_w(s)$, in other words, on the set of overlapping

⁴Note that in this work we assume only 20 MHz single-channel transmissions.

Scenario	Conf.	$\bar{\Gamma}$ (SFCTMN)	$\bar{\Gamma}$ (11axHDWLANSim)
1	C1	56.90 Mbps	56.94 Mbps
	C2	113.23 Mbps	113.23 Mbps
	C3	62.43 Mbps	62.43 Mbps
2	C2	0.73 Mbps	0.00 Mbps
	C4	62.43 Mbps	62.43 Mbps
3	C1	56.62 Mbps	56.62 Mbps
	C1 & C5	113.23 Mbps	113.23 Mbps

Table 1: Performance in each scenario achieved through different configurations. Each cell contains the performance computed by using CTMNs and Komondor, respectively. Komondor results are extracted from 1,000 s simulations.

WLANs that transmit simultaneously. The information contained in a given state, therefore, refers to the inter-WLAN interactions in that situation.

In order to estimate the average throughput experienced by each WLAN in a given scenario, we must first estimate the fraction of time the system spends in each state ($\vec{\pi}$). We define π_s as the probability of finding the system at state s . In continuous-time Markov processes with stationary distribution, $\vec{\pi}$ is given by solving the system of equations $Q\vec{\pi} = 0$, where the matrix item Q is the infinitesimal generator of the CTMN. Given $\vec{\pi}$, the average downlink throughput of WLAN w in a given state s can be defined as

$$\Gamma_w(s) := \begin{cases} E[L]\mu_w(s)\pi_s, & \text{SINR}_w(s) > \text{CE} \\ 0, & \text{otherwise} \end{cases}$$

where $\text{SINR}_w(s)$ is the SINR perceived by the receiving STA in WLAN w in state s , and CE is the capture effect threshold. Therefore, the resulting average downlink throughput that a given WLAN w experiences can be computed as $\Gamma_w = \sum_{s \in \mathcal{S}} \Gamma_w(s)$.

3.3 Analysis

To underline the potential of adjusting both the transmit power and the CST to enable SR in overlapping WLANs, we next introduce the main performance issues and anomalies that characterize IEEE 802.11 networks. Before, and in order to further analyze these issues, we introduce the set of scenarios shown in Figure 3. This set of scenarios is evaluated under different static configurations (shown in Table 3(d)), each one referring to a specific combination of channel, transmit power and CCA. Such combinations (from $C1$ to $C5$) refer to specific configurations of the set of allowed values that a given WLAN can choose, which are detailed in A, together with simulation parameters. Results shown in Table 1 were obtained by applying the analytical model presented in Section 3.2⁵ and the 11axHDWLANSim simulator.⁶

- **Exposed-terminal problem:** two or more WLANs are not able to transmit simultaneously due to the inter-WLAN interference, which is higher than the CCA threshold at the transmitter. However, the receiver would be able to properly decode the data of interest, even in presence of other transmissions. In *Scenario 1* (Figure 3(a)), the exposed-terminal problem occurs if all the WLANs use configuration $C1$. Such a situation is solved if WLANs apply configuration $C2$, which consists in increasing the CCA in a way that both AP_A and AP_B can transmit simultaneously while using the same transmit power. In this case, both WLANs receive the same interference, but, by using a higher CCA, it is dismissed and does not force contention. Similarly, if WLANs reduce both the transmit power and

⁵All of the source code used in this work is open (Wilhelmi, 2018) under the GNU General Public License v3.0, encouraging sharing of knowledge between potential contributors.

⁶The source code of 11axHDWLANSim is open under the GNU General Public License v3.0 and can be found at <https://github.com/wn-upf/Komondor>

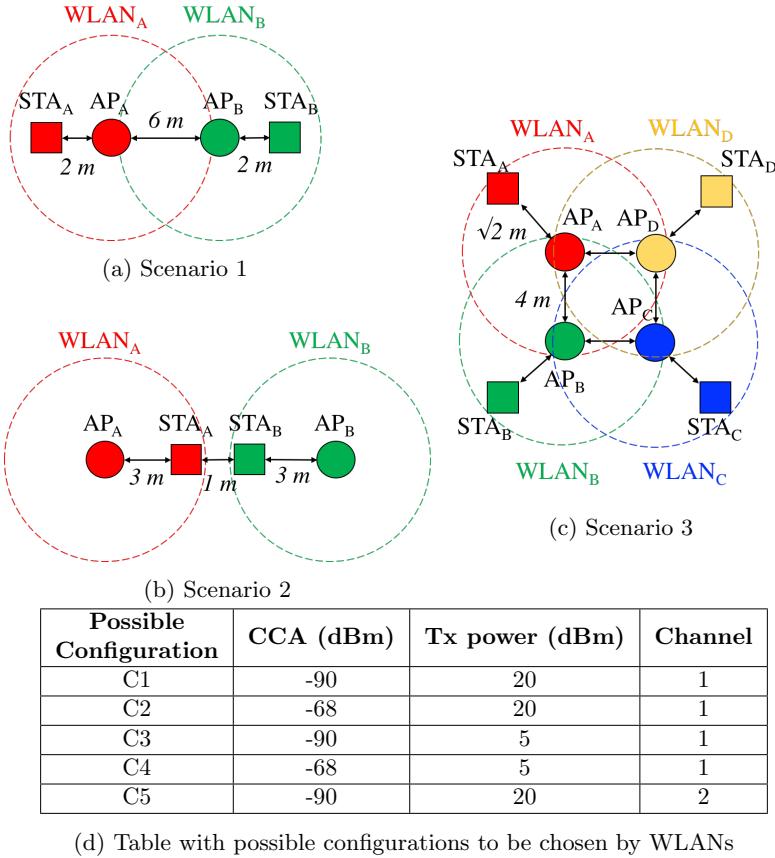


Figure 3: Scenarios for characterizing inter-WLAN interactions.

the sensitivity (configuration *C3*), the number of parallel transmissions can be increased. However, a lower performance compared to *C2* is achieved due to the use of a lower MCS.

- **Hidden-terminal problem:** occurs when two nodes that are not visible each other transmit simultaneously (not necessarily to the same destination), thus producing collisions. In particular, packet losses occur when the sensed interference at a given receiver results in a SINR lower than its capture effect. The hidden-terminal problem is framed in *Scenario 2* (Figure 3(b)) when both WLANs use configuration *C1*. As a result, AP_A and AP_B can transmit simultaneously due to the CCA condition. However, if they do so, the SINR experienced at both STA_A and STA_B falls below their capture effect, thus leading to a wrong packet decoding. Such a situation is improved when AP_A and AP_B apply configuration *C4*, which allows reducing the sensitivity area (higher CCA) and the generated interference (lower transmit power).
- **Contending nodes:** similarly to the exposed-terminal problem, the channel is underutilized if one or more WLANs must defer their transmissions when another one is transmitting. In this case, increasing the CST and/or decreasing the transmit power in an appropriate manner may help at reducing the number of contending nodes. As a result, the number of parallel transmissions is maximized. This phenomena, in addition to be closely related to the exposed-terminal problem shown in *Scenario 1*, can be further observed in *Scenario 3* (Figure 3(c)). In this case, in addition of using TPC and/or CST adjustment, we maximize SR by providing a proper channel allocation, which is done by combining configurations *C1* and *C5*. Hence, configurations are assigned so that frequency reuse is maximized.

Moreover, there are situations that may require the opposite. That is to say, to force contention between nodes in order to prevent collisions. Such a situation occurs in *Scenario 2* when configuration *C2* is employed,⁷ thus leading to zero throughput in both WLANs due to the collisions by hidden node. However, if a contending situation is forced by either increasing the transmit power or decreasing the CST (which occurs when setting configuration *C4*), then the network performance is increased.

- **Flow starvation:** a given WLAN may be deprived of accessing the channel in case of noticing an excessive interference from other WLANs that do not sense each other. Such a phenomena can be solved by tuning both the transmit power and the CST. However, due to the nature of the problem, it may require some level of collaboration, since interfering nodes are completely unaware on the damage caused to the most vulnerable WLANs in terms of sensed interference. Flow starvation is studied in detail in Section 4.
- **Asymmetries:** finally, it is important to remark the consequences of existing asymmetries in a wireless network, which are mostly generated by the different situations of coexisting WLANs. The performance of a given WLAN is basically limited by its geographical location and possible configurations. Accordingly, there can be WLANs more privileged than others, so that the interference they sense is generally lower, thus experiencing a higher performance. Therefore, due to the spatial interactions generated by certain transmit power and CST levels, asymmetries may lead to a monopolization of the channel by dominant WLANs (i.e., enjoying better conditions than others). The effect of asymmetries is studied in detail in Section 5.

As shown in the previous simulations, modifying either the transmit power or the CST in a WLAN may have severe implications on different communication aspects due to the utilization of CSMA/CA. While TPC allows to adjust the generated interference, CST adjustment aims to modify the sensitivity area. It is worth to mention that SR can be enhanced if short-range communications are held, which can be achieved if using the minimum necessary transmit power and the maximum possible CST. Conversely, longer-range communications can be achieved when using a high transmit power and a low CST. Increasing the area of operation is useful to minimize performance issues such as flow starvation and collisions by hidden nodes. Table 2 summarizes the intuitive effects of TPC and CST adaptation in WLANs.

Action	Effect		
	Exposed nodes	Hidden nodes	Data Rate
↑ Power	↑	↓	↑
↓ Power	↓	↑	↓
↑ CST	↓	↑	↓
↓ CST	↑	↓	↑

Table 2: Effects of TPC and CST adjustment.

4 Multi-Armed Bandits for Decentralized Spatial Reuse

Due to the nature of the CSMA/CA protocol - especially hampered in high-density scenarios - and the rigidity of the current configurations used by wireless devices (Akella et al., 2007), network overlapping drives into many problems and situations that result into poor throughput performance. Our goal is to provide a solution that enhances SR in an online fashion. To this end, we model the problem in which

⁷There is a significant difference in the throughput when applying *C2* to *Scenario 2* between the CTMNs model and **11axHDLANSim** simulator. The fact is that CTMNs consider the time spent in each state. In this case, the dominant state is the one in which both WLANs transmit and experience collisions. However, the time spent in states where individual transmissions are held is considered, even if it is very small. In practice, transmissions affected by overlapping interference would result into null throughput, which is shown via the **11axHDLANSim** simulator.

multiple WLANs contend for a common set of resources through adversarial MABs. The adversarial MAB problem (Auer et al., 1995) frames the scenario in which different learners compete for the same resources simultaneously. In particular, after taking an action, a given learner is granted with a reward that depends on the others' actions, i.e., the joint action profile.

MABs have been shown to properly deal with the exploration-exploitation trade-off when the uncertainty level is very high (Auer et al., 2002; Audibert et al., 2009; Scott, 2010), which properly addresses the decentralized SR problem studied in this work. Note that learning WLANs do not have information about the environment, which in addition is adversarial because of the competition for the channel resources. Moreover, the interactions shown by CSMA/CA-based devices lead to a very complex problem in terms of finding the optimal solution. Moreover, the lack of data and delay-sensitive constraints prevent using more powerful techniques such as convex optimization or even Deep Learning (DL). These techniques are computationally expensive and require from a lot of training data. In the decentralized SR problem presented here, none of the requirements can be satisfied. To alleviate this, MABs attempt to maximize the achieved performance while the learning procedure is being carried out.

For the remainder of this work, we consider that the concepts of WLAN and agent can be indistinctly exchanged, since WLANs act as learners by collecting knowledge regarding their possible configurations and the experienced throughput. In practice, WLANs accumulate knowledge of a given selected action by observing its performance during a certain amount of time, i.e., a learning iteration. Consequently, the accuracy of long-term estimations depends on for how long the output of a given action is observed. The analysis of the necessary time to successfully monitor the channel is out of the scope of this paper. Thus, we assume perfect long-term estimations regarding the actions' performance. Furthermore, due to the lack of coordination between WLANs, the abovementioned learning procedure would be done in a disorganized way. Accordingly, from a global network perspective, agents would pick actions at any time within a learning iteration, since they are not synchronized in practice. However, and for the sake of simplicity, we consider that WLANs select an action at the beginning of each iteration, so that we can properly capture the performance associated with the different actions (recall that long-term estimates of actions are considered). Therefore, the moment at which adversarial agents select an action is irrelevant to our analysis.

Figure 4 illustrates the inclusion of agents into WLANs, which operate on top of CSMA/CA, as well as the aforementioned learning procedure (Figure 4(b)). As shown, both agents act within each learning iteration. Initially, an agent observes the performance of the WLAN, which depends on overall network configuration. With such an information, the agent updates the estimate of each action and selects a new one accordingly. This procedure is repeated at the beginning of a new iteration. For the scenario shown in Figure 4(a), there is an overlapping between the two WLANs during the initial iteration, and simultaneous transmissions cannot be held. According to this information, a new action is chosen by both Agents A and B, which turns out to enable SR, thus allowing a higher number of successful data transmissions.

Roughly, the SR problem in IEEE 802.11 WLANs can be modeled through adversarial MABs as follows:

- Let $\mathcal{N} = \{1, \dots, N\}$ be the set of potentially overlapping WLANs.
- Each WLAN can choose from a range of actions $\mathcal{A} = \{1, \dots, K\}$, which refer to combinations of \mathcal{C} non-overlapping frequency channels, \mathcal{P} transmit power levels, and \mathcal{S} sensitivity levels.
- Initially, the estimate reward of each action available in any WLAN, $k \in \{1, \dots, K\}$, is set to 0.
- At every iteration, each WLAN selects an arm randomly, according to its action selection-strategy, which in this work is considered to be Thompson sampling.
- After choosing an action k at iteration t , each WLAN observes the reward generated by the environment, $r_{k,t}$, which is based on the experienced throughput that depends on *i*) its own action and *ii*) the actions made by the overlapping WLANs.
- The new information is used for updating the knowledge on the available arms.

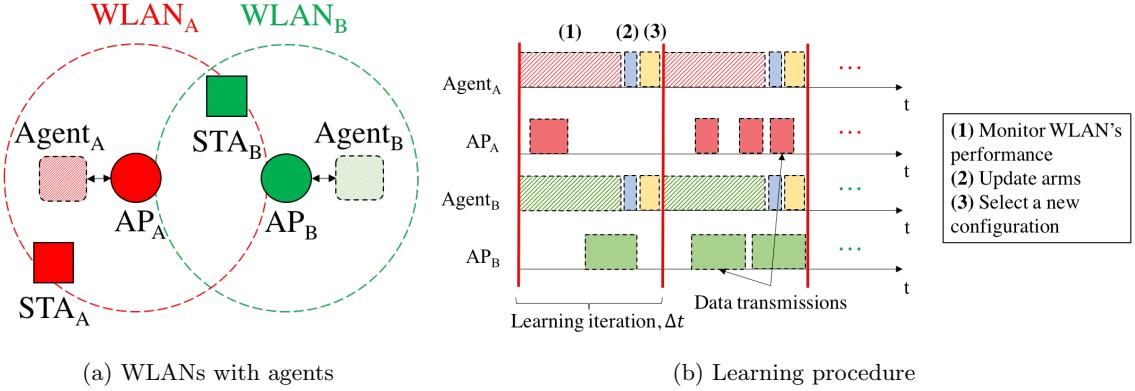


Figure 4: Agents integration in WLANs. (a) Scenario with two potentially overlapping WLANs, (b) Learning procedure followed by agents according to the performance observed in their associated WLAN.

The goal of an agent, then, is to maximize the reward function, which is equivalent to minimize the accumulated regret. In particular, the accumulated regret $R_{w,T}$ that a given WLAN w experiences until time T can be characterized as follows:

$$R_{w,T} = \sum_{t=1}^T (r_{w,t}^* - r_{w,t}),$$

where $r_{w,t}^*$ is the optimal reward granted by the best possible action in iteration t , and $r_{w,t}$ is the reward granted by the actual action chosen by WLAN w at that iteration. Since we face an adversarial setting, the process of minimizing the regret is highly influenced by the others' behavior. This raises concerns about the existence of an equilibrium in which the area throughput is fairly maximized.

For practical application of MABs in WLANs, the reward experienced by a given learner must be normalized, ideally, by the optimal reward $r_{w,t}^*$. This procedure is key to assess the potential of the played actions. But faced by the impossibility of providing such a value for every spatial distribution, which would require an exhaustive search, we define an upper bound consisting in the throughput that a given WLAN obtains in isolation (this concept is further described in Sections 4.1.1 and 4.1.2). Finally, it is important to remark that we consider time-invariant rewards, i.e., a given action always leads to the same reward.

4.1 Thompson Sampling

Thompson sampling has been employed in this work as the action-selection strategy used by WLANs. The motivation behind this choice is that Thompson sampling has been shown to grant excellent performance in front of other well-known policies such as UCB or EXP3, when applied into wireless networks. In (Wilhelmi et al., 2017), it was shown to converge fast to the optimal solution in terms of proportional fairness for adversarial environments, thus reducing the temporal variability observed for other exploration-exploitation mechanisms. Essentially, Thompson sampling (Thompson, 1933) is a Bayesian algorithm that constructs a probabilistic model of the rewards and assumes a prior distribution of the parameters of said model. Given the data collected during the learning procedure, Thompson sampling keeps track of the posterior distribution of the rewards and pulls arms randomly in a way that the drawing probability of each arm matches the probability of the particular arm being optimal. In practice, this is implemented by sampling the parameter corresponding to each arm from the posterior distribution, and pulling the arm yielding the maximal expected reward under the sampled parameter value.

For the sake of practicality, we apply Thompson Sampling using a Gaussian model for the rewards with a standard Gaussian prior as suggested in (Agrawal and Goyal, 2013). By standard calculations, it

can be verified that the posterior distribution of the rewards under this model is also Gaussian with mean $\hat{r}_k(t) = \frac{\sum_{w=1:k}^{t-1} r_k(w)}{n_k(t)+1}$ and variance $\sigma_k^2(t) = \frac{1}{n_k(t)+1}$, where $n_k(t)$ is the number of times that arm k was drawn until the beginning of round t . Henceforth, implementing Thompson sampling in MABs amounts to sampling a parameter θ_k from the Gaussian distribution $\mathcal{N}(\hat{r}_k(t), \sigma_k^2(t))$ and choosing the action k with the highest value. Our implementation of Thompson sampling to the WLAN problem is detailed in Algorithm 1.

Algorithm 1: Implementation of MABs (Thompson sampling) in a WLAN

```

1 Function Thompson Sampling ( $\mathcal{A}$ );
  Input :  $\mathcal{A}$ : set of possible actions in  $\{1, \dots, K\}$ 
2 initialize:  $t = 0$ , for each arm  $k \in \mathcal{A}$ , set  $\hat{r}_k = 0$  and  $n_k = 0$ 
3 while active do
4   For each arm  $k \in \mathcal{A}$ , sample  $\theta_k(t)$  from normal distribution  $\mathcal{N}(\hat{r}_k, \frac{1}{n_k+1})$ 
5   Play arm  $k = \operatorname{argmax}_{1,\dots,K} \theta_k(t)$ 
6   Observe the throughput experienced  $\Gamma_t$ 
7   Compute the reward  $r_{k,t}$ 
8    $\hat{r}_{k,t} \leftarrow \frac{\hat{r}_{k,t} n_{k,t} + r_{k,t}}{n_{k,t} + 2}$ 
9    $n_{k,t} \leftarrow n_{k,t} + 1$ 
10   $t \leftarrow t + 1$ 
11 end

```

In this paper, the reward is defined in two different ways, which are described in the following subsections.

4.1.1 Selfish Reward

The first considered reward aims to characterize a selfish behavior, which allows to purely represent the decentralized and adversarial SR problem. Through selfish learning, several WLANs attempt to learn the best configuration for their own gain, regardless of the performance experienced by neighboring networks. In fact, WLANs ignore the existence of other learners, which may have different goals. Henceforth, the reward $r_{w,t}$ that a given learner w experiences at iteration t is computed according to the throughput $\Gamma_{w,t}$ it experiences:

$$r_{w,t} = \frac{\Gamma_{w,t}}{\Gamma_w^*},$$

where Γ_w^* is a normalization value that refers to a certain upper bound reward that WLAN w can experience. In the selfish case, the optimal upper bound is given for any configuration that maximizes the individual performance of a given WLAN, regardless of the performance of other WLANs. It is important to remark that it may not be possible for the learner to know such an upper bound (further discussed in Section 4.2.1). In consequence, and for the rest of this paper, we define the upper bound reward to be the throughput that a given WLAN would obtain in isolation.

Selfish learning in WLANs has been shown to potentially increase SR while leading to collaborative results, provided that competitors enjoy equal opportunities (Wilhelmi et al., 2017). However, unfairness issues may be unleashed when dealing with significant asymmetries in terms of nodes location. As a result, WLANs in a dominant position may learn a performance maximization strategy at the expense of harming the weaker ones. By extension, competition among nodes is prone to lead to suboptimal configurations, so that the optimal action ends up being hidden to learners. In this sense, from the learner's point of view, the right action may not be robust enough against the environment, as a result of being susceptible to outer aggressive actions. Furthermore, learning selfishly in an adversarial environment may be detrimental in terms of temporal throughput variability.

4.1.2 Environment-Aware Reward

To overcome the unfairness situations that may be generated by selfish learning, we propose the environment-aware reward, which takes into consideration the effects that the actions of a given learner have on the environment (i.e., on the overlapping WLANs). To this end, we assume that WLANs are able to estimate the others' performance by listening to their activity on the channel. In practice, estimating the throughput experienced by overlapping WLANs may have limitations and lead to inaccurate values. Nevertheless, we assume perfect estimation to purely study the benefits and drawbacks of environment-aware learning. The analysis of dealing with inaccurate estimations is left as future work.

By assuming the availability of environmental information, we define an environment-aware reward that aims to fairly enhance the area throughput. Henceforth, rather than letting WLANs use their own performance, we propose that the reward experienced by each WLAN includes some notion of fairness. Three well-known fairness metrics are: *i*) Jain's Fairness Index (JFI) of the throughput, *ii*) Proportional Fairness (PF) of the throughput, and *iii*) max-min throughput. Throughout this paper, we are considering only the latter, since the JFI does not necessarily maximize aggregate performance, and the PF is very varying.⁸ As a result, the reward $r_{\mathcal{O},t}$ that a set of \mathcal{O} overlapping networks experience in iteration t is given by:

$$r_{\mathcal{O},t} = \frac{\min_{w \in \mathcal{O}} \Gamma_{w,t}}{\Gamma_{\mathcal{O},t}^*},$$

where $\min_{w \in \mathcal{O}} \Gamma_{w,t}$ is the minimum throughput experienced in the set of overlapping WLANs \mathcal{O} . The upper bound reward $\Gamma_{\mathcal{O},t}^*$ is shared, and refers to the configuration that grants the maximum max-min throughput. Again, since this knowledge may not be known at the learner side, we consider the set of throughputs in isolation for each WLAN in \mathcal{O} . Then, the max-min throughput value is taken as the shared optimal reward.

4.2 Considerations of Decentralized Learning in WLANs

The classical MAB problem frames the scenario whereby an agent interacts with the environment. The agent's goal is to maximize the long-term reward according to the actions it plays, regardless of any external factor. However, the presence of other agents in the adversarial learning problem adds an extra layer of complexity. That is the case of decentralized SR, where different potentially overlapping WLANs aim to find the best configuration by their own.

The competition unleashed by the adversarial setting can be formulated from a game theoretical perspective. It is important to concentrate on the possible equilibria that can be achieved for a given game, which can be defined by the set of competitors and their strategies. Of course, reaching an equilibrium whereby performance is maximized is limited to the conflicts that may crop up as a result of the clashing strategies followed by different players. For instance, the performance of a set of overlapping WLANs can be significantly limited if they use aggressive strategies in terms of interference generation, since a suboptimal equilibrium may be reached. In case of using the selfish reward presented in this paper, aggressive strategies in terms of interference would be preferred by learners, specially in dense environments. Note that WLANs seek to maximize their individual throughput, regardless the performance of the other networks. This particular scenario is further analyzed in Section 5.1.3. Moreover, it is possible that an equilibrium cannot be found in a decentralized manner. The main reason lies in the scarcity of the resources being shared, and in the individual requirements of each WLAN. In that case, if greedy strategies were employed, WLANs would alternate good and bad performing actions.

As a consequence to the adversarial setting unleashed in the decentralized SR problem, some important implications must be considered with regards to practical application of MABs to WLANs. Note, as well, that even if using an environment-aware reward that promotes collaboration (WLANs share a common goal), learning in a decentralized way may result into some other performance limitations. Such an issue is studied in Section 4.2.2. In essence, implications are noticed on the action-selection procedure, i.e., the set of rules

⁸Very different results may lead to the same (or very similar) PF value, which may have consequences on the learning procedure followed by WLANs. For instance, regarding the performance of two WLANs, a completely fair distribution of [50, 50] Mbps leads to a similar PF than a much more unfair distribution of [120, 20] Mbps.

and constraints according to which a given agent learns from the environment. Such a followed procedure is key to determine the potential of a given algorithm in terms of achievable performance and convergence guarantees. In the decentralized SR problem, the action selection procedure is held in a disorganized way, since every agent attempts to learn by its own. Such a situation leads to highly-varying environments in which an intensive action-selection procedure is held. This may severely impact on the learning process followed by any learner, and is worsened as the number of overlapping learners increases.

Regarding the learning process, on the one hand, a sublinear regret cannot always be guaranteed because of the intensive competition among networks. The speed at which regret is minimized strongly depends on the scenario. Because of the adversarial setting, a zero-regret configuration may be not be found, even if it exists. As a direct consequence, learners may suffer an increased variability on the experienced reward. Such a statement differs from the current work in multi-player MABs for opportunistic spectrum access, where strategies can be defined for sublinear regret minimization. First of all, unlike the SR problem, actions' performance can be binary modeled when attempting to access the channel, thus allowing to extract much more meaningful information regarding the environment: if selecting a given channel leads to a high number of collisions, the learner can easily infer that it is saturated. In the SR case, however, much more complex interactions may occur and have implications on a per-WLAN basis. Then, it is the aim of this work to provide insights on the application of decentralized learning to maximize the performance of a wireless network.

Finally, and related to regret minimization, assessing convergence in a WLAN (i.e., stop acting) may not be possible for the SR problem, thus impacting on the performance of higher communication layers. In particular, we can determine that a WLAN has learned which is the optimal action if it experiences a regret below a given threshold. However, due to the adversarial setting, such a condition may not hold, or may not be accomplished before the environment changes.

4.2.1 Reward Definition

A reward function describes how an agent should ideally behave, which allows conducting its activity towards maximizing (or minimizing) a given performance metric, i.e., the learner shapes a policy according to the obtained rewards. By extension, a precise definition of the reward allows to improve the learning procedure, since the reward perfectly matches with the desired goal. In that case, convergence can be improved, and the probability of falling into a local minimum is lower.

Unfortunately, defining a reward function in practice may become a very complex task. On the one hand, the optimal performance that can be achieved by a given individual or set of agents may not be known, thus hindering the learning procedure. On the other hand, the reward-based policies constructed by agents may be limited because of the competition among nodes (either for the selfish or the environment-aware approaches). Such dependencies can result in dominance positions of certain policies above others, which may obfuscate the optimal solution (obtained by the non-dominant policies). Moreover, reaching the optimal behavior is subject to the convexity of the joint reward function, which is not the case for the presented SR problem.

Now, in order to illustrate the impact of approximating the reward in the SR problem, let us consider a simple scenario (depicted in Figure 5(a)) and focus only on the selfish reward type (previously defined in Section 4.1.1). In particular, we place two WLANs that apply Thompson sampling selfishly. The range of possible actions in terms of CCA and transmit power levels are defined in Table 3 (included in A). We compare the usage of a reasonable upper bound reward (given a decentralized environment) in front of the optimal performance that can be actually achieved (computed by brute force). For the former, based on IEEE 802.11ax PHY capabilities (refer to simulation parameters in A), we use the theoretical data rate that can be achieved in case of using the maximum MCS.⁹ Note, as well, that this data rate may not correspond to the actual optimal performance due to several factors such as nodes position and inter-WLAN interactions. However, we refer to the utilization of a fixed MCS as an illustrative example of a practical upper bound that could be used in real networks. In contrast, we will use the throughput in isolation as an upper bound later in Section 5.

⁹We assume that the maximum data rate is achieved in case of using a single user (SU) transmission through a 1024-QAM MCS and a coding rate of 5/6. According to the IEEE 802.11ax standard parameters, this leads to a data rate of 114.37 Mbps.

Figures 5(b) and 5(c) show the experienced regret and throughput, respectively, experienced by two overlapping WLANs when applying Thompson sampling selfishly during 100 iterations. In order to emphasize on the effects of using an inaccurate upper bound, one of the WLANs (namely, WLAN_B) has stronger limitations than the other one (namely, WLAN_A), whose AP-STA distance is shorter. Such a situation makes WLAN_B more vulnerable in front of interference and prevents it to achieve the highest achievable throughput due to the SINR sensed at the receiver. As shown in Figure 5(c), WLAN_B experiences a higher throughput variability in case of using an approximated upper bound reward, rather than using the actual information for this concrete scenario. This can be also noticed in Figure 5(b), where the regret experienced by WLAN_B grows linearly if the actual optimal performance is unknown. In contrast, WLAN_A is able to use the maximum MCS due to its privileged situation, thus showing similar performance both for known and approximated upper bounds.

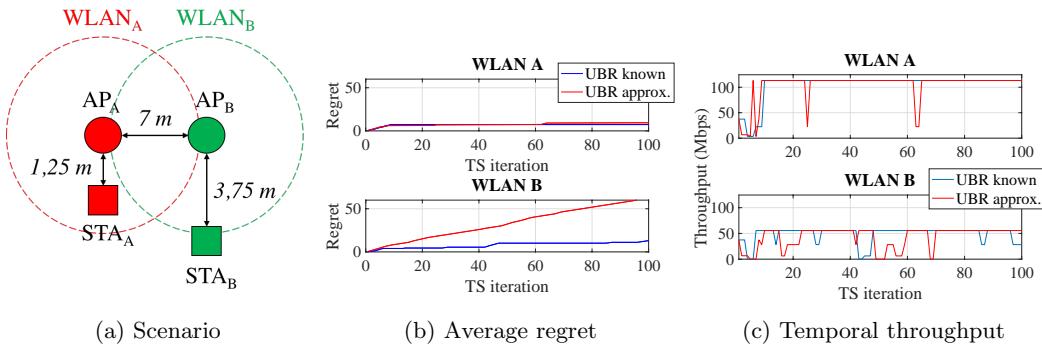


Figure 5: Upper bound reward considerations when applying selfish Thompson sampling (100 iterations are considered). (a) Scenario with two asymmetric WLANs in terms of maximum capacity, (b) Temporal regret experienced by each WLAN when the actual upper bound reward (UBR) is known (blue) or not (red), (c) Temporal throughput experienced by each WLAN when the actual UBR is known (blue) or not (red).

When defining an upper bound reward, we have seen that false expectations may lead to non-convergence, which may have severe implications in the temporal variability of the experienced performance.

4.2.2 Neighbors Identification when Applying an Environment-Aware Reward

In environment-aware learning, WLANs take the others' performance into account during the reward generation process. However, estimating the throughput of neighboring WLANs raises the following question: which are the potentially overlapping WLANs that each learner should consider? The fact of dealing with complex spatially-distributed environments hinders answering to that question, since interactions in overlapping WLANs are not trivial to be derived for the SR problem, and change with time. As a result, for a given learner, it is hard to identify the set of potentially overlapping WLANs whose performance must be taken into account. For convenience, let us refer to a particular set of overlapping networks as a cluster. In accordance to that, a WLAN applying clustering refers to the procedure whereby it considers the performance of other overlapping networks during the reward generation process. Note, as well, that an overlap between two networks may occur by different reasons. For instance, one may consider that two WLANs overlap if the mutual generated interference exceeds a given threshold, which may not necessarily be the capture threshold. For the rest of this paper, we assume that WLANs sharing a reward only take consideration of those generating a level of interference greater than the CCA threshold on their own. Furthermore, we assume bidirectional interactions, even in presence of asymmetries.

To showcase the importance of properly defining a list of neighbors (i.e., clusters), let us define a simple scenario in which two WLANs are independent to one another in terms of interference. Such a scenario has the particularity that one WLAN has limited performance due to the AP-STA distance. Therefore, we aim to study the effects of learning by either considering all the environment (long-range cluster) or just the

interfering devices (short-range cluster). On the one hand, we establish a soft establishment rule, where a neighbor is considered if the received power is higher than a very low decision threshold. In practice, this is equivalent to not considering any neighbors establishment rule, so that the max-min fairness involves all the WLANs in the presented scenario. On the other hand, short-range clustering by SINR is done (previously introduced in Section 4.1.2), which means that the performance of a given WLAN is considered by another one if the power received from the former is greater than the latter's CCA threshold.

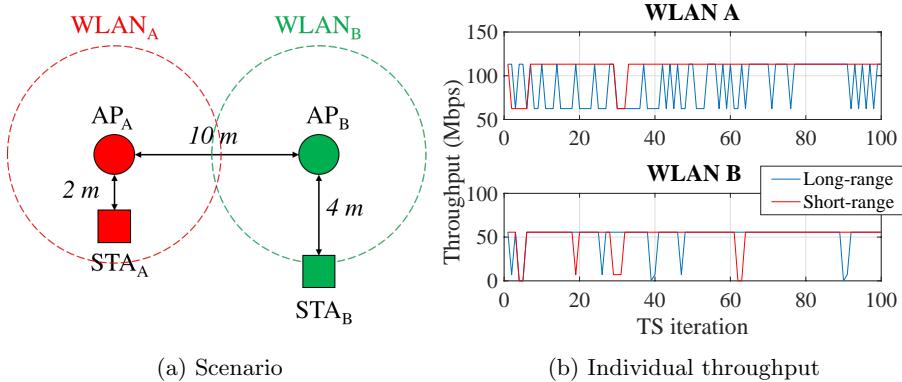


Figure 6: Neighbors establishment considerations when applying environment-aware Thompson sampling (100 iterations are considered). (a) Scenario with two independent WLANs in terms of interference, (b) Temporal throughput experienced by each WLAN when using long-range (blue) and short-range clustering (red).

As shown in Figure 6(b), short-range clustering grants better results in terms of temporal variability, since WLAN_A does not consider WLAN_B during the learning procedure (the CCA condition does not hold). Otherwise, in case that long-range clustering is applied, WLAN_A cannot determine that WLAN_B is not a potential overlapping network, i.e., the actions of the latter do not impact to performance of the former. If long-range clustering is applied, WLAN_A considers the throughput of WLAN_B to learn the performance of each action. This prevents the former to distinguish which are the best actions for itself. Therefore, good and bad performing actions are alternated because of the capacity limitation of WLAN_B (it never becomes satisfied).

Despite of the remarkable benefits of short-range clustering-based methods, determining neighbors lists is not trivial in dense WLAN scenarios. In particular, the proposed approach in which the SINR is used to determine interfering nodes fails in capturing additive interference situations. Such kind of interference appears when a given network is only affected when two or more WLANs transmit simultaneously. To illustrate this concept, we focus on the scenario shown in Figure 7(a), where additive interference generates flow-in-the-middle starvation to a WLAN located in the middle of the other two. In Figure 7(b) we show the results of applying environment-aware Thompson sampling for both short-range and long-range clustering. This time we have considered the results after 1,000 learning iterations, since we are interested in showing the long-term performance achieved in both situations.

As shown in Figure 7, the short-range clustering approach fails because additive interference affecting WLAN_B cannot be captured by the set of overlapping WLANs. Note, as well, that WLAN_A and WLAN_C always sense the channel free, regardless of the networks that are currently transmitting. This allows them to experience the highest possible throughput. In contrast, when agents consider all the WLANs in the environment (long-range clustering), the starvation at WLAN_B is noticed by the others. As a result, collaboration is enabled and the max-min throughput is maximized at the expense of the aggregate performance.

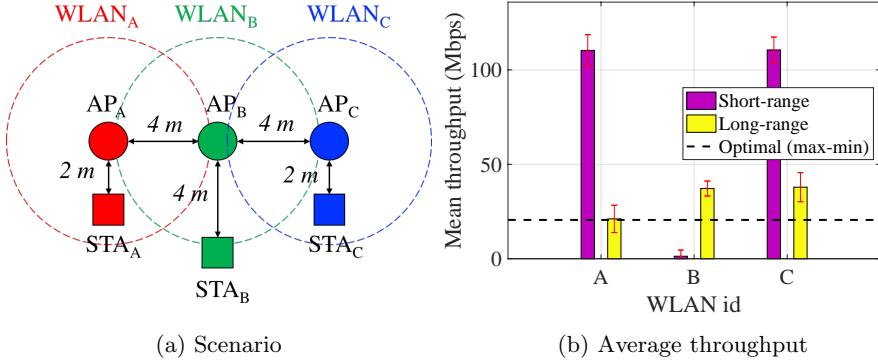


Figure 7: Issues on neighbors establishment when applying environment-aware Thompson sampling (1,000 iterations are considered). (a) Scenario in which WLAN_B is prone to suffer from flow starvation, (b) Average throughput per WLAN when using long-range (yellow) and short-range clustering (purple). The standard deviation of the average throughput between iterations is shown in red, and the black dashed line indicates the shared goal.

4.2.3 Learning in Dynamic WLANs

In addition to the abovementioned learning implications, it is interesting to study the effects of applying MABs in dynamic wireless environments. To that purpose, we frame a scenario in which a new WLAN appears after some other WLANs have already learned the optimal configuration. In particular, we use the scenario shown in Figure 7(a), and consider that WLAN_B is activated half-way through the simulation. Figure 8 shows the max-min throughput achieved when all the WLANs apply Thompson sampling (the environment-aware method is considered). WLAN_B is activated at iteration 500, point at which it is expected that WLAN_A and WLAN_C have gathered enough information for maximizing their performance during the initial phase.

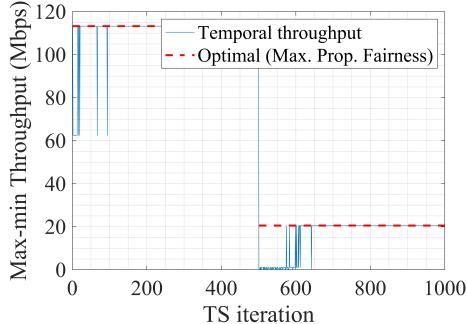


Figure 8: Thompson sampling application in a dynamic scenario where WLAN_B appears in iteration 500 (1,000 iterations are considered). The black dashed line indicates the shared goal.

As shown in Figure 8, when WLAN_B is activated, Thompson sampling adapts to the new situation. It reaches the new optimal goal after a reasonable number of iterations (similar to the initial learning phase). In particular, Thompson sampling needs some time to reshape the already defined probability distributions. The time the algorithm takes to adapt itself can be enhanced if changes in the environment are tracked (e.g., by sensing beacons from new WLANs). Hence, additional information provided to the learning algorithm may boost the exploration of the new optimal actions. In contrast, the implications in terms of performance

provoked by changes in the environment are hard to track in practice. Moreover, the procedure to be followed when a change is detected at the algorithm level is not easy to derive. There exists a trade-off between the past and the present information, which must be carefully balanced to achieve optimal performance.

In summary, learning in dynamic environments raises several questions about the validity and expiration of the learned data. This analysis is out of the scope of this paper, so we leave it as future work. In anticipation of this research topic, we highlight that dynamic MABs have been previously studied in (Hartland et al., 2006). Also, with a higher degree of relation to the algorithms shown in this work, we find Dynamic Thompson Sampling (DTS) (Gupta et al., 2011), which is shown to adapt faster to changes in the environment than the traditional Thompson sampling. Roughly, DTS promotes adaptive exploration by tracking the reward probabilities of each arm, which is useful to give emphasis to recent observations.

5 Performance Evaluation

In this Section we evaluate both selfish and environment-aware decentralized learning strategies. To that purpose, we first study the behavior shown by WLANs in representative scenarios when applying both kinds of learning. Then, we generalize those results through simulations in random high-density scenarios.

5.1 Selfish vs Environment-Aware Learning

Selfish and environment-aware strategies are now evaluated in scenarios describing different casuistic. In order to assess the performance achieved in WLANs by applying each strategy, we define the optimal result as: *i*) the maximum individual throughput that a given WLAN can achieve (regardless of the others' performance), *ii*) the maximum throughput that each WLAN can achieve by ensuring the max-min principle. Note, as well, that such values are computed by brute force in the following illustrative scenarios. It is also worth noting that such an optimal performance may not be achieved due to the interactions between WLANs, but special attention will be given to the behavior of each learning approach in relation to that.

For the sake of highlighting Thompson sampling performance in front of other online learning techniques, in this subsection we also provide the results of applying the ε -greedy action-selection strategy. In contrast to Thompson sampling, ε -greedy selects the action with highest absolute performance with probability $1 - \varepsilon$ (exploitation), where ε is within 0 and 1. Otherwise, a random action is selected with probability ε (exploration). Note, as well, that the parameter ε is initialized to 1 and dynamically adjusted as a function of the number of iterations, as done in (Auer et al., 2002).

5.1.1 Learning in Presence of Asymmetries

Wireless networks are not always symmetric in terms of nodes location, so that different WLANs may not enjoy the same opportunities when tackling the environment. Such an issue is more common to occur in dense environments where the diversity of deployments is high. In these situations, attempting to maximize spectral efficiency in a selfish way can be detrimental in terms of fairness, especially if there are WLANs in worse conditions than others. Conversely, the environment-aware approach is expected to solve the imbalance between WLANs by maximizing the max-min throughput.

To illustrate the effect of applying both selfish and environment-aware rewards in an asymmetric network, let us retrieve the simple 2-WLANs asymmetric scenario used in Section 4.2.1 (now shown in Figure 9(a)). In this scenario, each AP is separated d_{AP_A,AP_B} meters from the other one. The distance between an AP and its associated STA is d_{AP_A,STA_A} and d_{AP_B,STA_B} , respectively, so that $d_{AP_A,AP_B} > d_{AP_B,STA_B} > d_{AP_A,STA_A}$.

The results of applying both Thompson sampling and ε -greedy methods for 10,000 iterations are shown in Figures 9(b) and 9(c). While the former indicates the average throughput experienced by each WLAN, the latter illustrates the temporal variability at WLAN_A.¹⁰

As Figure 9(b) reveals, learning selfishly allows WLAN_A to experience the highest possible throughput, both using Thompson sampling and ε -greedy. However, WLAN_B suffers from starvation because none of

¹⁰We show the performance of only one WLAN as it is representative of the effects we illustrate here.

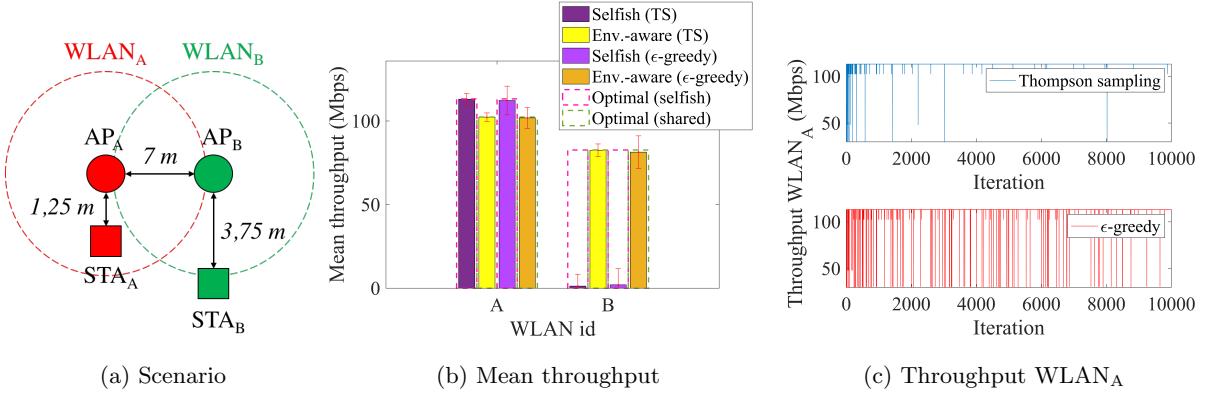


Figure 9: Fairness issues in both selfish and environment-aware Thompson sampling (10,000 iterations are considered). (a) Scenario in which WLAN_B is prone to suffer from starvation, (b) Average throughput per WLAN for selfish and environment-aware learning, both for Thompson sampling (TS) and ϵ -greedy. The standard deviation of the average throughput between iterations is shown in red. The pink and green dashed lines indicate the maximum performance achieved per WLAN regarding both selfish and environment-aware optimal solutions, respectively. (c) Temporal throughput achieved by WLAN_A when learning selfishly through Thompson sampling and ϵ -greedy, respectively.

its possible actions allows to palliate the effects of WLAN_A's aggressive configuration. In contrast, when both WLANs use the environment-aware strategy, the optimal max-min throughput is achieved, so that the starvation problem in WLAN_B is solved. In exchange, WLAN_A sacrifices a portion of its maximum achievable throughput, since it uses a less aggressive configuration. The difference between Thompson sampling and ϵ -greedy lies in the experienced temporal throughput variability, which is significantly higher for the latter method (refer to Figure 9(c)). Such a variability entails that WLANs experience a slightly lower mean throughput.

As shown in this subsection, selfish learning is prone to generate flow starvation. However, it is worth noting that it is a very common situation in real dense deployments, even if configurations remain static.

5.1.2 Learning on Equal Terms

We previously analyzed the effect of applying selfish and environment-aware learning in an asymmetric deployment. However, that might not represent other topologies where competing WLANs are in similar conditions. Therefore, we now showcase the potential of applying RL in dense scenarios where WLANs can access to the channel on equal terms.

For that, we consider a symmetric grid formed by 4 WLANs (Figure 10(a)), which can choose from the same range of CST and transmit power levels. In this scenario, there exists an optimal configuration that can be reached by each WLAN, regardless of the others' actions. Therefore, in case all the WLANs discover the optimal action, a Nash Equilibrium is conformed, so that no individual can obtain further benefits by deviating from its strategy. As previously done, we focus on the average throughput achieved by each WLAN (Figure 10(b)), and the temporal throughput in WLAN_A (Figure 10(c)).

Our results show that all the WLANs are able to rapidly find the configuration that grants the maximum possible throughput, for each of the proposed learning methods. A collaborative behavior between WLANs occur despite learning selfishly. The primary reason of such a collaboration lies in the symmetries found in the scenario, and on the ability of each WLAN to compete for resources in a fair manner. Finally, and similarly to what is shown in Subsection 5.1.1, ϵ -greedy is shown to lead to a significantly higher variability than Thompson sampling.

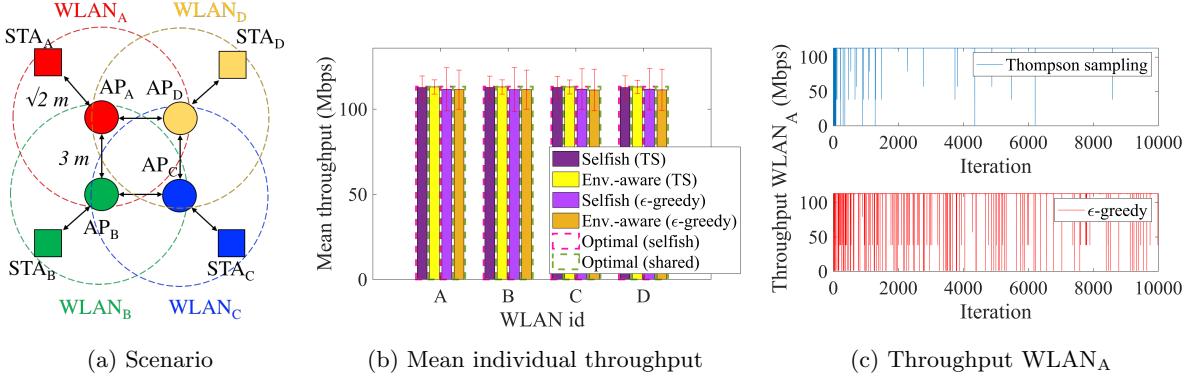


Figure 10: Potential of both selfish and environment-aware Thompson sampling (10,000 iterations are considered). (a) Scenario in which STAs are placed conservatively regarding inter-WLAN interference, (b) Average throughput per WLAN for Selfish and Environment-aware learning, both for Thompson sampling (TS) and ϵ -greedy. The standard deviation of the average throughput between iterations is shown in red. The pink and green dashed lines indicate the maximum performance achieved per WLAN regarding both selfish and environment-aware optimal solutions, respectively. (c) Temporal throughput achieved by WLAN_A when learning selfishly through Thompson sampling and ϵ -greedy, respectively.

5.1.3 Competition Effects

The scenario shown in Section 5.1.2 frames a conservative environment in which the inter-WLAN interference is low, i.e., APs belonging to different WLANs are distant enough, and STAs are reasonably close to their AP. However, if we refer to a less idyllic situation, applying RL may not be as effective as before. In particular, we are interested in showing the effects of using both selfish and environment-aware strategies in highly competitive environments. In those cases, it may happen that the optimal global solution becomes obfuscated because the action-selection procedure is held individually.

To show the implications of intensive competition between WLANs, let us propose the nodes distribution shown in Figure 11(a). In this scenario, interactions between WLANs are more prone to generate performance issues, since all the STAs are more exposed to inter-WLAN interference. In particular, the optimal solution for both individual performance and max-min throughput is obtained only if all the WLANs use the minimum transmit power and the maximum sensitivity. The results of applying both selfish and environment-aware strategies are shown in Figures 11(b) and 11(c).

As it can be observed from Figure 11(b), none of the WLANs is able to reach the optimum performance, neither for the selfish nor the environment-aware reward. We identify the fact that actions are selected individually as the main cause of such a performance inefficiency.

Regarding selfish learning, WLANs choosing the optimal configuration are more susceptible to be affected by inter-WLAN interference (asymmetries between WLANs are generated). First of all, since the optimal configuration entails using the minimum transmit power, the generated interference is minimized. As a result, the rest of WLANs can properly operate on the channel (they sense it free). However, these same WLANs cannot distinguish between the right and the harmful action from the global perspective, since both options lead to the optimal individual throughput (at the expense of harming the WLAN that is behaving properly). Henceforth, selfish WLANs are prone to act aggressively (i.e., use a high transmit power and limit the sensitivity area) in high-interference situations, which leads to obfuscate the optimal solution (even in terms of selfishness). In short, acting selfishly in this kind of scenario is not as effective as providing a certain level of collaboration that allows to identify the optimal global configuration.

Moreover, and similarly to the selfish approach, using an environment-aware metric is not enough to properly maximize the spectral efficiency. Despite WLANs act according to a joint reward, the same limitations occur due to the weakness of optimal actions in front of the environment. When a given WLAN

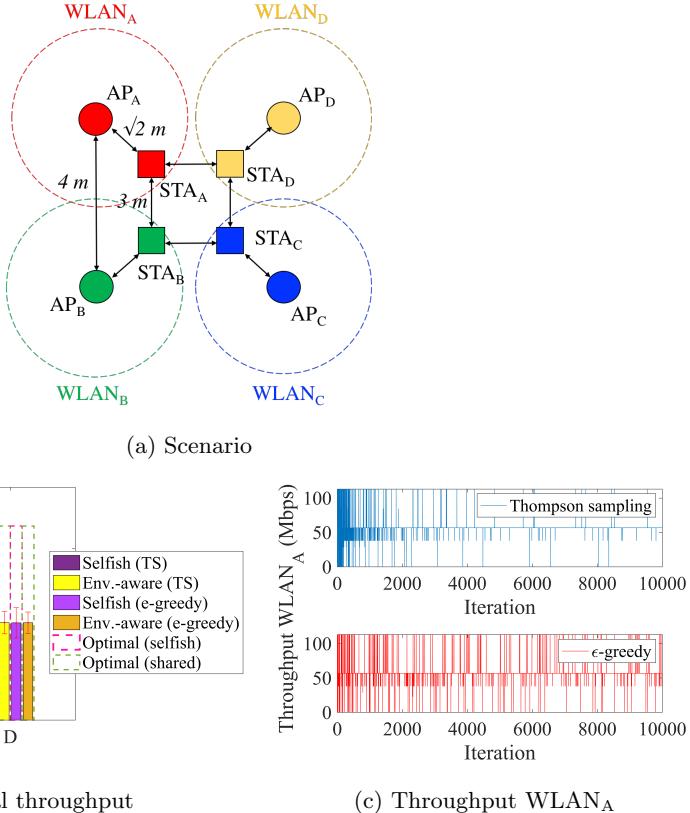


Figure 11: Competition issues in both selfish and environment-aware Thompson sampling (10,000 iterations are considered). (a) Scenario in which STAs are placed in a greedy way regarding inter-WLAN interference, (b) Average throughput per WLAN for Selfish and Environment-aware learning, both for Thompson sampling (TS) and ϵ -greedy. The standard deviation of the average throughput between iterations is shown in red. The pink and green dashed lines indicate the maximum performance achieved per WLAN regarding both selfish and environment-aware optimal solutions, respectively. (c) Temporal throughput achieved by WLAN_A when learning selfishly through Thompson sampling and ϵ -greedy, respectively.

selects the optimal action, probably it would not obtain the highest possible reward, since it is subject to the others' configuration. Since WLANs learn independently (even if the others' throughput is considered), the probabilities for all to choose the optimal action are very low. In consequence, the learning capacity is limited as for the selfish approach. To conclude, Figure 11(c) shows that Thompson sampling is more stable than ϵ -greedy, in terms of temporal throughput variability.

5.2 Random Scenarios

In order to further analyze the effects of applying both selfish and environment-aware strategies, we propose using 50 random scenarios, containing $N = \{2, 4, 6, 8\}$ WLANs in a $10 \times 10 \times 5$ m area (i.e., an AP every 250, 125, 83.33 and 62.5 m³, respectively). WLANs are uniformly randomly distributed in the scenario, as well as STAs are randomly located between 1 and 3 meters away from their AP. Configurations are assigned so that WLANs use the same channel by default, and maximum sensitivity and transmit power. Such a configuration has been previously shown to be common in real deployments (Akella et al., 2007). Further details regarding the generation of random scenarios can be found in A.

Unlike in previous results, we now consider applying MABs for only 500 iterations. The main reason lies

in showing the gains that can be achieved by applying MABs for short periods, i.e., before the environment significantly changes. Note that requiring large periods of time for reaching an equilibrium may not be feasible in real wireless deployments, because of the channel and users variability.

We first show the average results obtained by each approach in Figure 12, which are compared to the static situation. The latter considers that WLANs use the initial assigned configuration. In addition, all the WLANs use the same channel (namely, channel 1). For performance evaluation, we focus on the average throughput, the max-min throughput and the JFI. Moreover, due to the impossibility of using the actual upper bound reward for each configuration, we use *i*) the throughput in isolation as a maximum performance reference for the selfish strategy, and *ii*) for the environment-aware strategy, the minimum throughput noticed among the individual performances of the potentially overlapping WLANs, so that their throughput in isolation is considered.

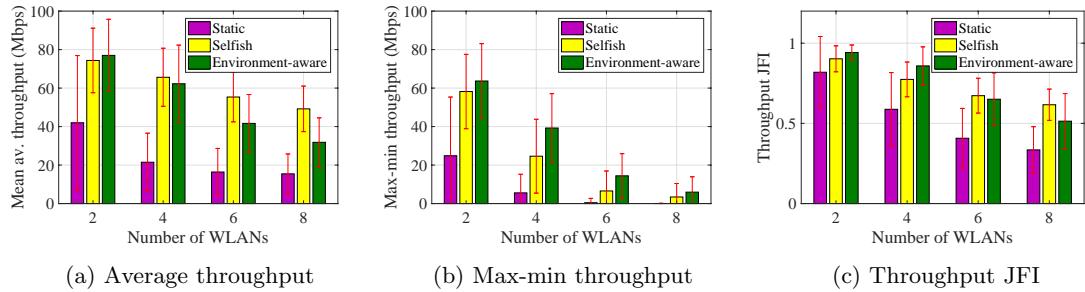


Figure 12: Average results of applying 500-iteration Thompson sampling in 50 random scenarios for $\{2, 4, 6, 8\}$ potentially overlapping WLANs. (a) Mean average throughput with standard deviation (in red) per WLAN in each scenario when using static (purple), selfish learning (yellow) or environment-aware learning (green), (b) Mean average max-min throughput with standard deviation (in red) per WLAN in each scenario when using static (purple), selfish learning (yellow) or environment-aware learning (green), (c) Mean JFI with standard deviation (in red) in each scenario when using static (purple), selfish learning (yellow) or environment-aware learning (green)

As shown, the average throughput obtained per scenario through selfish Thompson sampling outperforms the static configuration (refer to Figure 12(a)), which is evidence of the poor spectral efficiency achieved in current deployments. In all the cases, using MABs allows to maximize the static performance. In addition, we can observe that selfish learning grants higher throughput than the environment-aware as density increases. Regarding fairness, the selfish approach is shown to work better in average (refer to Figure 12(c)), because WLANs in a bad situation are able to self-adjust themselves in a competitive environment. However, this is not directly related to the max-min throughput, which is the goal of the environment-aware approach (refer to Figure 12(b)). Unfortunately, guaranteeing a certain minimum throughput to the less privileged WLANs in terms of interference becomes more challenging as the number of overlapping nodes increases. Such an issue is highly conditioned by the distance between the AP and the STA of a given WLAN.

Finally, to further illustrate the enhancements achieved by applying learning, we plot the average throughput obtained for different learning phases in Figure 13. By showing the performance experienced for each interval of 100 iterations, we aim to emphasize on the progressive gains achieved along the learning operation. As previously mentioned, wireless environments are highly varying, thus a fast convergence is essential for any learning algorithm.

For both strategies, we observe a big gain experienced during the first two intervals, which becomes stable from that point onwards. Thus, even if an equilibrium is not reached, a significant increase of the average performance is rapidly experienced. Regarding environment-aware learning, a greater enhancement is provided when density is low (two and four WLANs). However, when density increases, selfish learning achieves a higher average performance earlier. The key reason lies in the fact that max-min throughput is more likely to be low as the number of overlapping devices increases. Therefore, the learning procedure is

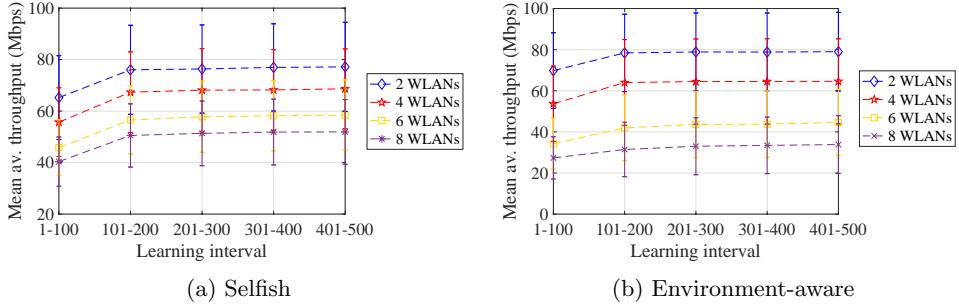


Figure 13: Mean average throughput with standard deviation achieved during specific intervals. 500 iterations are considered for both selfish and environment-aware Thompson sampling in 50 random scenarios for $\{2, 4, 6, 8\}$ overlapping WLANs. (a) Results for the selfish strategy, (b) Results for the environment-aware strategy.

slowed down due to the impossibility of finding an appropriate solution that alleviates the poor performance achieved by the most vulnerable WLANs. In opposite, learning selfishly speeds up performance trading fairness off.

6 Conclusions

In this work, we addressed the potential and feasibility of applying decentralized online learning to wireless networks, as a contribution to the debate about whether future WLANs should remain decentralized or evolve towards centralized mechanisms (such as in cellular networks). To that purpose, we delved into the SR problem in IEEE 802.11 WLANs and presented a practical learning-based application to overcome it. In particular, we modeled the problem through MABs and showed two strategies based on the Thompson sampling action-selection method. The first one is based on learning selfishly, where the reward that an agent obtains after playing an action is granted according to its own throughput, regardless of the performance achieved by the overlapping WLANs. The second one, referred to as environment-aware learning, quantifies how good actions are based on the max-min throughput achieved in the network. This can be done by inferring the performance obtained by the surrounding WLANs.

By using the SR problem as a guiding thread, we analyzed the main considerations that must be done when applying decentralized learning methods to wireless communications problems. Among them, we highlighted practical issues such as convergence assessment or the difficulties on developing an appropriate reward generation system. Finally, we evaluated two learning policies in terms of fairness and throughput, so as to show their potential and major implications. Despite learning selfishly has been shown to generate unfair situations, its potential at maximizing the aggregate performance is very promising in certain scenarios. Moreover, even if environment-aware methods allow to solve fairness-related issues, the fact of learning in a decentralized way is not a guarantee for finding the best-performing configuration. In addition, environment-aware learning may severely limit the aggregate performance in benefit of few WLANs.

As a final conclusion, we remark the potential of applying uncoordinated MABs in dense WLANs, thus bringing hope for decentralized deployments in front of centralized systems. However, for practical application, such a kind of mechanisms are required to take the environment into consideration, since selfish approaches are prone to generate unfair situations. Therefore, other important challenges such as inter-WLANs communication must be overcome. Moreover, the utilization of collaborative approaches raises several questions regarding fairness ascertainment. For instance, is it worth to significantly reduce the performance of many WLANs in benefit to less privileged ones in terms of location?

Future work will also consider the use of beamforming to improve spatial reuse. By defining multiple beamforming sectors (i.e., 4 or 8), multiple simultaneous transmissions can be performed from different

APs if they select non-interfering sectors for transmitting. Thus, it adds another degree of freedom which combined with transmission power, CCA, and channel allocation adaptation, which may be done per sector, could further contribute to improve the overall system performance.

Acknowledgment

This work has been partially supported by the Spanish Ministry of Economy and Competitiveness under the Maria de Maeztu Units of Excellence Programme (MDM-2015-0502), by the Catalan Government SGR grant for research support (2017-SGR-1188), by the European Regional Development Fund under grant TEC2015-71303-R (MINECO/FEDER), and by a Gift from the Cisco University Research Program (CG#890107, Towards Deterministic Channel Access in High-Density WLANs) Fund, a corporate advised fund of Silicon Valley Community Foundation.

The authors would like to thank the anonymous reviewers, which thorough work and insightful comments were of great help to improve this paper.

Appendices

A Wireless Environment

Here we provide details on the wireless environment used to simulate IEEE 802.11 WLANs behavior. First, physical medium effects are modeled by following the specification provided in the IEEE 802.11ax standard for residential scenarios (Merlin et al., 2015), which includes specific path-loss and shadowing models. We have chosen this scenario because it is very representative for next-generation dense and chaotic deployments. In particular, since we refer to random scenarios, we capture the essence of the 11ax residential scenario by proposing a model that takes into account the walls and floor frequencies, rather than the actual location of walls and floors. Accordingly, the power loss PL_d in such an environment is given by:

$$PL_d = 40.05 + 20 \log_{10} \left(\frac{f_c}{2.4} \right) + 20 \log_{10}(\min(d, 5)) + I_{d>5} \cdot 35 \log_{10} \left(\frac{d}{5} \right) + 18.3F^{\frac{F+2}{F+1}-0.46} + 5W$$

where f_c is the frequency in GHz, d is the distance between the transmitter and the receiver in meters, and F and W are the average number of floors and walls traversed per meter, respectively. The original scenario considers a 5-floor building, with twenty apartments of $10 \times 10 \times 3$ meters size per floor. Regarding adjacent channel interference, we consider that consecutive channels are non-overlapping.

Note, as well, that the data rate at which a transmitter sends data is subject to the signal strength in the receiver, which in this work is assumed to be known. IEEE 802.11ax parameters are used, so that modulations range from BPSK to 1024-QAM (iee, 2015). Table 3 details the parameters used, which include PHY and MAC specifications (tga, September 2017).

Parameter	Description	Value
\mathcal{C}	Set of channels	1 / 2
\mathcal{S}	Set of sensitivity thresholds	-68 dBm / -90 dBm
\mathcal{T}	Set of transmit power values	5 dBm / 20 dBm
$d_{AP,STA}^{min} / d_{AP,STA}^{max}$	Min/max distance AP - STA	1 m / 3 m
(x, y, z)	3D map dimensions in each axis	(10, 10, 5) m
W	Channel bandwidth	20 MHz
F	Central frequency	5 GHz
SUSS	Spatial streams per user	1
G_{tx}	Transmitting gain	0 dBi
G_{rx}	Reception gain	0 dBi
N	Floor noise level	-95 dBm
CE	Capture Effect threshold	10 dBm
T_s	Symbol duration	9 μ s
DIFS/SIFS	DIFS and SIFS duration	34 μ s / 16 μ s
CW _{min} /CW _{max}	Min/max contention window	16 / 16
N_{agg}	Number of packets aggregated	64
L_{DATA}	Length of a data packet	12000 bits
L_{RTS} / L_{CTS}	Length RTS and CTS packets	160 bits / 112 bits
L_{MAC}	Length MAC header	272 bits
L_{SF}	Length Service Field (SF)	16 bits
L_{MPDU}	MPDU delimiter	32 bits
L_{Tail}	Length tail	6 bits
L_{BACK}	Length block ACK	240 bits
T_{RTS}	RTS packet duration	$20 \cdot 10^{-6} + \frac{L_{SF}+L_{RTS}+L_{Tail}}{R} T_s$ s
T_{CTS}	CTS packet duration	$20 \cdot 10^{-6} + \frac{L_{SF}+L_{CTS}+L_{Tail}}{R} T_s$ s
T_{DATA}	Data packet duration	$36 \cdot 10^{-6} + SUSS \cdot 16 \cdot 10^{-6} + \frac{(L_{SF}+N_{agg} \cdot (304+L_{DATA})+L_{Tail})}{R} T_s$ s
T_{BACK}	Block ACK duration	$20 \cdot 10^{-6} + \frac{L_{SF}+L_{BACK}+L_{Tail}}{R} T_s$ s
\mathcal{T}	Traffic model	Full buffer (downlink)

Table 3: Simulation parameters

References

References

- IEEE 802.11-15/1070r4 - 1024 QAM proposal. *IEEE*, 802:802–11, 2015.
- IEEE p802.11ax/d2.0, September 2017.
- S. Agrawal and N. Goyal. Further optimal regret bounds for thompson sampling. In *Artificial Intelligence and Statistics*, pages 99–107, 2013.
- A. Akella, G. Judd, S. Seshan, and P. Steenkiste. Self-management in chaotic wireless deployments. *Wireless Networks*, 13(6):737–755, 2007.
- A. Anandkumar, N. Michael, A. K. Tang, and A. Swami. Distributed algorithms for learning and cognitive medium access with logarithmic regret. *IEEE Journal on Selected Areas in Communications*, 29(4):731–745, 2011.
- J.-Y. Audibert, R. Munos, and C. Szepesvári. Exploration–exploitation tradeoff using variance estimates in multi-armed bandits. *Theoretical Computer Science*, 410(19):1876–1902, 2009.
- P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire. Gambling in a rigged casino: The adversarial multi-armed bandit problem. In *Foundations of Computer Science, 1995. Proceedings., 36th Annual Symposium on*, pages 322–331. IEEE, 1995.

- P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2-3):235–256, 2002.
- S. Barrachina-Muñoz, F. Wilhelmi, and B. Bellalta. To overlap or not to overlap: Enabling channel bonding in high density wlans. *arXiv preprint arXiv:1803.09112*, 2018a.
- S. Barrachina-Muñoz, F. Wilhelmi, and B. Bellalta. Performance analysis of dynamic channel bonding in spatially distributed high density wlans. *arXiv preprint arXiv:1801.00594*, 2018b.
- B. Bellalta. IEEE 802.11 ax: High-efficiency WLANs. *IEEE Wireless Communications*, 23(1):38–46, 2016.
- B. Bellalta. Throughput analysis in high density wlans. *IEEE Communications Letters*, 21(3):592–595, 2017.
- M. Bennis and D. Niyato. A Q-learning based approach to interference avoidance in self-organized femtocell networks. In *GLOBECOM Workshops (GC Wkshps), 2010 IEEE*, pages 706–710. IEEE, 2010.
- M. Bennis, S. Guruacharya, and D. Niyato. Distributed learning strategies for interference mitigation in femtocell networks. In *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*, pages 1–5. IEEE, 2011.
- L. Chen. A distributed access point selection algorithm based on no-regret learning for wireless access networks. In *Vehicular Technology Conference (VTC 2010-Spring), 2010 IEEE 71st*, pages 1–5. IEEE, 2010.
- M. A. Ergin, K. Ramachandran, and M. Gruteser. Understanding the effect of access point density on wireless LAN performance. In *Proceedings of the 13th annual ACM international conference on Mobile computing and networking*, pages 350–353. ACM, 2007.
- N. Gupta, O.-C. Granmo, and A. Agrawala. Thompson sampling for dynamic multi-armed bandits. In *2011 10th International Conference on Machine Learning and Applications Workshops*, pages 484–489. IEEE, 2011.
- C. Hartland, S. Gelly, N. Baskiotis, O. Teytaud, and M. Sebag. Multi-armed bandit, dynamic environments and meta-bandits. 2006.
- H. Li. Multi-agent q-learning of channel selection in multi-user cognitive radio systems: A two by two case. In *Systems, Man and Cybernetics, 2009. SMC 2009. IEEE International Conference on*, pages 1893–1898. IEEE, 2009.
- K. Liu and Q. Zhao. Distributed learning in multi-armed bandit with multiple players. *IEEE Transactions on Signal Processing*, 58(11):5667–5681, 2010.
- S. Maghsudi and S. Stańczak. Channel selection for network-assisted D2D communication via no-regret bandit learning with calibrated forecasting. *IEEE Transactions on Wireless Communications*, 14(3):1309–1322, 2015a.
- S. Maghsudi and S. Stańczak. Joint channel selection and power control in infrastructureless wireless networks: A multiplayer multiarmed bandit framework. *IEEE Transactions on Vehicular Technology*, 64(10):4565–4578, 2015b.
- S. Merlin, G. Barriac, H. Sampath, L. Cariou, T. Derham, J. Le Rouzic, R. Stacey, M. Park, R. Porat, N. Jindal, et al. Tgax simulation scenarios. *doc.: IEEE*, pages 802–11, 2015.
- J. Nie and S. Haykin. A Q-learning-based dynamic channel assignment technique for mobile communication systems. *IEEE Transactions on Vehicular Technology*, 48(5):1676–1687, 1999.
- J. Rosenski, O. Shamir, and L. Szlak. Multi-player bandits—a musical chairs approach. In *International Conference on Machine Learning*, pages 155–163, 2016.

- N. Rupasinghe and İ. Güvenç. Reinforcement learning for licensed-assisted access of LTE in the unlicensed spectrum. In *Wireless Communications and Networking Conference (WCNC), 2015 IEEE*, pages 1279–1284. IEEE, 2015.
- O. Sallent, J. Pérez-Romero, R. Ferrús, and R. Agustí. Learning-based coexistence for LTE operation in unlicensed bands. In *International Conference on Communication Workshop (ICCW)*, pages 2307–2313. IEEE, 2015.
- S. L. Scott. A modern bayesian look at the multi-armed bandit. *Applied Stochastic Models in Business and Industry*, 26(6):639–658, 2010.
- G. Smith. Dynamic sensitivity control-v2. *IEEE*, 802:802–11, 2015.
- W. R. Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294, 1933.
- F. Wilhelmi. Potential and pitfalls of multi-armed bandits for decentralized spatial reuse in wlans. https://github.com/fwilhelmi/potential_pitfalls_mabs_spatial_reuse
Commit: e761e6e, 2018.
- F. Wilhelmi, B. Bellalta, C. Cano, A. Jonsson, G. Neu, and S. Barrachina-Muñoz. Collaborative spatial reuse in wireless networks via selfish bandits. *arXiv preprint arXiv:1705.10508*, 2017.

Komondor: a Wireless Network Simulator for Next-Generation High-Density WLANs

Sergio Barrachina-Muñoz, Francesc Wilhelmi*,
Ioannis Selinis, and Boris Bellalta

Abstract

Komondor is a wireless network simulator for next-generation wireless local area networks (WLANs). The simulator has been conceived as an accessible (ready-to-use) open source tool for research on wireless networks and academia. An important advantage of Komondor over other well-known wireless simulators lies in its high event processing rate, which is furnished by the simplification of the core operation. This allows outperforming the execution time of other simulators like ns-3, thus supporting large-scale scenarios with a huge number of nodes. In this paper, we provide insights into the Komondor simulator and overview its main features, development stages and use cases. The operation of Komondor is validated in a variety of scenarios against different tools: the ns-3 simulator and two analytical tools based on Continuous Time Markov Networks (CTMNs) and the Bianchi's DCF model. Results show that Komondor captures the IEEE 802.11 operation very similarly to ns-3. Finally, we discuss the potential of Komondor for simulating complex environments – even with machine learning support – in next-generation WLANs by easily developing new user-defined modules of code.

1 Introduction

The Institute of Electrical and Electronics Engineers (IEEE) 802.11 Wireless Local Area Networks (WLANs) are evolving fast to satisfy the new strict requirements in terms of data rate and user density. In particular, various IEEE 802.11 amendments have been introduced in the past few years or are under active development to accommodate the need for higher capacity, exponential growth in number of devices, and novel use-cases. [1]. An example of next-generation high-density deployment is depicted in Fig. 1 where multiple WLANs are allocated with different channels and dynamic channel bonding (DCB) policies.

Of particular interest is the IEEE 802.11ax (11ax) amendment [2], that is under active development and which was introduced to address the demands and challenges that WLANs will face in the congested 2.4/5 GHz bands [3].

*The contribution of the first two authors is the same.

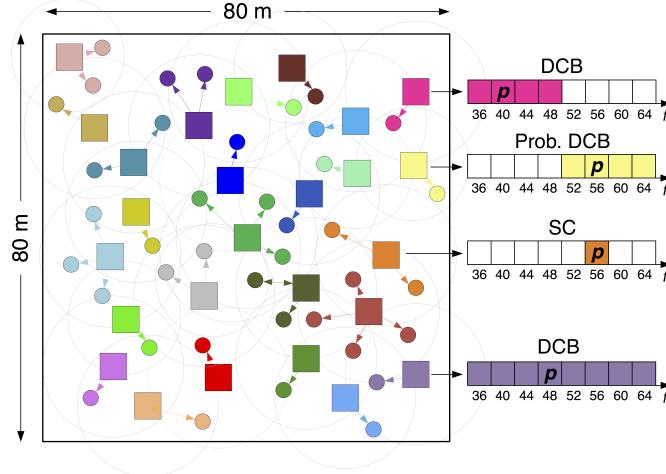


Figure 1: Dense scenario composed of 25 WLANs. Note that each WLAN has its own channel allocation and DCB policy.

Other important amendments for next-generation wireless networks are the IEEE 802.11ay [4] and EXtreme Throughput (XT) 802.11 [5], which aim to exploit the 60 GHz and ≤ 6 GHz frequency bands, respectively. Amendments like the aforementioned ones lay the foundation of next-generation WLANs by including new features such as multiple-antenna techniques like Downlink/Uplink Multi-User Multiple-Input-Multiple-Output (DL/UL MU-MIMO), spatial reuse techniques like BSS coloring, and efficient use of channel resources like DL/UL Orthogonal Frequency Division Multiple Access (OFDMA). Therefore, it becomes necessary to provide reliable simulation tools able to assess the performance and behavior of next-generation WLANs in multiple scenarios/cases, especially in high-density deployments.

In this paper, we present Komondor,¹ an open source, event-driven simulator based on the CompC++ COST library [6]. Komondor is focused on fulfilling the need for assessing the novel features introduced in recent and future amendments, which may be endowed with applications driven by machine learning (ML). The motivation for developing and building the presented wireless network simulator is motivated by:

- i) The lack of analytical models for capturing next-generation techniques in spatially distributed and/or high-density deployments.
- ii) The lack of next-generation WLAN-oriented simulators.

¹All of the source code of Komondor, under the GNU General Public License v3.0., is open, and potential contributors are encouraged to participate. The repository can be found at <https://github.com/wn-upf/Komondor>

- iii) The complexity of extending simulators comprising an exhaustive implementation of the physical (PHY) layer.
- iv) The large (or intractable) execution time required by other simulators to simulate high-density deployments.
- v) The need for conveniently incorporating ML-based agents in the simulation tool.

In short, Komondor is designed to efficiently implement new functionalities by relying on flexible and simplified PHY layer dependencies, to be faster than most off-the-shelf simulators, and to provide reliable simulations and a gentle learning curve to new users.

2 Wireless network simulators

Wireless network simulators can be categorized into continuous-time and discrete-event. On the one hand, continuous-time simulators continuously keep track of the system dynamics by dividing the simulation time into very small periods of time. On the other hand, in discrete-event simulators, events are used to characterize changes in the system. Accordingly, for the latter, events are ordered in time and normally allow running faster simulations than continuous-time simulators. In addition, discrete-event simulators allow tracing events with higher precision.

From the family of discrete-event driven network simulators, only a few ones are publicly available. OMNET++ [7] is a component-based C++ simulation library that is not open-source and is used for modeling communication networks and distributed multiprocessor systems. OPNET is another commercial simulator that allows the integration of external components. NetSim [8] was conceived to provide an accurate simulation model oriented to the world wide web. To that purpose, the simulator was written in Java, which compromises simulation time with programming flexibility. When it comes to open source network simulators, a MATLAB-based link-level simulator was presented in [9] for supporting the IEEE 802.11g/n/ac/ah/af technologies. The ns-2 simulator [10] is another network simulator known for its accuracy and the integration with the network animator. Finally, the ns-3, which was introduced in 2006 to replace the ns-2, presents significant advantages over the ns-2 due to its detailed simulation features, becoming very popular among the research community [11]. Table 1 highlights in a nutshell the most important characteristics of the overviewed network simulators and Komondor.

Among the family of overviewed discrete-event simulators, we highlight the ns-3 open-source simulator due to its popularity and use it as a baseline for comparing against Komondor. Despite the plethora of features that are supported

²Although ns-2 and ns-3 do not provide a default graphical animation tool, there are tools supporting live animation, e.g., PyViz or NetAnim for ns-3 and NAM for ns-2.

³blueAn integration with OpenAI Gym has been recently provided to ns-3 [12], but the ML-based operation is not part of the simulator.

Table 1: Comparison of wireless network simulators.

Simulator	Open-source	Source lang.	Complexity	GUI	11ax features	ML/based module
ns-3	Yes	C++	High	No ¹	Partial	No ²
ns-2	Yes	C++/OTcl	Low	No ¹	No	No
OMNET++	No	C/C++	Medium	Yes	No	No
OPNET	No	C++	Medium	Yes	No	No
NetSim	No	Java	Low	Yes	No	No
Komondor	Yes	C/C++	Low	No	Partial	Yes

in ns-3, it has some inherent limitations, such as the high complexity for developing new features/models as an extension of the simulator core. In particular, compatibility with the already existing/supported models is required and must be carefully ensured. For example, beamforming for previous mature amendments (i.e. IEEE 802.11n/ac) is not available yet, owing to the effort required to integrate it. Moreover, the integration of new features strongly depends on the willingness of the community to contribute to the development.

With respect to the IEEE 802.11ax operation – rates and support for information elements are being developed – the implementation is mostly based on the Draft 1.0 [13]. Such a draft dates from 2016 and does not include most of the core IEEE 802.11ax functionalities. At the time of submitting this paper, only the Single-User Protocol Data Unit (SU PPDU) and MIMO with up to four antennas are supported in ns-3, whereas OFDMA and MU-MIMO are not supported in the official distribution [14].

Apart from the official resources, we find few ns-3 works publicly available that support IEEE 802.11ax features, which may (or may not) be integrated into future releases. For example, we highlight the works with regard to the OFDMA that have been carried out by Getachew Redieiteab et al. (based on the IEEE 802.11ax specification framework document [15]) and Cisco [16]. However, none of these works completely follow the latest developments in the IEEE 802.11ax standard and have not been validated through extensive simulations and testbed results, as had previously occurred with the OFDM [17]. In addition to OFDMA, the spatial reuse operation (i.e., BSS Color [18]) is under active development, whereas extensions of the capture effect have been applied to ns-3 to follow the IEEE 802.11ax guidelines and studied in [19] and in a testbed [20].

3 Komondor Design Principles

3.1 Architecture

Komondor aims to realistically capture the operation of WLANs. Henceforth, it reproduces actual transmissions on a per-packet basis. To that purpose, Komondor is based on the COST library, which allows building interactions between components (e.g., wireless nodes, buffers, packets) through synchronous and

asynchronous events. While the former are messages explicitly exchanged between components through input/output ports, the latter are based on timers. In practice, components perform a set of operations until a significant event occurs. For instance, a node that is decreasing its backoff may freeze it when an overlapping node occupies the channel. The beginning and end of such a transmission are examples of significant events, whereas decreasing the backoff counter is not. Nevertheless, events may be triggered by different timers. In the previous example, a node’s transmission begins once the backoff timer terminates (i.e., the backoff timer triggers the beginning of the transmission), while the end of the transmission is triggered by the packet transmission timer. Fig. 2 shows the schematic of a COST component, which is composed of imports, exports, and a set of timers.

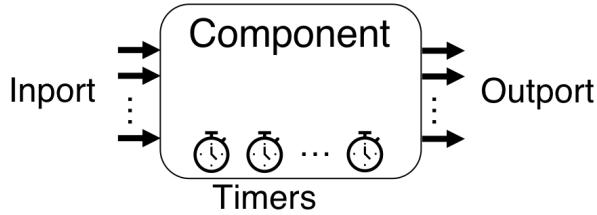


Figure 2: COST component. While imports and exports allow to directly communicate with other components, timers trigger events specific to the component.

3.2 IEEE 802.11 Features

Komondor entails a long-term project in which several contributors are involved. That is, the simulator is continuously evolving to include novel techniques and generally improve performance. The current version of Komondor (v2.0) includes the following fully tested IEEE 802.11ax features:

- **Distributed coordination function (DCF):** the Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) captures the basic Wi-Fi operation for accessing the channel. Moreover, Contention Window (CW) adaptation is considered.
- **Buffering and packet aggregation:** several traffic generator models are implemented in Komondor such as deterministic, Poisson or full-buffer. Besides, multiple media access control protocol data unit (MPDU) can be aggregated into the same PLCP protocol data unit (PPDU) in order to reduce the generated communication overheads.
- **Dynamic channel bonding (DCB):** multiple channel widths can be selected during transmissions by implementing DCB policies in order to

maximize the spectrum efficiency. Some of these policies were already evaluated in [21], [22].

- **Modulation coding scheme (MCS) selection:** the MCS is negotiated between any transmitter-receiver pair according to the Signal-to-Interference-and-Noise Ratio (SINR), thus supporting multiple transmission rates.
- **Ready-to-send/Clear-to-send (RTS/CTS) and Network Allocation Vector (NAV):** virtual carrier sensing is implemented in order to minimize the number of collisions by hidden-nodes.

Future development stages are under progress including other features such as OFDMA, MU-MIMO transmissions, beamforming, spatial reuse, and ML-based configuration.

3.3 Execution Flowchart

Komondor is composed of several modules that allow performing simulations with a high degree of freedom. Fig. 3 summarizes the operational mode of Komondor from a user’s point of view. A more detailed user’s guide providing a quick-start and guided execution examples is available in the Komondor’s Github repository.

3.3.1 Input and Setup/Start

as for the execution console command for starting Komondor simulations, arguments are designed in a simple and efficient way. Examples of console arguments are the file names of the inputs, the activation flags of the logs, the simulation time and the random seed. In addition, input files (in CSV format) are used to define the environment and have been conceived in a way that the user can easily modify important simulation parameters such as the traffic load, the path-loss model, or the data packet size. Once the environment is generated and nodes are initialized, traffic is exchanged between nodes until the simulation time runs out.

3.3.2 Stop and Output

when the simulation finishes, the closing is handled and statistics are gathered. Then, extensive and detailed performance statistics are per default provided by Komondor (e.g., throughput, delay, spectrum utilization, or collisions). Moreover, the user can efficiently include as much as metrics as desired.

3.4 States and events

The Komondor’s core operation is based on states, which represent the status (or situation) in which a node can be involved. A state diagram summarizing both states and transitions is shown in Fig. 4. Roughly, a given node starts

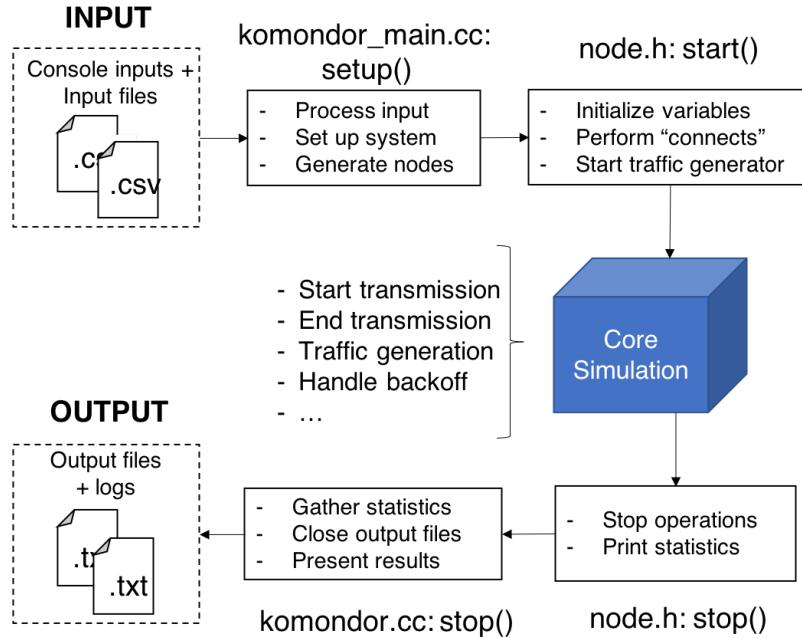


Figure 3: Komondor execution flowchart.

in the SENSING state, where multiple events can occur (e.g., a new packet is buffered or a new transmission is detected). Then, according to the noticed event, the node transits to the corresponding state.

3.4.1 States

we depict below each state and how a node must behave in front of new events.

- **SENSING:** a node senses the channel with two main purposes. First, to follow the CSMA/CA operation to gain access to the channel (in case there is backlogged data in the buffer/s). Second, to wait for incoming transmissions, so that either carrier sensing or receiving procedures are held. In case of being immersed in a backoff procedure, a node detecting a “new transmission” event would sense the power received in its primary channel, and assess whether to freeze the backoff countdown. Similarly, whenever an “end transmission” event occurs, the channel is sensed in order to determine whether the backoff counter can be resumed or remain paused.
- **TRANSMIT:** a transmitter node is currently transmitting a packet. No matter what events may occur, during the packet transmission, the node

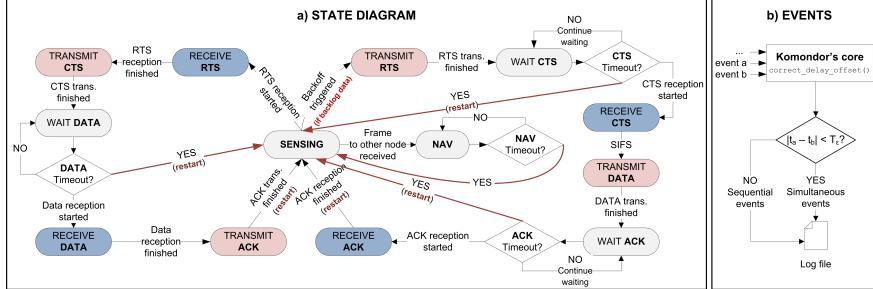


Figure 4: Komondor’s state diagram and events. a) States are reachable by different transitions. b) Simultaneous events are properly processed through delay offset correction.

blocks its receiver capabilities and remains in the same state until the transmission is finished.

- **RECEIVE:** when a node is receiving and decoding an incoming packet, it will behave in front of a new event according to its implication in the channel of interest. Of especial importance are those new transmission events triggered by other nodes that have gained access to the medium. Specifically, if a new transmission generates enough interference, the ongoing reception will be discarded, thus leading to a packet loss.
- **WAIT states:** these states allow modeling the situations where a node that transmitted a packet is expecting for the corresponding response. Namely, after transmitting RTS, CTS or DATA packets, the transmitter will wait for the corresponding CTS, DATA or ACK/BACK packets, respectively. If the response packet is not received before the corresponding timeout is triggered, the transmitter assumes that either the transmitted packet or the response packet is lost and resets to SENSING state. Wait states are particularly useful to detect packet losses when anomalies in the network (e.g., hidden terminal problem) occur.
- **NAV:** when a node enters in NAV state due to the successful reception of a frame addressed to a different destination, it sets a NAV timer and keeps listening to its primary channel. If a new frame is successfully received during the NAV, the timer is updated, provided that the new NAV time is larger than the current remaining time.

3.4.2 Events

each time a node performs an action that can affect the system (e.g., it starts transmitting a frame), an event is announced. Events in Komondor are lined up on the time axis and handled by the core entity. Events management is similar in ns-3. However, the latter exhibits a significant limitation, since events that

are scheduled at the exact same time can be executed in any order. Such a development feature may lead to unpredictable results and is incompatible with real-world situations in which events can occur simultaneously. Some inconsistencies may occur in case that the execution order affects multiple simultaneous events (e.g., two packets arriving at the exact same time). To solve this, Komondor, which is also a discrete-event simulator, employs temporal variables to compare the exact timestamps at which two or more events were generated. As a result, Komondor is able to successfully simulate the behavior of simultaneous events while keeping the logic of the states.

3.5 Developing new modules

Komondor has been conceived to be easily modified and extended. In particular, several modules have been provided to represent different simulation capabilities (e.g., propagation, channel access or traffic generation). Accordingly, Komondor can be potentially extended to support the operation of other IEEE 802.11 amendments such as 11n, 11ac, 11ad or 11ay. In addition, ML-based modules can also be introduced. A complete manual can be found at the Komondor’s repository.

4 Validation

In this Section, we validate the operation of Komondor and show its potential for dealing with high-density scenarios. In particular, we show the reliability of the simulator, despite its reduced complexity of the PHY.⁴ The validation of the Komondor’s operation is done through a set of illustrative scenarios, and our results are compared with the ones obtained with ns-3.⁵ In addition to ns-3, a mutual validation is performed with the Continuous Time Markov Networks (CTMNs) modeling introduced in [24], and which is extended for spatially distributed networks in the Spatial-Flexible Continuous Time Markov Network (SFCTMN) framework [21]. As for high-density scenarios, we make use of the Bianchi’s DCF analytical model [25] to validate the results in fully-overlapping deployments, where all the nodes are within the carrier sense of the others. The results shown in the following subsections were obtained according to the parameters defined in Table 2. The duration of the RTS, CTS and data frame is computed as follows:

⁴For instance, channel effects are assumed to remain static during the whole transmission of a given frame, and the propagation delay is considered to be negligible.

⁵Details on the ns-3 implementation used in the simulations presented throughout this paper can be found at <https://github.com/wn-upf/Komondor/tree/master/Documentation/Validation/ns-3>. For instance, this implementation includes the 11ax residential scenario propagation loss [23] and has a PLCP training duration updated according to the 11ax amendment [2].

$$T_{\text{RTS}} = T_{\text{PHY-leg}} + \left\lceil \frac{L_{SF} + L_{\text{RTS}}}{L_{s,l}} \right\rceil \sigma_{\text{leg}},$$

$$T_{\text{CTS}} = T_{\text{PHY-leg}} + \left\lceil \frac{L_{SF} + L_{\text{CTS}}}{L_{s,l}} \right\rceil \sigma_{\text{leg}},$$

$$T_{\text{D}} = T_{\text{HE-SU}} + \left\lceil \frac{L_{SF} + L_{\text{MH}} + N_{\text{agg}} L_{\text{D}}}{L_{s,l}} \right\rceil \sigma.$$

Note that full-buffer traffic is assumed in all the scenarios throughout this work for comparative purposes. Moreover, we have considered the residential path-loss model recommended in the IEEE 802.11ax [23], which inflicts high losses due to its large number of obstacles (e.g., walls).

4.1 Analyzing toy Scenarios

Komondor has been conceived as a friendly and ready-to-use wireless network simulator that can be used by researchers and teachers to study fundamental networking issues. In particular, scenarios and environment configurations can be conveniently modified through structured input files. The scenarios proposed in this Section are a clear example of toy scenarios where different networking concepts such as *flow starvation* or *additive interference* take place. Furthermore, a given user can easily analyze WLAN scenarios through the implemented logs generation system and statistics reporting. Accordingly, particular phenomena in the PHY and medium access control (MAC) layers can be tracked (e.g., channel contention, packet collisions, physical carrier sensing, energy detection, or buffer dynamics).

4.2 Basic Operation

We first aim to validate the basic IEEE 802.11 operation of the DCF implemented in Komondor when RTS/CTS is applied. For that, we consider a single Access Point (AP) scenario (we name it *Scenario 1*) with one and two stations (STAs), where full-buffer downlink traffic is held. The two-STAs case allows us to assess the proper behavior of Komondor in presence of multiple STAs. To validate this scenario, we compare the Komondor results with the ones provided by ns-3 and the SFCTMN framework. Fig. 5 shows the simulation results obtained from each tool, for packet aggregation ($N_{\text{agg}} = 64$) and no-aggregation ($N_{\text{agg}} = 1$). We note that the average throughput obtained by each simulation tool is almost identical, either for packet aggregation or not. In addition, having multiple STAs leads to the same result as for a single one since the destination STA is picked at random in every transmission.

4.3 Complex inter-WLAN interactions

In order to validate the behavior of Komondor in front of more complex inter-WLAN interactions, we now focus on the three-WLANs scenarios shown in

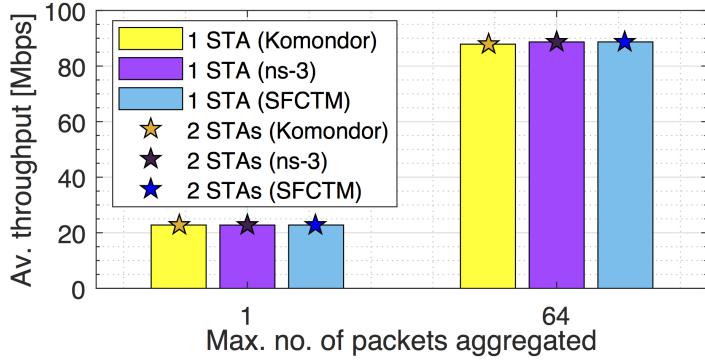


Figure 5: Average throughput experienced by the WLAN of *Scenario 1*, for $N_{\text{agg}} = 1$ and $N_{\text{agg}} = 64$. Results obtained from each simulation tool are shown.

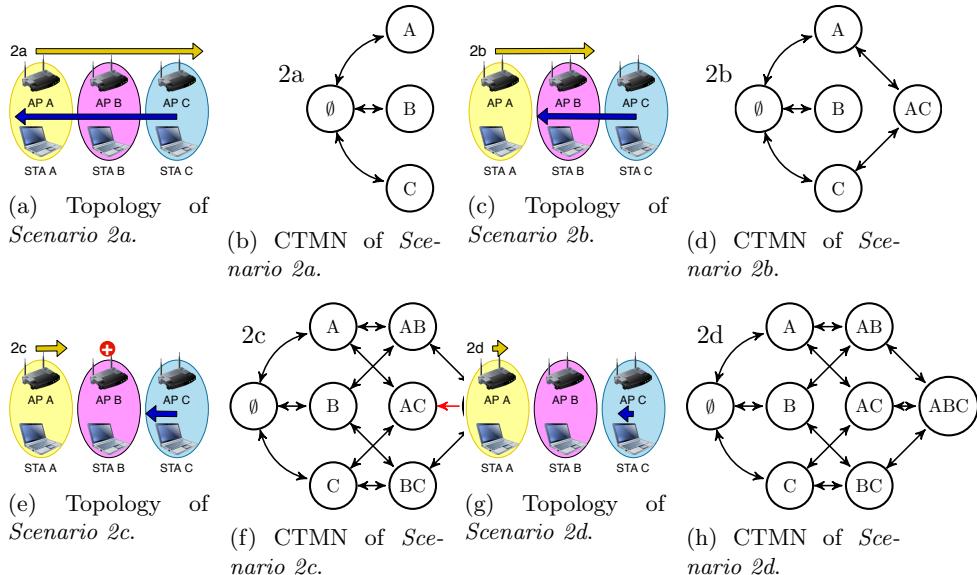


Figure 6: Topologies and corresponding CTMNs of scenarios 2a-2d. The yellow and blue arrows represent the area of interference from transmitters in WLANs A and C, respectively, whereby medium contention is forced.

Fig. 6. We name them *Scenario 2a-2d*. The interactions occurring in such scenarios are illustrated through CTMNs, where states⁶ represent the WLANs that are currently transmitting. Note that each of these scenarios reflects different situations that are of particular interest since they generalize different

⁶Note that CTMN states are not related by any means to Komondor states.

well-known phenomena in wireless networks:

- **Fully overlapping (Fig. 6a):** all the nodes cause contention to all the others when transmitting. For that, the distance between consecutive APs and between AP and STA of the same WLAN is set to $d_{AP,AP} = d_{AP,STA} = 2$ m, respectively.
- **Flow starvation (Fig. 6c):** contention is triggered in a pair-wise manner, so that WLAN_A and WLAN_C do not interfere each other. For that, the distance is set to $d_{AP,AP} = 4$ m and $d_{AP,STA} = 2$ m. Note that this case could be also extended to show a hidden node effect if AP_A or AP_C were intended to transmit to a STA located at the location of AP_B.
- **Potential overlap (Fig. 6e):** contention only occurs at WLAN_B when both WLAN_A and WLAN_C transmit concurrently. Otherwise, the channel is sensed as free. Note that, in this case, packets are successfully transmitted in WLAN_B whenever it access the channel. The distances are $d_{AP,AP} = 5$ m and $d_{AP,STA} = 2$ m for WLAN_A and WLAN_C, and $d_{AP,STA} = 3$ m for WLAN_B.
- **No overlapping (Fig. 6g):** none of the nodes causes contention to any other when transmitting. That is, every WLAN operates like in isolation. The distances in this case are $d_{AP,AP} = 10$ m and $d_{AP,STA} = 2$ m.

The average throughput experienced by each WLAN in each scenario is shown in Fig. 7. As previously done, we compare the performance of Komondor with ns-3 and SFCTMN. Note that results gathered by both Komondor and ns-3 are very similar in all the cases. Concerning the differences in the average throughput values estimated by both simulators and SFCTMN, we observe two phenomena with respect to backoff collisions in topologies of *Scenario 2a* and *2c*. First, in *2a*, the throughput is slightly higher when the capture effect condition is ensured. This is due to the fact that concurrent transmissions (or backoff collisions) are permitted and captured in the simulators. Second, the most notable difference is given in *2c*, which is caused by the assumption of continuous time backoffs in the CTMN. These are clear examples of the limitations of the analytical tool.

4.4 High-density and simulator performance

Finally, we assess the performance of Komondor when dealing with high-density scenarios. Notice that being able to simulate scenarios with a large number of nodes is a key feature due to the ever-increasing trend towards short-range and dense deployments. In this situation, we show the results of different fully-overlapping scenarios, ranging from 1 to 50 WLANs, each consisting in of one AP and one STA. The validation is performed against the Bianchi's analytical model and ns-3. The MCS for all the WLANs is set to 256-QAM. Fig. 8 shows the results in terms of throughput (average and aggregate) and collision probability obtained for fully overlapping networks of different sizes. For comparison

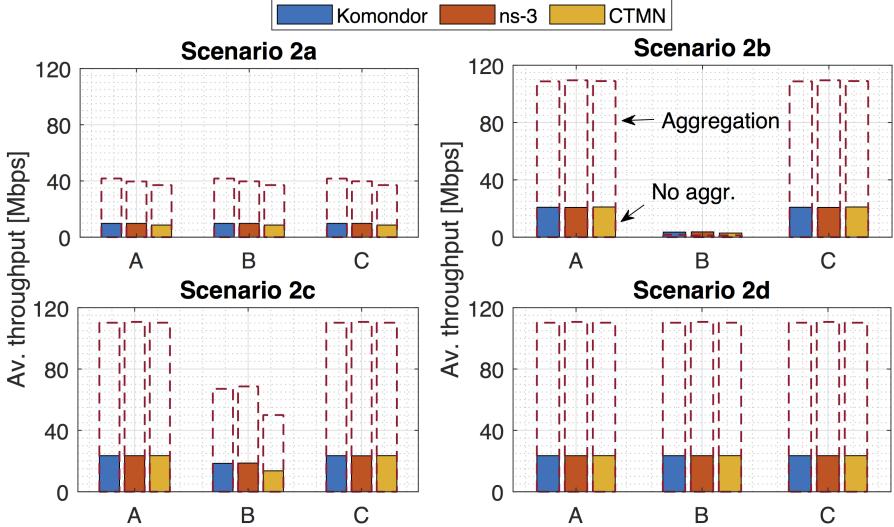


Figure 7: Average throughput experienced by each WLAN in scenarios 2a-2d. $N_{\text{agg}} = 1$ and $N_{\text{agg}} = 64$ are represented through solid bars and dashed lines, respectively.

purposes, the simulation time used in each scenario has been set to 100 seconds, for both Komondor and ns-3. Notice that such a fully overlapping setting frames a worst-case situation regarding packet collisions. This impacts on the number of events and the simulation time as the network density increases. Nevertheless, much more positive results are expected to be achieved in more realistic non-fully overlapping dense scenarios.

As shown, Komondor maintains its accuracy with respect to Bianchi's model, even when dealing with a lot of nodes. Regarding ns-3, slight differences are noticed in the collisions probability due to the error rate model, where collisions are based on the dropped RTS frames and the use of the Extended Interframe Space (EIFS). Moreover, differences in the throughput increase with the number of nodes, as previously addressed in [26].

To conclude this section, we provide insights into the execution complexity of Komondor. Fig. 9 shows the execution time and the number of generated events in Komondor and ns-3 for each number of WLANs.⁷ As shown, the execution complexity of ns-3 is significantly higher than in Komondor. We identify the cause of this difference to be the complex PHY implementation in ns-3, which leads to a larger number of generated events.

⁷Note that the execution time is strongly dependent of the computer used and its status at the moment of performing the simulation. In our case, we used an Intel Core i5-4300U CPU @ 1.9 GHz x 4 and 7.7 GiB memory.

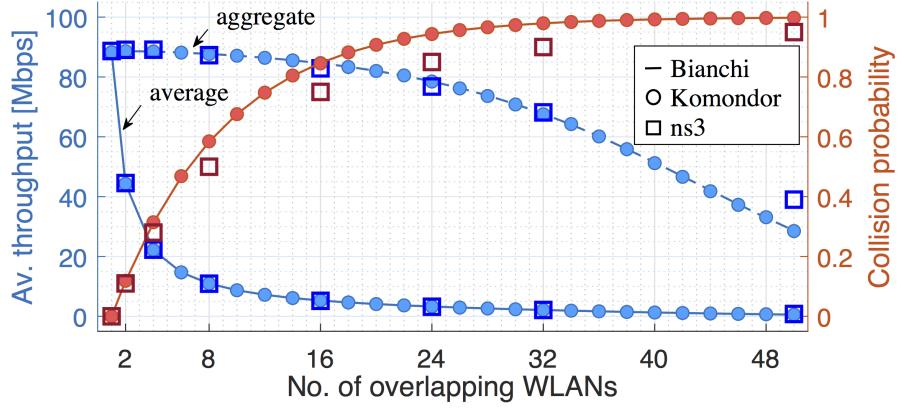


Figure 8: Throughput (average and aggregate) and collision probability vs. number of overlapping WLANs. Only some ns-3 points are plotted for the sake of visualization.

5 Komondor and potential use cases

Apart from small deployments consisting of few WLANs under single-channel operation [27], more complex scenarios capturing DCB or high-density scenarios have been already validated and analyzed by using Komondor. In this section, we briefly discuss further potential uses such as the implementation of next-generation WLAN techniques or the inclusion of learning agents to perform efficient spectrum access and spatial reuse.

5.1 Potential usage

Complex wireless environments can be already extensively simulated by Komondor as a result of its reduced computational complexity in comparison to other well-known simulators such as ns-3. A prominent example of a complex scenario mixing both high-density deployments and DCB is discussed in [21], where authors assessed the performance of different DCB policies versus node density (see Fig. 1). In [22], a similar deployment is analyzed while considering different traffic loads. A set of scenarios including DCB is shown in Fig. 10, which were validated in Komondor’s validation report v0.1.⁸ New features like spatial reuse, MIMO, beamforming and MU communications through OFDMA and/or MU-MIMO are currently under development.

⁸Komondor’s validation report v0.1: https://github.com/wn-upf/Komondor/blob/master/Documentation/Other/validation_report_v01.pdf.

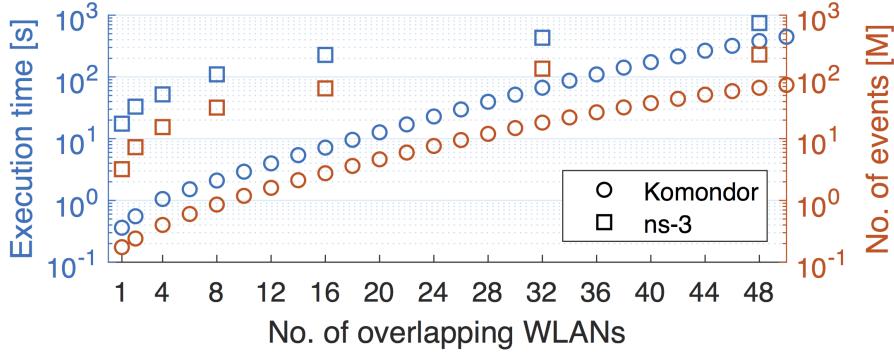


Figure 9: Execution time and number of generated events vs. number of overlapping WLANs.

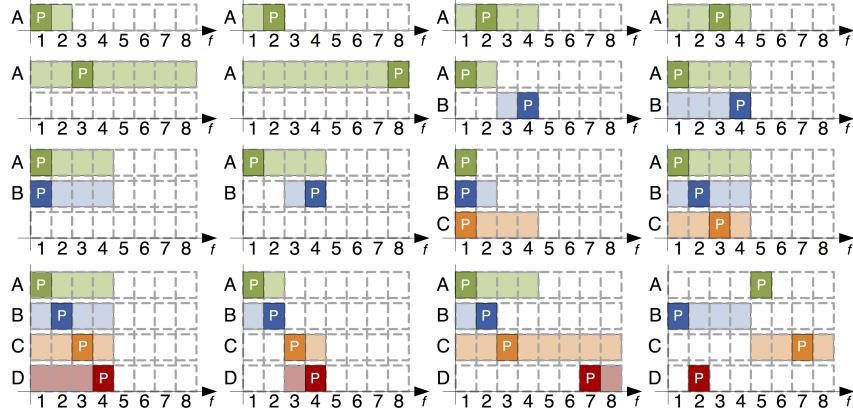


Figure 10: Scenarios with different DCB capabilities.

5.2 Machine learning agents

In addition to simulating advanced techniques proposed by the latest IEEE 802.11 amendments, Komondor permits including intelligent agents. In particular, agents are embedded to APs (see Fig. 11a) to perform the following operations (see Fig. 11b): *i*) monitor WLAN's performance, *ii*) run an implemented learning method, and *iii*) suggest new configurations to be applied by the WLAN, according to generated knowledge.

The application of intelligent agents has been previously studied in [27, 28], where decentralized learning is employed to both Transmit Power Control (TPC) and Carrier Sense Threshold (CST) adjustment.

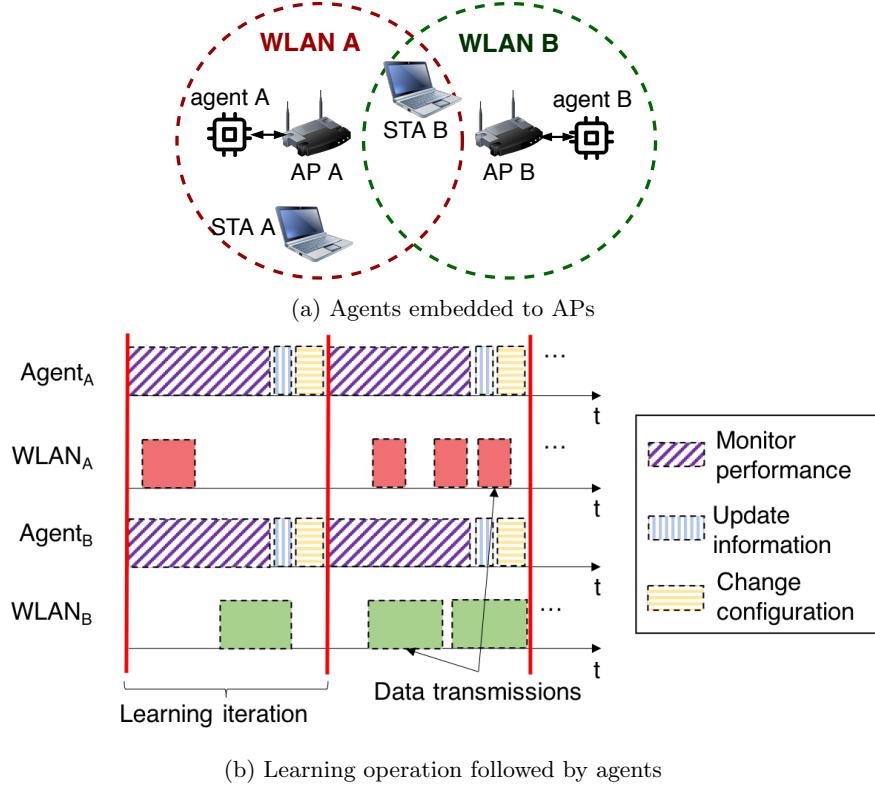


Figure 11: ML-based operation implemented in Komondor.

6 Conclusions

In this work, we presented Komondor, a wireless network simulator that stems from the need of providing a reliable and low-complexity simulation tool able to capture the operation of novel WLAN mechanisms like DCB or spatial reuse. The operation of Komondor has been validated against the ns-3 simulator and analytical tools such as CTMNs and Bianchi's DCF model. In this regard, we have shown its effectiveness when dealing with high-density scenarios, thereby outperforming ns-3 with respect to the simulation time. The provided validation is fundamental for the next development stages, which contemplate the inclusion of novel techniques in WLANs that have not been fully implemented in other well-known simulators. Some future implementations contemplate OFDMA, MU-MIMO, and the spatial reuse operation, naming a few among others. Finally, we have discussed the potential of Komondor regarding complex scenarios and ML integration. In particular, a preliminary ML-based architecture is already implemented, so that intelligent agents can rule self-configuring operations at different communication levels.

Acknowledgments

This work has been partially supported by a Gift from CISCO University Research Program (CG#890107) & Silicon Valley Community Foundation, by the Spanish Ministry of Economy and Competitiveness under the Maria de Maeztu Units of Excellence Programme (MDM-2015-0502), and by the Catalan Government under grant SGR-2017-1188. The work by S. Barrachina-Muñoz is supported by an FI grant from the Generalitat de Catalunya.

References

- [1] I. Selinis, K. Katsaros, M. Allayioti, S. Vahid, and R. Tafazolli, “The Race to 5G Era; LTE and Wi-Fi,” *IEEE Access*, vol. 6, no. 1, pp. 56 598–56 636, December 2018.
- [2] “IEEE p802.11ax/d2.0,” pp. 1–596, November 2017.
- [3] B. Bellalta, “IEEE 802.11ax: High-efficiency WLANs,” *IEEE Wireless Communications*, vol. 23, no. 1, pp. 38–46, February 2016.
- [4] Y. Ghasempour, C. R. da Silva, C. Cordeiro, and E. W. Knightly, “IEEE 802.11 ay: Next-generation 60 GHz communication for 100 Gb/s Wi-Fi,” *IEEE Communications Magazine*, vol. 55, no. 12, pp. 186–192, 2017.
- [5] L. C. et al., “EXtreme Throughput (XT) 802.11,” 2018.
- [6] G. Chen and B. K. Szymanski, “Reusing simulation components: cost: a component-oriented discrete event simulator,” in *Proceedings of the 34th conference on Winter simulation (WSC): exploring new frontiers*. Winter Simulation Conference, 2002, pp. 776–782.
- [7] A. Varga and R. Hornig, “An overview of the OMNeT++ simulation environment,” in *Proceedings of the 1st international conference on Simulation tools and techniques for communications, networks and systems & workshops*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2008, p. 60.
- [8] A. K. Rathi and A. J. Santiago, “The new NETSIM simulation program,” *Traffic engineering & control*, vol. 31, no. 5, 1990.
- [9] J. Milos, L. Polak, M. Slanina, and T. Kratochvil, “Link-level simulator for WLAN networks,” in *Link-and System Level Simulations (IWSLS), International Workshop on*. IEEE, 2016, pp. 1–4.
- [10] T. Issariyakul and E. Hossain, “Introduction to Network Simulator 2 (NS2),” in *Introduction to Network Simulator NS2*. Springer, 2012, pp. 21–40.

- [11] G. F. Riley and T. R. Henderson, “The ns-3 network simulator,” in *Modeling and tools for network simulation*. Springer, 2010, pp. 15–34.
- [12] P. Gawłowicz and A. Zubow, “ns3-gym: Extending openai gym for networking research,” *arXiv preprint arXiv:1810.03943*, 2018.
- [13] R. Stacey *et al.*, “Proposed TGax draft specification,” *IEEE 802.11-16/0024rl*, 2016.
- [14] “Wi-Fi design,” <https://www.nsnam.org/docs/models/html/wifi-design.html>, 2018.
- [15] R. Stacey, “Specification framework for tgax,” <https://mentor.ieee.org/802.11/dcn/15/11-15-0132-17-00ax-spec-framework.docx>, 2016.
- [16] Cisco, “NS3 Simulator of 802.11ax (ns3-802.11ax-simulator),” {<https://github.com/cisco/ns3-802.11ax-simulator>}, year=2017.
- [17] G. Pei and T. R. Henderson, “Validation of OFDM error rate model in ns-3,” *Boeing Research Technology*, pp. 1–15, 2010.
- [18] I. Selinis, M. Filo, S. Vahid, J. Rodriguez, and R. Tafazolli, “Evaluation of the DSC Algorithm and the BSS Color Scheme in Dense Cellular-like IEEE 802.11ax Deployments,” in *2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Sept 2016, pp. 1–7.
- [19] I. Selinis, K. Katsaros, S. Vahid, and R. Tafazolli, “Exploiting the Capture Effect on DSC and BSS Color in Dense IEEE 802.11 ax Deployments,” in *Proceedings of the Workshop on ns-3*. ACM, 2017, pp. 47–54.
- [20] E. Khorov, A. Kureev, I. Levitsky, and A. Lyakhov, “Testbed to Study the Capture Effect: Can We Rely on this Effect in Modern Wi-Fi Networks,” in *2018 IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom)*, June 2018, pp. 1–5.
- [21] S. Barrachina-Munoz, F. Wilhelmi, and B. Bellalta, “Dynamic Channel Bonding in Spatially Distributed High-Density WLANs,” (*forthcoming*) *IEEE Transactions on Mobile Computing*, 2019.
- [22] S. Barrachina-Muñoz, F. Wilhelmi, and B. Bellalta, “To Overlap or not to Overlap: Enabling Channel Bonding in High-Density WLANs,” *Computer Networks*, vol. 152, pp. 40 – 53, 2019.
- [23] S. Merlin, et al., “TGax Simulation Scenarios,” doc. IEEE 802.11-14/0980r16, 2016.
- [24] B. Bellalta, A. Zocca, C. Cano, A. Checco, J. Barcelo, and A. Vinel, “Throughput analysis in csma/ca networks using continuous time markov networks: a tutorial,” in *Wireless Networking for Moving Objects*. Springer, 2014, pp. 115–133.

- [25] G. Bianchi, “Performance analysis of the IEEE 802.11 distributed coordination function,” *IEEE Journal on selected areas in communications*, vol. 18, no. 3, pp. 535–547, 2000.
- [26] R. Patidar, S. Roy, T. R. Henderson, and M. Mehrnoush, “Validation of wi-fi network simulation on ns-3,” 2017.
- [27] F. Wilhelmi, S. Barrachina-Muñoz, B. Bellalta, C. Cano, A. Jonsson, and G. Neu, “Potential and pitfalls of multi-armed bandits for decentralized spatial reuse in wlans,” *Journal of Network and Computer Applications*, vol. 127, pp. 26–42, 2019.
- [28] F. Wilhelmi, C. Cano, G. Neu, B. Bellalta, A. Jonsson, and S. Barrachina-Muñoz, “Collaborative spatial reuse in wireless networks via selfish multi-armed bandits,” *Ad Hoc Networks*, 2019.

Table 2: Parameters considered in the presented scenarios.

Parameter	Description	Value
f_c	Central frequency	5 GHz
$ c $	Basic channel bandwidth	20 MHz
MCS	11ax MCS index	0-11
G_{tx}	Transmitting gain	0 dB
G_{rx}	Reception gain	0 dB
$PL(d)$	Path loss (Residential scenario)	see [23]
N	Background noise level	-95 dBm
σ_{leg}	Legacy OFDM symbol duration	4 μ s
σ	OFDM symbol duration (GI-32)	16 μ s
N_{sc}	Number of subcarriers (20 MHz)	234
N_{ss}	Number of spatial streams	1
T_e	Empty slot duration	9 μ s
T_{SIFS}	SIFS duration	16 μ s
T_{DIFS}	DIFS duration	34 μ s
T_{PIFS}	PIFS duration	25 μ s
$T_{PHY-leg}$	Legacy preamble duration	20 μ s
T_{HE-SU}	HE single-user field duration	100 μ s
T_{ACK}	ACK duration	28 μ s
T_{BACK}	Block ACK duration	32 μ s
T_{PPDU}^{\max}	Max. PPDU duration	5484 μ s
$L_{s,l}$	Size OFDM symbol (legacy)	24 bits
L_D	Data packet size	11728 bits
N_{agg}	No. of frames in an A-MPDU	1, 64
L_{RTS}	Length of an RTS packet	160 bits
L_{CTS}	Length of a CTS packet	112 bits
L_{SF}	Length of service field	16 bits
L_{MH}	Length of MAC header	320 bits
CW	Contention window (fixed)	15

Spatial Reuse in IEEE 802.11ax WLANs

Francesc Wilhelmi, Sergio Barrachina-Muñoz, Cristina Cano,
Ioannis Selinis, and Boris Bellalta

Abstract

Dealing with massively crowded scenarios is one of the most ambitious goals of next-generation wireless networks. With this goal in mind, the IEEE 802.11ax amendment includes, among other techniques, the Spatial Reuse (SR) operation. This operation encompasses a set of unprecedented techniques that are expected to significantly boost the performance of Wireless Local Area Networks (WLANs) in dense environments. In particular, the main objective of the SR operation is to maximize the utilization of the medium by increasing the number of parallel transmissions. Nevertheless, due to the novelty of the operation, its performance gains remain largely unknown. In this paper, we first provide a gentle tutorial of the SR operation included in the IEEE 802.11ax. Then, we analytically model SR and delve into the new kind of MAC-level interactions that occur among network devices. Finally, we provide a simulation-driven analysis to showcase the potential of SR in a variety of deployments, comprising different network densities and traffic loads. Our results show that the SR operation can significantly improve the medium utilization, especially in scenarios under high interference conditions. Moreover, our results demonstrate the non-intrusive design characteristic of SR, which allows enhancing the number of simultaneous transmissions with a low impact on the environment. We conclude the paper by giving some thoughts on the main challenges and limitations of the IEEE 802.11ax SR operation, as well as on the most prominent research gaps and future directions.

1 Introduction

Due to the popularity and ease of deployment of IEEE Wireless Local Area Networks (WLANs), it is becoming increasingly common to find multiple Basic Service Sets (BSSs) within the same overlapping areas. Unfortunately, the most typical channel access mechanism based on Carrier Sense Multiple Access (CSMA) was not designed to support a huge number of contending devices, thus resulting in low performance.

In order to improve the performance of WLANs, several amendments have been conceived along the past few years. Earlier IEEE 802.11 standards, e.g., 11n (2009) and 11ac (2013), defined the concepts of High Throughput (HT) and Very High Throughput (VHT) devices, respectively. These standards defined new functionalities to be included at that time, such as Channel Bonding (CB).

More recently, the Task Group ax (TGax) was created to develop the IEEE 802.11ax-2021 (11ax) standard [1], which belongs to the group of standards for next-generation WLANs (e.g., IEEE 802.11aq, IEEE 802.11ad, IEEE 802.11ay). Through the definition of High Efficiency (HE) WLANs, the 11ax aims to improve network efficiency in dense deployments. To that purpose, it includes several novel techniques, such as Orthogonal Frequency Division Multiple Access (OFDMA), Downlink/Uplink Multi-User Multiple-Input-Multiple-Output (DL/UL MU-MIMO), and the Spatial Reuse (SR) operation. We refer the reader to the works in [2–5] for an overview of the major novelties proposed in the IEEE 802.11ax standard.

In this paper, we focus on the 11ax SR operation [6], which seeks to increase the number of parallel transmissions and therefore improve spectral efficiency. In order to do so, the amendment introduces Carrier Sense Threshold (CST) adjustment for the detected inter-BSS transmissions¹), which is performed through two different mechanisms: *i*) OBSS Packet Detect (PD)-based SR, and *ii*) Parametrized Spatial Reuse (PSR). The main difference between the two mechanisms lies in the degree of collaboration among BSSs for identifying SR-based opportunities (further details are provided in Sections 3 and 4). Both mechanisms include Transmission Power Control (TPC) to limit the additional interference produced by simultaneous transmissions.

Figure 1 summarizes the components that constitute the 11ax SR operation, which are described in detail throughout this paper.

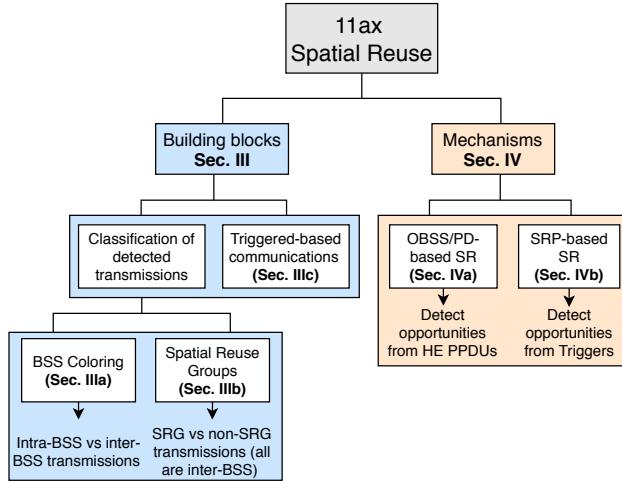


Figure 1: Summary of the 11ax SR operation.

To illustrate the potential SR enhancement in an OBSS, let us focus on Fig. 2. Unlike typical coverage representations in wireless networks, throughout

¹In the following, we will use intra-BSS or inter-BSS to refer to the transmissions detected from the same or from a different BSS, respectively.

this paper, we consider the carrier sense area of each device (instead of the generated interference). For that, our representation assumes that the transmission power used by any device is fixed and that all the BSSs use the same frequency channel, which allows focusing on the spatial interactions only. Accordingly, the dashed circles in Fig. 2 indicate the transmitters that can be detected by the node of interest. In our example, both Access Points (APs) can simultaneously transmit to their corresponding stations (STAs), provided that they use the enhanced combination of CST and transmission power (bold line). In contrast, parallel transmissions are not possible when using the default configuration.

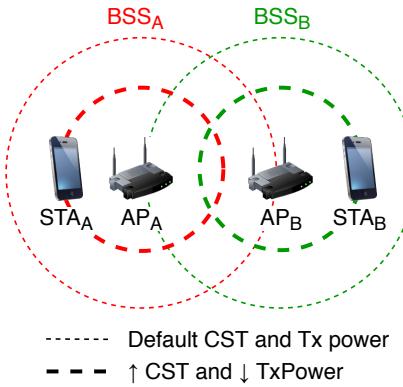


Figure 2: SR enhancement through CST adjustment and TPC. The carrier sensing area of each transmitter is graphically represented by the dashed lines.

In spite of the apparent benefits of the SR operation, its actual potential is still unknown. The fact is that SR depends on multiple factors, such as the network topology, the type of propagation effects, or the type of radio used by devices [7]. Moreover, sensitivity adjustment and power control may result in asymmetric links that can potentially lead to unfairness situations [8]. In some cases, dynamic sensitivity and transmission power adjustment have been shown to significantly increase the network performance and to contribute to reducing the effects of the well-known hidden and exposed terminal problems [9]. However, in some other cases, these problems may be exacerbated [10]. Indeed, modifying either the CST or the transmit power can worsen the hidden/exposed terminal problems by generating flow starvation and asymmetries.

Figure 3 shows in an intuitive manner the effect of increasing and decreasing both the transmission power and the sensitivity in WLANs. For instance, increasing the sensitivity of a device may contribute to accessing the channel more often since the listening area is reduced. However, this can lead to observing a higher number of collisions by hidden-node. Moreover, using a more aggressive channel access policy may expose the receivers to a higher level of interference, thus requiring the utilization of more robust Modulation and Coding Scheme (MCS).

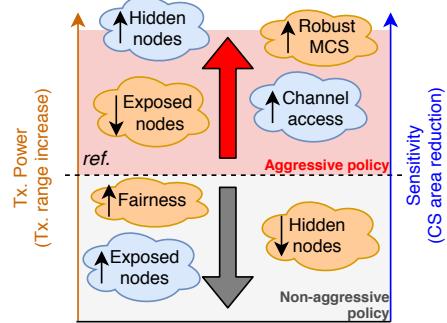


Figure 3: Effects of different policies with regards to sensitivity adjustment and transmission power control.

As discussed, dealing with the spatial dimension has different kinds of implications and leads to complex inter-BSS interactions that are hard to predict beforehand. Indeed, the 11ax SR operation is one of the least studied features in next-generation WLANs and only a few works have devised its potential. Firstly, the authors in [11] evaluated the benefits of using dynamic sensitivity thresholds for inter-BSS transmissions, given a fixed transmit power. Secondly, the work in [4] exhaustively surveyed the 11ax amendment, thus providing an overview of the first drafted SR operation. Moreover, it provided some results on applying SR in both indoor and outdoor scenarios, showing a higher potential for indoor deployments. Similarly, the authors in [12] introduced the contents of the 11ax SR as they are described in the amendment. In addition, they provided a performance evaluation based on the adjustment of the inter-BSS sensitivity threshold. Their results showed significant gains when applying SR, especially for dense scenarios.

Unlike in [4, 11, 12], in this paper we delve into the 11ax SR operation in more detail since we consider the two different SR operations included in the 11ax amendment. In addition, our analysis of the 11ax SR is not limited to the technical information included in the amendment. Instead, we accompany our descriptions with illustrative use cases, thus bringing a new perspective that allows devising the real utility behind the operation. We thus go beyond the definition of the specification, shedding light on its purpose, benefits, and challenges.

Our aim in this paper is to provide a comprehensive tool for researchers interested in the topic and to analyze the potential of the SR operation in future WLANs. Besides, we focus on the potential gaps in the standard to be filled by the research community. The main contributions of this paper lie in the description, analysis, and evaluation of the 11ax SR operation. In particular:

1. We provide a gentle, exhaustive, and comprehensive overview of the SR operation included in the 11ax amendment.

2. We analytically model the 11ax SR operation for the sake of capturing the new kinds of inter-BSS interactions and understanding their implications. The results of this model are verified with the Komondor 11ax-based simulator [13].
3. We study the potential performance gains of the 11ax SR operation through simulations.
4. We delve into the gaps and gray areas existing in the current 11ax SR operation and provide forecasts of future research directions in the field.

The remainder of this document is structured as follows. Section 2 surveys the related work on SR in WLANs. Section 3 describes the specifications and procedures to enable 11ax SR, whereas Section 4 details the operation itself. Section 5 presents an analysis-based study of 11ax SR in simple scenarios, which is extended in Section 6 by simulating dense deployments. Section 7 identifies the gaps and research opportunities found within the 11ax SR operation and explores potential ways forward. Finally, Section 8 provides some concluding remarks.

2 Spatial Reuse techniques in IEEE 802.11 WLANs

The problem of dynamic sensitivity and transmission power adjustment has been previously addressed in multiple ways. On the one hand, we find centralized solutions such as the ones proposed in [14–16], where the SR operation is controlled and mandated from the APs. Among these, we highlight [15], which uses a method based on Neural Networks (NN) to compute the best combination of sensitivity and transmit power to be used by all the BSSs in a given scenario. Nonetheless, centralized approaches require coordination and extra overhead, which is usually impractical.

On the other hand, SR has been addressed through a decentralized perspective in [10, 17–20]. Most of the decentralized strategies rely on collecting feedback on several performance metrics (e.g., sensed interference, packets lost, etc.). While works such as [17–19] propose adaptive mechanisms to adjust the CST and/or the transmission power, some others like [10, 20] provide probabilistic approaches based on Reinforcement Learning (RL) for finding the best possible configuration.

Concerning IEEE 802.11ax WLANs, the Dynamic Sensitivity Control (DSC) scheme was proposed to be included in the standard, but it was never incorporated. The performance of DSC was evaluated in [21–23]. Furthermore, the authors in [24, 25] combined DSC with BSS color schemes to devise further improvements in WLANs.

The current 11ax SR operation has nonetheless been studied to a lower extent. Based on the OBSS/PD-based SR operation, the work in [26] proposed

a new mechanism to adjust the OBSS/PD threshold.² This mechanism, so-called Control OBSS/PD Sensitivity Threshold (COST), differs from DSC in terms of the information available in 11ax nodes. In this case, nodes need to be aware of changes in the neighboring BSSs.

Unlike previous works, we focus on the IEEE 802.11ax SR operation defined in Draft v4.0 and delve into its potential through analytical modeling and a simulation tool. Moreover, we identify potential gaps and research opportunities with regard to the amendment.

3 IEEE 802.11ax Spatial Reuse Operation: Building Blocks

Before delving into the 11ax SR mechanisms, we first describe the enabling concepts and features. In particular, the 11ax SR operation can be understood through **BSS coloring** and **Spatial Reuse Groups (SRG)**. In addition, we introduce the **Triggered-based (TB) transmissions** upon which the PSR operation is based.

3.1 BSS coloring

BSS coloring is a key enabler of the 11ax SR operation, whereby HE nodes can rapidly identify the source of a given transmission. In case that BSS coloring is adopted in an OBSS, a given device can effectively determine whether the channel is occupied by another device belonging to the same BSS (intra-BSS transmission, same color) or from a different one (inter-BSS transmission, different color). The BSS color, which is determined by the AP and is included in the preambles of Wi-Fi frames,³ is a value in the range of 1 to 63. It remains static until the AP considers to change it. When noticing a BSS color overlap (i.e., two different BSSs use the same color), a new color may be chosen by the affected APs.

The method for selecting a new color is out of the scope of the 11ax amendment, but the advertising operation is actually defined. An HE AP may announce a new BSS color via the **BSS Color Change Announcement** element, which is carried in Beacon, Probe Response, and (Re)Association Response frames.

3.1.1 BSS color-based channel access rules

When detecting a transmission, an HE node can distinguish between intra and inter-BSS frames by rapidly inspecting the **BSS color** field that is carried in

²The OBSS/PD threshold refers to the sensitivity to be used for detected inter-BSS transmissions.

³The **BSS color** field is included in the Physical Layer Convergence Procedure (PLCP) header. See Appendix A for further details.

every HE PLCP Protocol Data Unit (PPDU).⁴ In particular, the default PD threshold (i.e., -82 dBm) is used for intra-BSS frames. So, from now onwards, we will refer to the default PD threshold simply as Clear Channel Assessment / Carrier Sense (CCA/CS). On the contrary, when inter-BSS frames are detected, more aggressive PD thresholds can be applied to increase the number of parallel transmissions. Those PD thresholds are termed **non-SRG OBSS/PD** and **SRG OBSS/PD**. The SRG OBSS/PD is used when spatial reuse groups are allowed, which is discussed in detail in Section 3.2.

To illustrate how BSS coloring can help at enhancing SR, let us consider the scenario shown in Fig. 4(a). We consider that AP_A is prone to suffer from flow starvation if it uses the default CCA/CS value, which entails being in the range of both AP_B and AP_C . Note that simultaneous downlink transmissions can be held in both BSS_B and BSS_C because the transmitters are not in the range of each other. This may lead to flow-in-the-middle starvation in BSS_A .

The flow-in-the-middle situation can be overtaken by AP_A if using an OBSS/PD value higher than the CCA/CS for inter-BSS frames, which would allow ignoring the transmissions sensed from AP_B and AP_C . This procedure is illustrated in Fig. 4(b), where AP_A first identifies the source of a detected transmission by inspecting its headers. Then, after detecting the source of the ongoing transmissions (indicated by the color), the non-SRG OBSS/PD threshold is applied. If the OBSS/PD threshold is high enough to ignore inter-BSS transmissions, AP_A can reset the PHY and continue the backoff procedure.

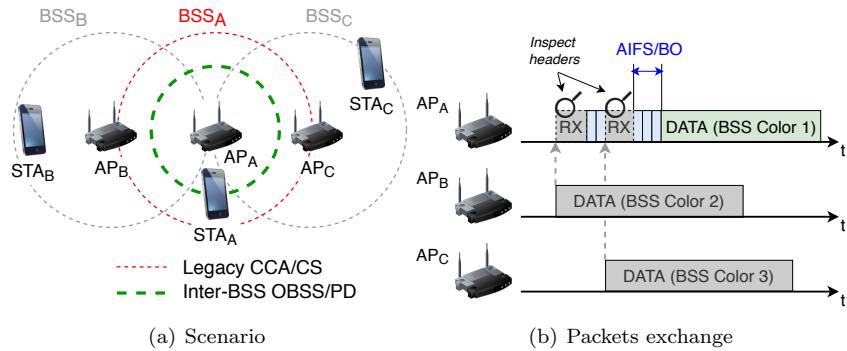


Figure 4: Channel access rules based on BSS coloring. In (b), the propagation delay is considered to be negligible.

3.1.2 Two NAVs

The SR operation provides significant changes in the virtual carrier sensing procedure since two different Network Allocation Vectors (NAVs), i.e., *intra-*

⁴If the **BSS color** is not announced, frames can be classified according to the **GROUP_ID** and **PARTIAL_AID** in VHT PPDUs, or the MAC address in the MAC header of legacy frames (i.e., predecessor amendments of the IEEE 802.11ac).

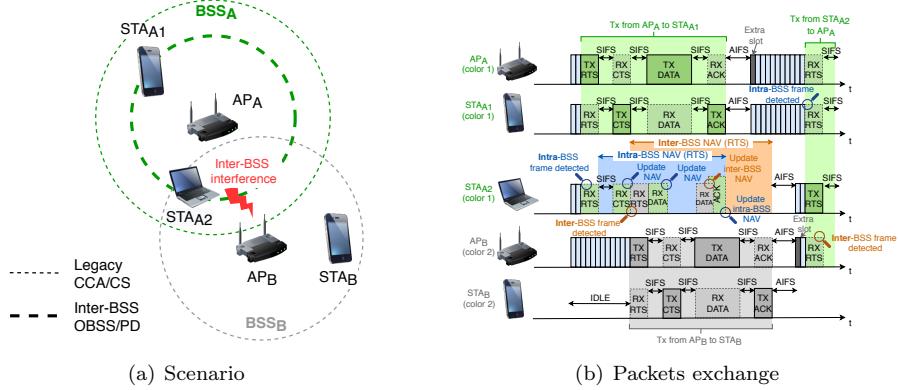


Figure 5: Two NAVs operation in an OBSS.

BSS NAV and *inter-BSS NAV*, have to be maintained for intra and inter-BSS frames, respectively. Accordingly, a given transmitter can decrease its backoff counter only if both NAV timers are set to zero. Otherwise, it must remain idle for at least the duration of the ongoing transmission(s),⁵ which had previously activated the virtual carrier sensing.

To showcase the utilization of two NAVs within the SR operation, we consider the scenario shown in Fig. 5(a), where packets are exchanged as illustrated in Fig. 5(b). In this scenario, BSS_A and BSS_B are within the same carrier sense area, provided that they both use the same CCA/CS value. In opposite, both BSSs can ignore each others' transmissions in case of using a higher OBSS/PD threshold.

Following the Distributed Coordination Function (DCF) operation, AP_A initiates a downlink transmission to STA_{A1} by first sending a Request-to-Send (RTS) frame. Then, STA_{A2} decodes the RTS frame and realizes that it was sent by an intra-BSS device (the BSS color field matches with its own color). As a result, STA_{A2} applies the default CCA/CS threshold to determine whether the channel remains idle or not. In this case, the sensed power is above the threshold, which makes STA_{A2} to apply virtual carrier sensing. It is also worth mentioning that the NAV timer is confirmed in STA_{A2} when STA_{A1} sends the Clear-to-Send (CTS) frame to AP_A, because they are in the sensing range.

In parallel to the abovementioned intra-BSS interactions, AP_B starts its own downlink transmission to STA_B. The initiating RTS frame sent by AP_B is also listened by STA_{A2}, which sets the inter-BSS NAV accordingly. On this occasion, the acknowledgments sent by STA_B are ignored by STA_{A2} due to the OBSS/PD threshold employed for inter-BSS transmissions. However, it has no impact on the modification of the inter-BSS NAV, which is properly set by the

⁵The duration used for setting the NAV is indicated in the Duration field of RTS/CTS frames or Physical layer Service Data Units (PSDU).

frames transmitted by AP_B .⁶ After both intra and inter-BSS NAV timers are over, STA_{A2} can perform an uplink transmission.

The utility behind maintaining two NAVs becomes evident for dense deployments. On the one hand, the intra-BSS NAV allows protecting STAs from intra-BSS transmissions, thus reducing the effect of certain anomalies such as the hidden-terminal problem. On the other hand, as a novelty, the inter-BSS NAV allows mitigating OBSS interference, which contributes to increasing the number of parallel transmissions.

3.2 Spatial Reuse Groups

To further enhance network efficiency, the 11ax amendment provides a mechanism for differentiating between two types of inter-BSS frames; that is to say, belonging or not to the same SRG. These groups can be formed by BSSs to achieve a more sophisticated SR operation. For instance, more aggressive channel access policies can be used for transmissions within the same SRG, in case that higher levels of interference could be supported by the nodes of the same SRG. Or it could be the other way around. A conservative policy can be employed for the sake of minimizing collisions by hidden nodes.

Despite the formation of SRGs is out of the scope of the amendment, differentiating between two OBSS/PD thresholds can be useful for capturing more subtle inter-BSS interactions. Note, as well, that SRGs could be formed online to address some issues detected by an entity controlling a set of APs (e.g., belonging to the same operator).

Figure 6 shows a deployment in which the formation of SRGs makes sense. In this case, the channel utilization can be enhanced by using the non-SRG OBSS/PD threshold to ignore inter-BSS transmissions (light dashed lines). However, using this threshold homogeneously leads STA_C to suffer from packet losses. In particular, STA_C cannot properly decode the information sent by AP_C when AP_A also occupies the channel. To solve this, BSS_A and BSS_C can form a group and employ a more conservative SRG OBSS/PD threshold to avoid simultaneous transmissions between the two BSSs. Notice that the non-SRG OBSS/PD threshold (which is more aggressive) can be still employed for transmissions held by any pair of BSSs involving BSS_B , thus increasing network efficiency. As shown, not only the formation of SRGs is a complex task, but also the definition of both non-SRG and SRG OBSS/PD thresholds.

To apply the SR operation based on SRGs, the involved HE nodes must have indicated support for this feature. With regards to HE STAs, they enable the SRG operation upon the reception of an activating **Spatial Reuse Parameter Set** element (further described in Appendix A.2.1) from their AP. Then, for the following detected PPDUs, both HE APs and STAs may differentiate between SRG and non-SRG PPDUs. Note that 11ax devices can also identify the source of non-HE transmissions. Therefore, not only HE devices are supported but also

⁶Note, as well, that the amendment also allows resetting the NAV if no CTS frame is received (i.e., a timeout occurs).

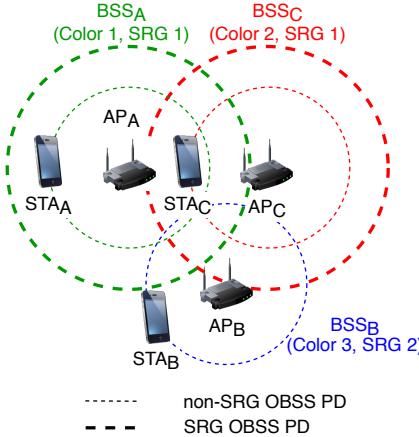


Figure 6: Spatial Reuse Groups in an OBSS.

legacy devices. The way of classifying frames according to the SRG is backward compatible with previous IEEE 802.11 amendments. Technically speaking, SRG identification is done as follows:

- For HE PPDUs, an HE STA inspects the BSS color and checks if it belongs to the same SRG. This information is kept on the SRG BSS Color Bitmap of the Spatial Reuse Parameter Set, which stores the different BSS colors that belong to the same SRG. The AP of a given BSS is responsible for maintaining the SRG BSS Color Bitmap up to date, and to inform STAs in case of noticing any change.
- When it comes to VHT PPDUs, inter-BSS transmissions are considered to belong to the same SRG if the GROUP_ID parameter (included in the RXVECTOR⁷) has a value of 0, and the bit in the SRG Partial BSSID Bitmap field corresponding to the numerical value of PARTIAL_AID⁸ (also included in the RXVECTOR) is set to 1.
- Finally, regarding other types of PPDUs, they are classified as SRG PPDUs if the BSSID information from a MAC Protocol Data Unit (MPDU) of the PPDUs is correctly received and the bit in the SRG Partial BSSID Bitmap field corresponding to the numerical value of BSSID is 1.

⁷The RXVECTOR constitutes a set of parameters that the PHY layer delivers to the MAC on receiving a PPDU.

⁸The PARTIAL_AID is an identifier which, similarly to the BSS color, is used by IEEE 802.11ac WLANs to quickly identify the source of a given transmission.

3.2.1 SRG-based Channel Access Rules

Differentiating between SRGs may provide further SR enhancements than considering only one type of inter-BSS frame. Despite the specific utilization of SRGs is also out of the scope of the 11ax amendment, we devise several situations where its application can be useful. As previously pointed out, one possibility is to establish groups for BSSs whose transmissions need to be protected. In other words, an HE STA detecting an SRG frame can implement a more conservative channel access policy. Conversely, a more aggressive policy can be applied for non-SRG PPDUs, thus increasing the number of parallel transmissions.

To illustrate the SRG-based channel access rules, let us retake the scenario shown in Fig. 6, where three overlapping BSSs potentially share the medium. While BSS_A and BSS_B belong to SRG 1, BSS_C belongs to SRG 2. Accordingly, different OBSS/PD thresholds are applied by BSS_A when detecting inter-BSS frames belonging to groups 1 or 2 (note that all the BSSs use different BSS colors).

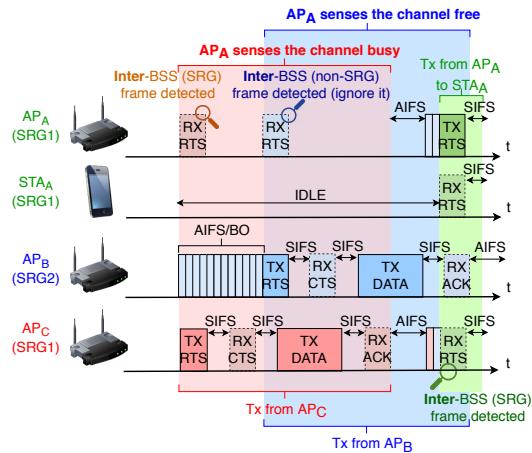


Figure 7: Packets exchange based on SRG channel access rules.

As shown in Fig. 7, transmissions from BSS_C (in blue) provoke that AP_A senses the channel busy. In contrast, packets detected from BSS_B (in red) are ignored by AP_A after PHY headers are properly inspected. In this example, a less restrictive OBSS/PD threshold is applied for SRG transmissions than for non-SRG ones. The fact is that STA_B is sufficiently far away from AP_A . Therefore, simultaneous transmissions between BSS_A and BSS_B are completely feasible. The opposite occurs for BSS_A - BSS_C interactions. Note that collisions may occur at STA_C if simultaneous transmissions among the abovementioned BSSs are held, thus requiring additional protection.

3.3 Triggered-based communications

As previously pointed out, one of the 11ax SR mechanisms relies on TB transmissions [27]. Roughly, in a TB communication, an AP schedules UL transmissions from one or more STAs. To that purpose, a Trigger Frame (TF) is sent by a given AP to indicate the group of users that are allowed to transmit during the current Transmission Opportunity (TXOP), along with other relevant information.

Fig. 8 illustrates an example of a TB transmission. After gaining access to the channel, the AP first sends a TF packet, which is received by HE STAs. Upon successful reception of the TF, STAs start simultaneous TB UL transmissions, which can be enabled by using multiple antenna technologies (i.e., MU-MIMO) or different OFDMA subcarriers. Once all the UL transmissions finish, the AP acknowledges all the packets with a multi-station block ACK (MACK).

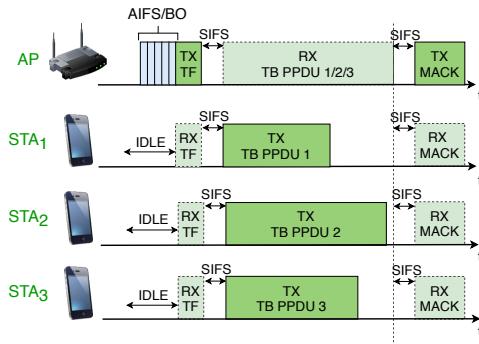


Figure 8: TB UL transmission held in a BSS.

SR in 11ax takes advantage of TB communications for detecting the so-called PSR opportunities. By inspecting an inter-BSS TF packet, an HE STA implementing PSR can determine the maximum allowed interference supported by the inter-BSS AP scheduling the transmission. As a result, it can transmit during the TXOP at a regulated transmission power. Further details on PSR are provided in Section 4.2.

Finally, it is worth mentioning that, before scheduling a UL transmission, APs can cancel the virtual carrier sensing of their STAs by sending a Contention Free End (CF-End) control frame. This is done to reduce the idle periods provoked by inter-BSS transmissions, thus enhancing network efficiency.

4 IEEE 802.11ax Spatial Reuse Operation

The IEEE 802.11ax SR operation is divided into two different mechanisms: *i)* OBSS/PD-based SR and *ii)* PSR. So far, we have described the elements

that enable both operations, thus providing insights on the potential of applying SR. In this Section, we show the technical details of IEEE 802.11ax SR, thus embodying the concepts that have been previously introduced in Section 3.

4.1 OBSS/PD-based Spatial Reuse

The OBSS/PD-based SR operation is based on CST and transmit power adjustment for detected inter-BSS frames. By knowing the source of an ongoing transmission, an HE STA may employ higher CST values for the sake of improving the probability of accessing the channel. In particular, upon PPDU reception, the MAC layer of a given device receives a notification from the PHY. At that moment, the node inspects the packet and, among several operations, it determines whether the PPDU is an intra-BSS or an inter-BSS frame. The latter may be subdivided into SRG or non-SRG frames, provided that SRGs are enabled.

4.1.1 General constraints

As a general rule, the OBSS/PD threshold that is used for detected inter-BSS frames cannot exceed a certain value. This upper bound is illustrated in Fig. 9, and is defined as follows:

$$\text{OBSS/PD} \leq \max \left(\text{OBSS/PD}_{\min}, \min \left(\text{OBSS/PD}_{\max}, \text{OBSS/PD}_{\min} + (\text{TX_PWR}_{\text{ref}} - \text{TX_PWR}) \right) \right),$$

where OBSS/PD_{\min} and OBSS/PD_{\max} are set to -82 dBm and -62 dBm, respectively, the reference power $\text{TX_PWR}_{\text{ref}}$ is set to 21 or 25 dBm, based on the capabilities of the device,⁹ and TX_PWR is the transmission power at the antenna connector in dBm of the HE node that identifies the SR-based opportunity.

Note that the OBSS/PD is defined for 20 MHz PPDUs received on the primary channel, but, in general, this value depends on the bandwidth used. In particular, the OBSS/PD increases 3 dB each time the channel width is doubled, as shown in Table 1.

Channel width	OBSS/PD
40 MHz	$\text{OBSS/PD}_{20\text{MHz}} + 3$ dB
80 MHz	$\text{OBSS/PD}_{20\text{MHz}} + 6$ dB
160 MHz or $80+80$ MHz	$\text{OBSS/PD}_{20\text{MHz}} + 9$ dB

Table 1: Effect of the channel width on the OBSS/PD threshold.

⁹The $\text{TX_PWR}_{\text{ref}}$ is set to 21 dBm at HE nodes which **Highest NSS Supported M1** field is equal or less than 1 . Otherwise, the $\text{TX_PWR}_{\text{ref}}$ is set to 25 dBm. The **Highest NSS Supported M1** subfield is part of the **Tx Rx HE MCS Support** field of the **HE Capabilities element**.

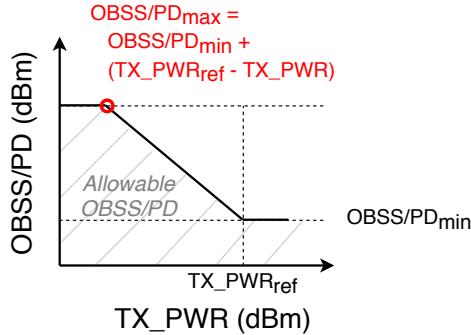


Figure 9: Graphical representation of the adjustment rules for OBSS/PD and transmission power [1].

4.1.2 SRG-based constraints

In addition to the general rules for the OBSS/PD, further constraints apply when using SRGs. In particular, an AP can define certain tolerance margins for setting both the SRG and the non-SRG OBSS/PD (see Tables 2 and 3). Those margins are referred to as minimum and maximum OBSS/PD offsets, respectively, and must verify:

- $-82 \text{ dBm} \leq -82 \text{ dBm} + \text{SRG OBSS/PD Min Offset dBm} \leq -62 \text{ dBm}$
- $\text{SRG OBSS/PD Min Offset} \leq \text{SRG OBSS/PD Max Offset}$
- $\text{SRG OBSS/PD Max Offset} + -82 \text{ dBm} \leq -62 \text{ dBm}$
- $\text{Non-SRG OBSS/PD Max Offset} + -82 \text{ dBm} \leq -62 \text{ dBm}$

OBSS/PD SR disallowed	Non-SRG Offset	Non-SRG OBSS/PD Min	Non-SRG OBSS/PD Max
Unspecified	Unspecified	-82	-62
0	0	-82	-62
0	1	-82	$-82 + \text{Non-SRG OBSS/PD Max off.}$
1	Don't care	-82	-82

Table 2: Minimum and maximum non-SRG OBSS/PD threshold (in dBm) to be used by a given HE STA, according to the information provided by the AP in parameters **OBSS/PD SR Disallowed** and **Non-SRG Offset Present**.

Note, as well, that the way of computing the exact SRG and non-SRG OBSS/PD values is not defined in the standard, thus opening the door to new contributions. In relation to this, the authors of [28] proposed using the Received Signal Strength Indicator (RSSI) of received beacons to compute it, so

SRG field	SRG OBSS/PD Min	SRG OBSS/PD Max
Unspecified	N/A	N/A
0	N/A	N/A
1	-82 + SRG OBSS/PD Min Offset	-82 + SRG OBSS/PD Max Offset

Table 3: Minimum and maximum SRG OBSS/PD values (in dBm) to be used by a given HE STA, according to the information provided by the SRG field. If SRG is not activated (or its value is unspecified), PPDU frames cannot be classified as SRG frames.

that $\text{OBSS/PD} = \text{RSSI} - \text{OBSS/PD}_{\text{margin}}$. This approach is similar to the DSC procedure described in Section 2.

4.1.3 Transmit power restriction

So far, we have referred to CST adjustment, but transmit power limitation is also an important part of the SR operation. In particular, a power restriction is imposed for any transmission occurring as a result of a detected SR opportunity (i.e., after ignoring a given inter-BSS frame through the OBSS/PD-based SR operation). By applying a power restriction, the standard aims to reduce the impact of these transmissions on other ongoing ones. The allowed transmit power is related to the OBSS/PD employed for detecting the SR opportunity. Simply put, the more inter-BSS transmissions can be ignored (by increasing the OBSS/PD), the less interference should be generated. The transmission power restriction lasts until the end of the SR opportunity identified by an HE node, which starts when its backoff reaches zero. Notice that this period depends on the duration of the active transmission(s) used for detecting the SR opportunities. The maximum allowed transmission power ($\text{TX_PWR}_{\text{max}}$) is given by:

$$\text{TX_PWR}_{\text{max}} = \text{TX_PWR}_{\text{ref}} - (\text{OBSS/PD} - \text{OBSS/PD}_{\text{min}}) \quad (1)$$

The previous equation holds for $\text{OBSS/PD}_{\text{max}} \geq \text{OBSS/PD} > \text{OBSS/PD}_{\text{min}}$. Otherwise, the maximum transmission power is unconstrained.

4.1.4 Example of OBSS/PD Spatial Reuse

To illustrate the OBSS/PD-based SR operation in detail, we propose the scenario shown in Fig. 10, from which we focus on $\text{STA}_{\text{C}2}$. In this scenario, several potential interfering devices (belonging to BSS_A and BSS_B) surround $\text{STA}_{\text{C}2}$. In particular, when using the default CCA/CS, all the APs are able to transmit simultaneously. However, $\text{STA}_{\text{C}2}$ may suffer flow starvation because of its unprivileged location.

To overcome the flow starvation issue, $\text{STA}_{\text{C}2}$ can ignore inter-BSS transmissions, provided that the appropriate OBSS/PD value is used. Through

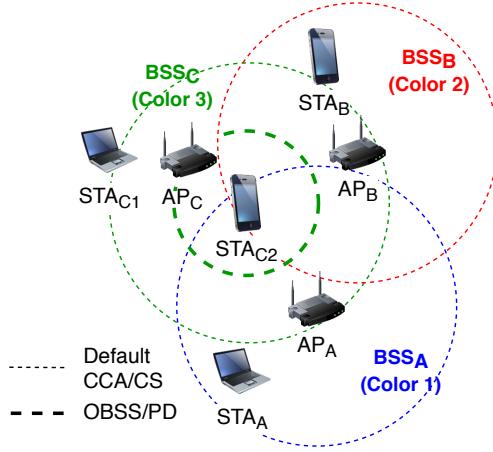


Figure 10: Scenario for showcasing the OBSS/PD SR operation.

OBSS/PD-based SR, STA_{C2} can detect SR opportunities when nodes from BSS_A and BSS_B transmit. Notice that any detected SR opportunity is subject to a power restriction. Since different OBSS/PD thresholds can be maintained for different inter-BSS transmissions (of SRG and non-SRG type), different power restrictions can be used. This has to be considered before transmitting, i.e., the most restrictive power limitation should be applied when accessing the channel.

Fig. 11 illustrates an example of packets exchange when OBSS/PD-based SR is enabled. The following particular interactions (displayed in yellow) are given:

1. STA_{C2} analyzes the RTS frame sent by AP_A and classifies it as an inter-BSS frame. As a result, it applies a given OBSS/PD value that allows sensing the channel idle ($RSSI_{A \rightarrow C2} < OBSS/PD$). However, STA_{C2} must take into account a first power restriction, which is given by Equation (1).
2. The same procedure is followed at STA_{C2} when detecting the RTS frame transmitted by AP_C . However, the transmission cannot be ignored this time because the source is an intra-BSS device (the default CCA/CS threshold is applied).
3. As for points 1) and 2), AP_B 's transmission is ignored by STA_{C2} because $RSSI_{B \rightarrow C2} < OBSS/PD$. Again, a new power restriction is considered.
4. Finally, STA_{C2} transmits by taking advantage of the detected SR opportunities. The transmission is nonetheless subject to a transmit power limitation, which is the more restrictive one among all the collected power restrictions (PRs). In particular, $TX\ PWR_{max} = \min(PR_1, PR_2)$.¹⁰

¹⁰Notice that, once STA_{C2} transmits under the power restriction, the ACK sent by STA_B

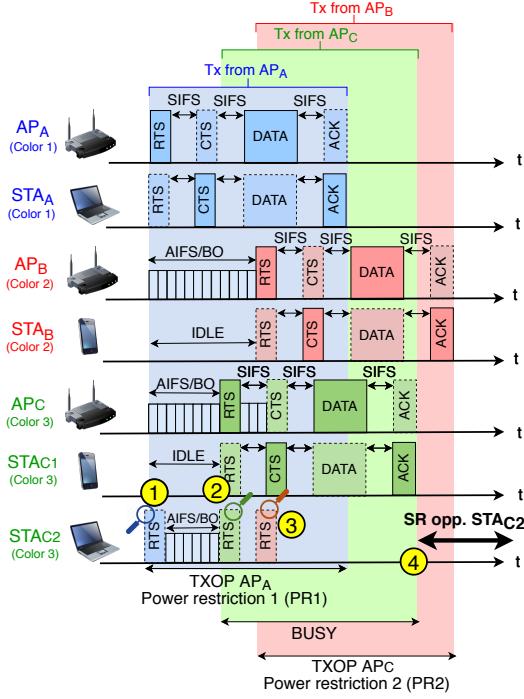


Figure 11: Example of the OBSS/PD-based SR operation. STA_{C2} applies different OBSS/PD values according to the source of detected transmissions.

4.2 Parametrized Spatial Reuse

The PSR operation is based on TB transmissions and requires cooperation among the participating BSSs. On the one hand, we find nodes taking advantage of PSR opportunities (i.e., the *opportunist*s). These nodes identify PSR opportunities from detected TB transmissions. On the other hand, we find the *transmission holders*, which perform TB transmissions and indicate support for the PSR operation. Notice that PSR opportunities can only be detected from the transmission holders that explicitly indicate support for the operation (e.g., in the headers of a TF).

When it comes to identifying PSR opportunities, an opportunist must check whether the TB PPDUs that follow a given TF packet can be ignored or not. To do so, the intended transmission power at the opportunist must not exceed the requirements imposed by the transmission holder. Those requirements are encapsulated by the latter through the `PSR_INPUT` parameter. This parameter is indicated in the TF and can take any of the discrete values shown in Table

can be ignored, so that a new power restriction is not defined.

4. The PSR INPUT is computed as follows:

$$\text{PSR INPUT} = \text{TX PWR}_{\text{AP}} + I_{\text{AP}}^{\max},$$

where $\text{TX PWR}_{\text{AP}}$ is the normalized transmit power in dBm at the output of the antenna connector, and I_{AP}^{\max} is a normalized value in dB that captures the maximum allowed interference at the transmission holder.¹¹

Value	Meaning	Value	Meaning
0	PSR_DISALLOW	8	PSR = -44 dBm
1	PSR = -80 dBm	9	PSR = -41 dBm
2	PSR = -74 dBm	10	PSR = -38 dBm
3	PSR = -68 dBm	11	PSR = -35 dBm
4	PSR = -62 dBm	12	PSR = -32 dBm
5	PSR = -56 dBm	13	PSR = -29 dBm
6	PSR = -50 dBm	14	PSR \geq -26 dBm
7	PSR = -47 dBm	15	PSR_AND_NON-SRG_OBSS-PD_PROHIBITED

Table 4: PSR subfield encoding for Trigger and HE TB PPDU frames [1].

Once an opportunist inspects the PSR value of the detected TF¹² and confirms that the intended transmission power is acceptable, it transmits during the duration of the TB PPDU(s) (indicated in the **Common Info** field). In particular, the intended transmission power must be below the value of PSR minus the Received Power Level (RPL), which is measured from the legacy portion of the TF (i.e., from PHY headers).

In order to illustrate the PSR operation, refer to the scenario that is shown in Fig. 12, where we focus on AP_A . The interference sensed in that AP from the overlapping transmitters, i.e. AP_B and AP_C , is -80 dBm and -75 dBm, respectively. According to that, Fig. 13 shows an example of packets exchange when applying PSR in AP_A . The key actions (displayed in yellow) are as follows:

1. AP_A detects a PSR opportunity from AP_B 's TF packet. Notice that the intended transmit power for the next queued packet of AP_A must be lower than the indicated PSR by AP_B minus the RPL. If so, AP_A 's backoff keeps counting down.
2. As soon as AP_C transmits a TF, the PSR opportunity previously detected by AP_A is canceled because the transmit power condition no longer holds. As a result, the channel is marked as busy and the backoff countdown is frozen.

¹¹In particular, I_{AP}^{\max} is computed as the target RSSI indicated in the TF minus the minimum SNR granting a 10% PER (based on the highest MCS to be used for transmitting the UL HE TB PPDU). A safety margin (set by the AP) is also included not to exceed 5 dB.

¹²The PSR can be extracted either from the **SPATIAL REUSE** field, which is included in the **Common Info** field of the Trigger frame, or the **SIG-A PSR** field of the HE TB PPDU.

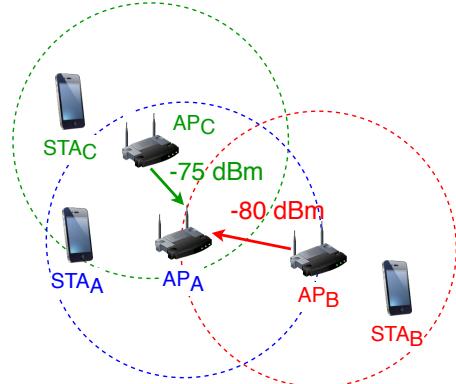


Figure 12: Scenario for showcasing the PSR operation.

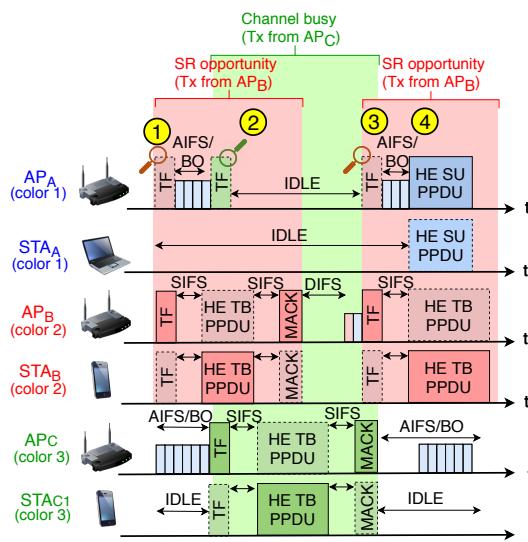


Figure 13: Packets exchange according to the PSR operation.

3. For a new transmission held by AP_B , a PSR opportunity is detected again, even if BSS_C is still transmitting. That opportunity may be used by AP_A as soon as BSS_C 's transmission finishes.
4. Once BSS_C 's transmission is over, AP_A keeps the backoff countdown and transmits according to the last detected PSR opportunity.

5 Model and Simulation of the 11ax Spatial Reuse Operation

Characterizing the IEEE 802.11ax SR operation is crucial to fully understand its implications and potential gains. However, it turns out to be a challenging task due to the complex (and still unknown) inter-BSS interactions generated by adjusting the sensitivity and the transmission power. To the best of our knowledge, none of the previous works have attempted to model the 11ax SR operation. Nevertheless, we find other works that analyze the impact of sensitivity adjustment and transmit power control in wireless networks. In this regard, we find SINR-based methods [7, 29], which allow characterizing radio links. However, these methods typically consider the worst-case interference (i.e., nodes are assumed to transmit permanently). This entails neglecting spectrum access coordination and hence losing insights on the MAC operation.

Another field that is attracting a lot of attention is Stochastic Geometry (SG), which models the random nature of dense wireless networks. In particular, SG allows defining a random set of nodes (typically, based on random point processes) and deriving statistical properties on them. In telecommunications, stochastic geometry has been widely applied to model the behavior of users and to estimate metrics such as the outage probability or the throughput per area [30]. Concerning SR, the works in [31–33] apply SG to model the effect of tuning the sensitivity threshold in WLANs. However, SG models are mainly focused on PHY layer effects and fail to capture the asymmetries that may take place on applying the SR operation.

To address the limitations posed by the abovementioned models, we propose characterizing SR through Continuous Time Markov Networks (CTMNs) [34, 35], which allow capturing both the PHY and MAC layers on estimating the throughput of a CSMA/CA network. The analytical model presented in this work aims to provide further insight into the effects of applying SR in next-generation BSSs.

Apart from the analytical model, we introduce the 11ax SR operation in the Komondor simulator [13].¹³ The Komondor simulator was conceived, among other purposes, to allow the low-cost integration of novel mechanisms included in new IEEE 802.11 standards. This is the case of 11ax SR, which has not been yet fully implemented in any other well-known simulator. At the time of publishing this article, SR is still being developed for ns-3.¹⁴ By comparing our simulation results with the analytical model, we expect to shed some light on the effects of using 11ax SR, particularly with regard to inter-BSS interactions. The simulation of SR allows addressing the assumptions done by the CTMNs model and extend the provided results in more realistic dense environments.

Before getting into the analysis of 11ax SR through CTMNs, it is important

¹³The implementation of SR can be found in Komondor v3.0, available in <https://github.com/wn-upf/Komondor/releases/tag/v3.0>.

¹⁴It is planned to be included in the following repository: <https://gitlab.com/nsnam/ns-3-dev>.

to mention that we have focused on the OBSS/PD-based operation described in Section 4.1, which has drawn much more attention than PSR. Therefore, from now onwards, we may refer to the OBSS/PD-based SR operation simply as SR. The implementation and modeling of PSR is left as future work.

5.1 Introduction to Continuous Time Markov Networks

The CTMN model captures the CSMA/CA operation used in IEEE 802.11 WLANs through states, which represent the set of potentially overlapping BSSs that are active at a given moment. Transitions between states occur when BSSs become active (i.e., they gain access to the medium) or when they abandon the channel (i.e., their transmission is finished). It is worth pointing out some assumptions made by the CTMN model. First, downlink traffic is considered and the interference produced in uplink transmissions (e.g., ACKs) is not considered. Second, the backoff procedure for accessing the medium is continuous in time. Thus, collisions due to backoff expiring at the same instant are not captured by the model. Despite those are unrealistic assumptions, the model is particularly useful to depict inter-AP interactions.

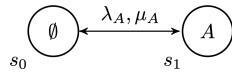


Figure 14: CTMN of BSS_A .

For the sake of illustration, let us consider Fig. 14, which represents the CTMN of a single BSS, namely BSS_A . In the CTMN, s_0 is the empty state (the channel is idle) and s_1 indicates that BSS_A is transmitting in a given channel. Regarding the transition rates between states, we find two different types: *i*) AP activates, and *ii*) AP finishes a transmission. While *i*) is related to the necessary time for a given node to access the channel (characterized by the arrival rate λ), *ii*) depends on the time spent by a given node for transmitting data (characterized by the service rate μ). Based on the transition probabilities, one can obtain the probability of every state, which allows computing the long-term throughput experienced by the different BSSs.

In this work, the 11ax SR operation has been implemented as part of the Spatial Flexible Continuous Time Markov Network (SFCTMN) framework [10, 36, 37].¹⁵ This framework allows generating the CTMN of a given WLAN deployment, based on the spatial distribution of nodes and their configuration (e.g., range of channels used, transmission power, sensitivity, etc.). It is important to remark that additive interference is considered, which results from the combination of different simultaneous interfering transmissions. Accordingly, we are able to characterize real deployments where spatially-distributed interactions occur. Moreover, traffic is considered to be saturated in all the nodes, so that pure SR-based interactions become more apparent.

¹⁵A dedicated Github branch of SFCTMN has been provided for single-channel spatial reuse [38].

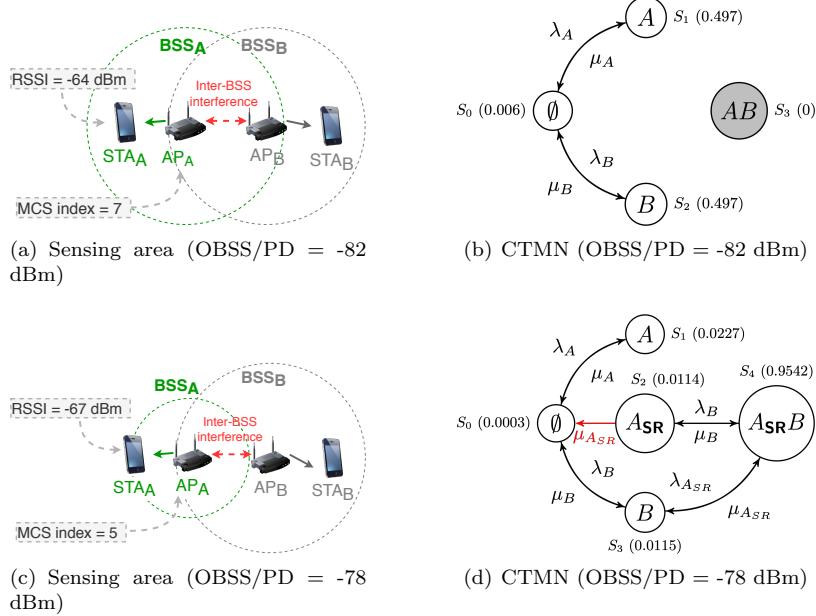


Figure 15: Representation of *Toy scenario 1* for different settings. (a) and (c) illustrate the inter-BSS sensing area of each transmitter for OBSS/PD equal to -82 dBm and -78 dBm, respectively, whereas (b) and (d) illustrate the corresponding inter-BSS interactions through CTMNs (unidirectional transitions are shown in red).

Concerning the 11ax SR operation, we have considered new states to represent the different sensitivity levels that each BSS can use based on the type of ongoing transmissions. Furthermore, the maximum allowed transmission power depends on the selected OBSS/PD threshold, which is also captured in the new SR states. Accordingly, the varying transmission capabilities of a given node (which depend on the MCS used) are represented.

5.2 Simple inter-BSS interactions

Through our proposed SR model, we start depicting the inter-BSS interactions that occur between two BSSs. With this aim, we introduce *Toy scenario 1*, in which only BSS_A implements SR. Fig. 15(a) and Fig. 15(c) illustrate the carrier sense areas that result from the default and the SR settings, respectively. The CTMNs capturing the inter-BSS interactions taking place in each setting are shown in Fig. 15(b) and Fig. 15(d), respectively. Notice that the long-term probability of each state is shown in parentheses.

As shown in Fig. 15(a) (CCA/CS is applied by both BSSs), AP_A and AP_B are in the carrier sense range of each other, so that parallel transmissions are not

possible. This can also be noticed in the CTMN representation (see Fig. 15(b)), where state s_3 (AB) cannot be reached from any other state. Nevertheless, both APs can transmit at a high rate because the maximum transmission power is used when accessing the medium. In particular, the STA in BSS_A observes an RSSI of -64 dBm, which allows AP_A using the MCS 7 for 20 MHz transmissions.

When it comes to the SR setting, it is possible to have simultaneous transmissions, provided that BSS_A uses an OBSS/PD value greater or equal than -79 dBm. As shown in Fig. 15(c), AP_A reduces its sensitivity area in case of detecting any transmission from BSS_B . However, ignoring inter-BSS transmissions entails a transmit power limitation. This results in poorer signal strength at the STA (RSSI = -67 dBm when the transmit power used by AP_A is 17 dBm), thus forcing to use a lower data rate. The SR operation is represented through the CTMN's model in Fig. 15(d), where new states appear (i.e. s_2 and s_4). These new states capture the situations in which the transmitter of BSS_A uses a higher OBSS/PD to ignore BSS_B 's transmissions (mode A_{SR} is used, instead of A). In particular, state s_2 (A_{SR}) can never be reached from the empty state since BSS_A is always expected to transmit under its default operation when the channel is idle.

The throughput achieved by BSS_A and BSS_B is shown Fig. 16 (left side), for each possible OBSS/PD value. The transmit power limitation that is imposed on each BSS is also illustrated (right side).

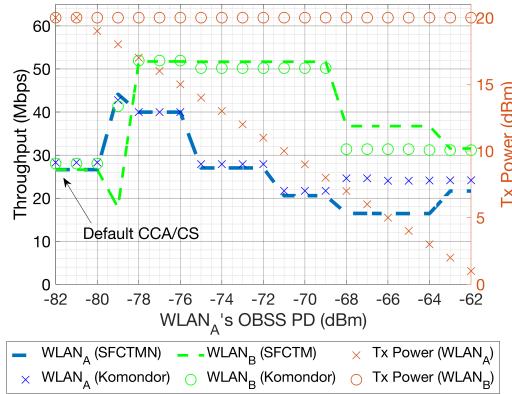


Figure 16: Effects of applying OBSS/PD-based SR in BSS_A of *Toy scenario 1*, for each possible OBSS/PD value. The transmission power is shown in red. Results are shown for both SFCTMN and Komondor.

As shown, both BSSs obtain the same performance for $\text{OBSS/PD} < -79$ dBm (they share the channel). To that point onwards, BSS_A is able to ignore BSS_B 's transmissions due to the SR operation. However, what might seem a worthy strategy for BSS_A turns out to be more beneficial to BSS_B . The latter,

except for OBSS/PD = -79 dBm,¹⁶ enjoys the highest possible throughput when BSS_A applies the SR operation. The fact is that BSS_A is forced to use a lower transmission power in case of transmitting when BSS_B is occupying the channel. Therefore, BSS_B will keep sensing the channel idle once its transmission finishes, provided that BSS_A is still subject to the transmission power restriction.

It is important to note that, in Fig. 16, there is a region (from OBSS/PD = -68 dBm to OBSS/PD = -64 dBm, both included) in which the SFCTMN is less accurate at capturing the actual OBSS behavior on using SR. For these OBSS/PD values, STA_A cannot decode any transmission from AP_A in state $A_{SR}B$. In turn, it can do that for the A_{SR} state. In particular, the transmit power limitation used by AP_A in the SR mode makes that STA_A perceives an insufficient signal-to-noise-plus-interference ratio (SINR) when BSS_B is also occupying the channel. The main reason is that the SFCTMN model considers that the throughput obtained in every state is independent of the others, and this condition does not hold for states $A_{SR}B$ and A_{SR} . In reality, AP_A is expected to abandon its transmission in state $A_{SR}B$ as soon as a timeout is noticed, thus spending a few time in the SR mode (transition $A_{SR}B$ to A_{SR} is unlikely). In contrast, the SFCTMN considers that much more time is spent in state A_{SR} since transmissions at that point are successful (but slow due to the low MCS used).

Now, let us consider the case where both BSSs apply the SR operation simultaneously. The CTMN for OBSS/PD \geq -79 dBm is shown in Fig. 17. For the sake of illustration, only transitions between states s_0 and s_1 are provided. As shown, both BSSs can act by using the default or the SR modes, thus generating a symmetric CTMN. In particular, we find two dominant states: s_5 and s_6 . These states are visited with the same probability (0.4482), which indicates that both BSSs alternate the default and the SR modes, thus obtaining the same throughput. However, in reality, one of the BSSs may monopolize the channel through under the default mode, while the other operates under the transmit power-constrained SR mode.

This phenomenon is properly captured in the Komondor simulator, where SR opportunities are identified on a per-packet basis. In this case, the BSS that accesses to the channel for the first time (e.g., BSS_A) is most likely to enjoy the maximum throughput. In contrast, the other BSS (e.g., BSS_B) transmits under the SR mode almost all the time (as a result of BSS_A 's activity), until they alternate roles. Notice that a single state between s_5 and s_6 is more likely to be monopolized as the transmission time becomes longer than the idle periods. In our case, we have very long transmission times in comparison to the idle time since we assume full-buffer traffic, packet aggregation, and short contention window (CW) values.

Fig. 18 shows the throughput achieved by each BSS when both apply SR, and for each OBSS/PD threshold. Again, the results have been extracted from both SFCTMN and Komondor. In order to show the long-term performance

¹⁶At that point, the transmission power limitation for OBSS/PD = -79 dBm is insufficient, so BSS_B senses the channel busy when BSS_A transmits under the SR mode.

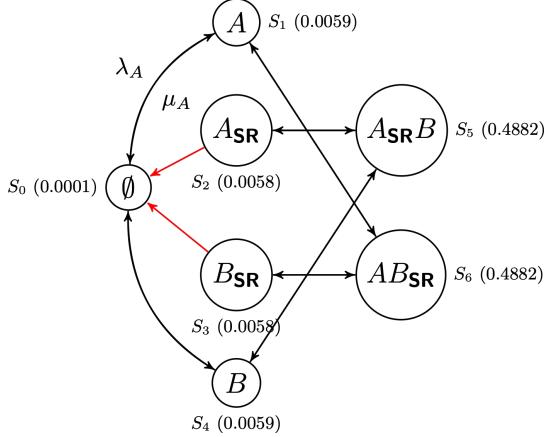


Figure 17: CTMN of *Toy scenario 1* when both BSSs apply OBSS/PD-based SR with $\text{OBSS}/\text{PD} \geq -79 \text{ dBm}$ (unidirectional transitions are marked in red).

of each BSS in the Komondor simulator, we have displayed the average values obtained from 100 simulations. As shown, the performance achieved by both BSSs is totally fair, due to the symmetry of the scenario. In particular, states in which SR is used are alternated, thus allowing each BSS to access the channel while the other is transmitting. As a result, the throughput of both BSSs can be further increased with respect to the case in which only one BSS applies SR. However, unlike the previous case, the long-term throughput never reaches the maximum possible throughput in isolation (the transmission power limitation prevents to do so).

5.3 Interactions among Spatial Reuse groups

Differentiating between SRGs may potentially enhance spectral efficiency due to the further inter-AP interactions that can be generated with an additional OBSS/PD threshold. In practice, devices belonging to the same SRG use a dedicated OBSS/PD threshold, namely SRG OBSS/PD. For the rest of inter-BSS transmissions, the non-SRG OBSS/PD threshold is used instead. One possible use case may lie in residential building apartments, where BSSs belonging to the same building form an SRG. For the rest of networks (e.g., public Wi-Fi in the street), other SRGs can be considered.

To illustrate the implications of using SR based on SRGs, let us focus on *Toy scenario 2*, which is depicted in Fig. 19(a). In this deployment, all the BSSs apply the SR operation and two different SRGs are created. In particular, BSSs belonging to the same SRG (i.e., BSS_A and BSS_B) are close to each other, such as in a residential building. Apart from that, BSS_C , belongs to another SRG.

The result of jointly applying OBSS/PD-based SR in *Toy scenario 2* is illustrated in Fig. 19(b), which plots the throughput achieved by each of the

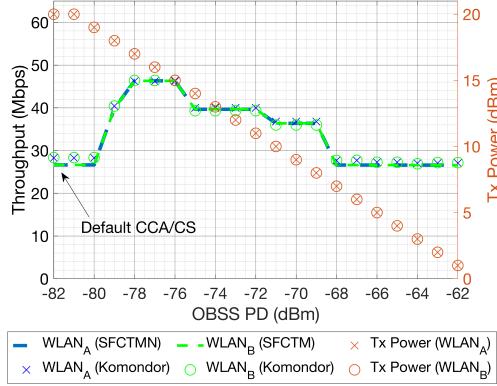


Figure 18: Effects of applying OBSS/PD-based SR in both BSSs of *Toy scenario 1*, for each possible OBSS/PD value. The transmission power is shown in red. Results are shown for both SFCTMN and Komondor.

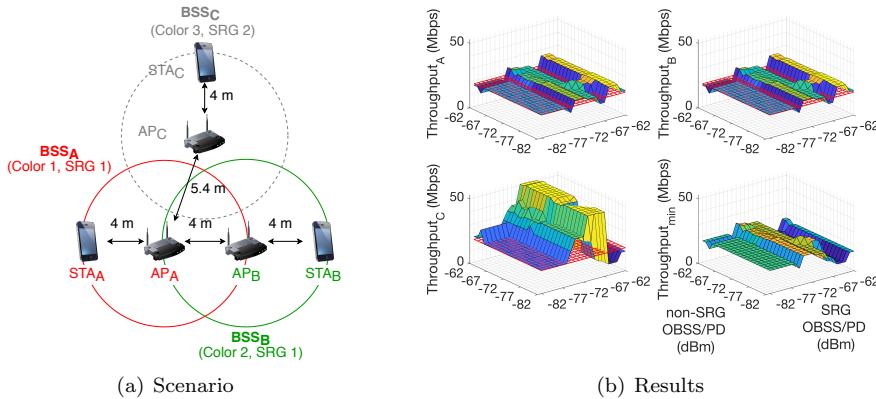


Figure 19: Results of applying the OBSS/PD-based SR operation in *Toy scenario 2*. In (b), the individual and max-min throughput are shown for each SRG and non-SRG OBSS/PD threshold. The red mesh indicates the performance achieved by using the default CCA/CS.

three BSSs, for each combination of SRG and non-SRG OBSS/PD thresholds. Notice that we have considered that all the BSSs use the same OBSS/PD values since the number of total combinations grows exponentially and is unfeasible to be plotted.

As shown, the throughput achieved by each BSS follows an irregular pattern due to the complex inter-BSS interactions that take place in this scenario. Moreover, it can be appreciated the clashing interests of each BSS, where the

individual performance is sometimes maximized at the expense of reducing the throughput of the others. For instance, if we focus on BSS_C , it obtains the maximum throughput when flow starvation is generated to BSS_A (the same occurs for BSS_B). However, this is not optimal in terms of fairness. The CTMN resulting from the flow starvation situation is shown in Fig. 20, which is given when all the BSSs use non-SRG OBSS/PD = -82 dBm and SRG OBSS/PD = -73 dBm.¹⁷

In case of considering the optimal max-min performance¹⁸, a completely different situation is observed. In this case, the max-min throughput is increased when every BSS can overtake a single detected inter-BSS transmission (regardless of its source) and access to the channel. This situation is fair and at the same time increases the overall performance. However, it occurs when all the inter-BSS transmissions are equally treated. Using SRGs can therefore improve the performance of certain nodes (belonging to the same group), but potentially leads to unfairness.

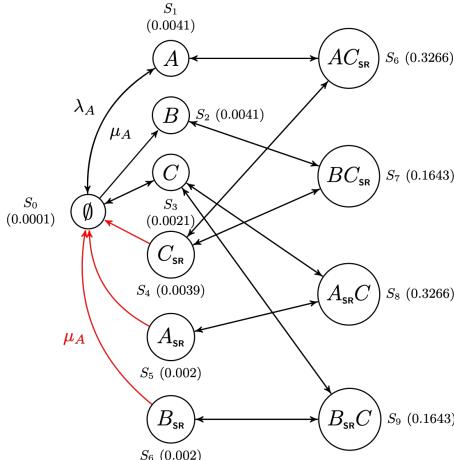


Figure 20: CTMN of *Toy scenario 2*, for non-SRG OBSS/PD = -73 dBm and SRG OBSS/PD = -82 dBm. The unidirectional transitions are marked in red, and subindex *SR* indicates the use of the non-SRG OBSS/PD threshold.

Table 5 provides a verification, for both SFCTMN and Komondor, of the results obtained in *Toy scenario 2*. For the sake of representation, we show the Root Mean Square Error (RMSE) for all the considered SRG and non-SRG OBSS/PD threshold values. As shown, the error for BSS_A and BSS_B is relatively small. In contrast, a higher error is obtained for BSS_C . This is

¹⁷The CTMN model captures the utilization of different OBSS/PD thresholds by considering that each BSS acts in three different ways (states), as a result of the employed OBSS/PD threshold: *i*) default CCA/CS, *ii*) SRG OBSS/PD, and *iii*) non-SRG OBSS/PD.

¹⁸The max-min throughput corresponds to the solution that maximizes the minimum throughput achieved by a set of BSSs.

strongly related to the fact that BSS_C belongs to a different SRG than BSS_A and BSS_B , which leads to different inter-BSS interactions. Moreover, dominant states may lead to situations that cannot be captured by the SFCTMN, as previously shown for *Toy scenario 1*. In particular, BSS_C in *Toy scenario 2* is prone to participate in these states because of its asymmetric location with respect to BSS_A and BSS_B .

	BSS_A	BSS_B	BSS_C
RMSE (Mbps)	6.02	6.03	18.42

Table 5: Verification of the results obtained in *Toy scenario 2* from the SFCTMN and Komondor.

6 Performance Evaluation

In this Section, we study the potential gains of SR in large-scale WLAN scenarios. With this aim, we leave the CTMNs-based analysis out and concentrate on simulation results. For the rest of this Section, each BSS is considered to be composed by an AP and a single STA, which are placed uniformly at random, as shown in Fig. 21.

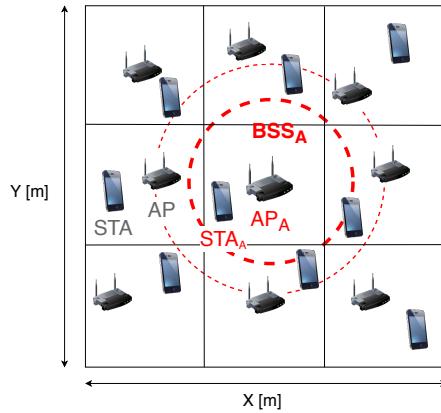


Figure 21: Random grid scenario containing 9 BSSs. The location of BSS_A is fixed at the center for the sake of analysis.

The simulation parameters are provided in Table 6. The scenario is divided into 9 cells, but the location of BSS_A is always fixed at the center of the scenario. For the rest of the APs and STAs, their position is randomly selected within their corresponding cell. The configuration of each BSS is set homogeneously, i.e., they all use the same channel, the default sensitivity is set to -82 dBm, and the default transmission power is set to 20 dBm. Notice that, for dense

deployments, BSS_A is expected to suffer a higher level of interference than the others, which allows us to assess the effectiveness of the SR operation in crowded environments.

	Parameter	Value
PHY	Central frequency, f_c	5 GHz
	Transmission gain, G_{tx}	0 dB
	Reception gain, G_{rx}	0 dB
	Path-loss (residential scenario), $\text{PL}(d)$	See ([39])
	Background noise level, N	-95 dBm
	Legacy OFDM symbol duration, σ_{leg}	4 μs
	OFDM symbol duration (GI-32), σ	16 μs
	Number of subcarriers (20 MHz), N_{sc}	234
	Number of spatial streams, N_{ss}	1
	Transmit power levels, \mathcal{T}	1 to 20 dBm (1 dBm steps)
MAC	Empty slot duration, T_e	9 μs
	SIFS duration, T_{SIFS}	16 μs
	DIFS/AIFS duration, $T_{\text{DIFS/AIFS}}$	34 μs
	PIFS duration, T_{PIFS}	25 μs
	Legacy preamble duration, $T_{\text{PHY-leg}}$	20 μs
	HE single-user field duration, $T_{\text{HE-SU}}$	100 μs
	ACK duration, T_{ACK}	28 μs
	Block ACK duration, T_{BACK}	32 μs
	Size OFDM symbol (legacy), $L_{s,l}$	24 bits
	Length of data packets, L_d	12,000 bits
	No. of frames in an A-MPDU, N_{agg}	64
	Length of an RTS packet, L_{RTS}	160 bits
	Length of a CTS packet, L_{CTS}	112 bits
	Length of service field, L_{SF}	16 bits
Misc.	Length of MAC header, L_{MH}	320 bits
	Contention window (fixed), CW	15
	Allowed sensitivity levels, \mathcal{S}	-82 to -62 (1 dBm steps)
	Traffic model, Λ	Downlink (UDP)
	Traffic generation ratio, l	1,000, 2,000, 10,000 pkts/s
	Map area (random scenario), A	625, 400, 225, 100 m^2

Table 6: Simulation parameters.

6.1 Network Density

To analyze SR based on the network density, we consider four different map sizes: sparse (25×25 m), semi-dense (20×20 m), dense (15×15 m) and ultra-dense (10×10 m). For each type of scenario, we provide 50 different deployments, in which APs and STAs are placed uniformly at random within their corresponding cell. BSS_A is the only one applying the SR operation. Since we compute all the possible OBSS/PD values to be used by BSS_A , each random deployment leads to $21 \times 4 \times 50 = 4,200$ different scenarios.

Figure 22 shows the average throughput achieved under the default and SR settings. In particular, we differentiate between the individual throughput of BSS_A and the average throughput of the other BSSs. For each network density,

we have tried all the possible OBSS/PD values and compared the best one to the default CCA/CS.

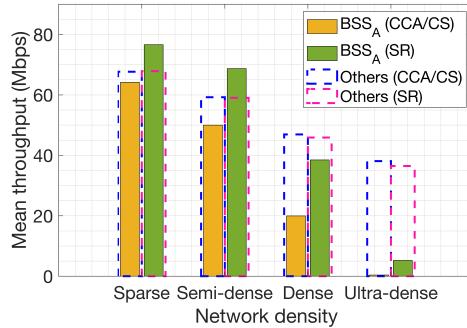


Figure 22: Mean throughput achieved with and without applying the SR operation in BSS_A, for each network density. Results show the mean throughput achieved by BSS_A and the rest of BSSs.

First of all, if we focus on the throughput that BSS_A experiences by default (amber solid bars), we notice a dramatic decrease as network density increases. Nevertheless, the SR operation allows BSS_A to significantly improve the throughput (displayed by the green solid bars). Note, as well, that the maximum improvement is experienced at the dense scenario (15×15 m). While the default performance is quite high for sparser scenarios, channel reutilization cannot be further improved at the ultra-dense scenario due to the high level of inter-BSS interference.

Apart from that, we observe that the average performance of the other BSSs (dashed bars) does not suffer radical changes for any of the network densities when BSS_A applies SR. This is a really positive result, which indicates that SR allows improving the individual performance without affecting the rest of devices that do not apply the operation.

6.2 Traffic load

Besides network density, traffic load is another interesting factor to be studied with regards to the SR operation. To that purpose, we focus on the second densest scenario, which has been previously shown to achieve the maximum gains of the SR operation. In particular, we provide three different traffic loads (l), which are the same for all the BSSs: *i*) low (1,000 packets/s, i.e., 12 Mbps), *ii*) medium (2,000 packets/s, i.e., 24 Mbps), and *iii*) high (10,000 packets/s, i.e., 120 Mbps). The traffic type considered is UDP in the downlink, which follows a Poisson distribution with λ equal to the traffic load considered in each case.

Fig. 23 compares the performance achieved by default and SR configurations, for the different considered traffic load values. As done before, the results show the individual performance of BSS_A and the average performance of the

rest of BSSs. In particular, Fig. 23(a) shows the maximum improvements achieved by BSS_A in terms of throughput. Notice that the SR configuration considers the OBSS/PD values that maximize BSS_A 's throughput. Based on that configuration, Fig. 23(b) shows the average channel occupancy (in %).

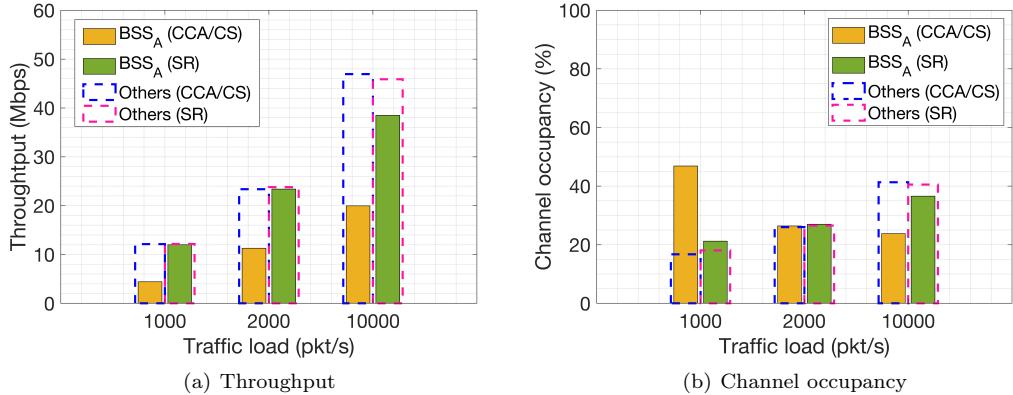


Figure 23: Mean throughput and channel occupancy achieved with and without applying the SR operation in BSS_A , for each traffic load. Results are shown for BSS_A and for the rest of BSSs (others).

As shown in Fig. 23(a), BSS_A obtains higher throughput gains as the traffic load increases. In particular, the highest gain is noticed for the largest traffic load (10,000 packets/s), which entails a saturation regime. This is a quite remarkable result since the interference noticed by BSS_A is much higher when all the surrounding devices are constantly transmitting due to their high traffic load. Regarding channel occupation (shown in Fig. 23(b)), an interesting phenomenon is observed for the lowest traffic load. The fact is that the legacy CCA/CS configuration provides a higher channel occupancy than the SR one. However, this is not translated into a higher throughput, due to the high number of experienced collisions. Notice that collisions entail a high number of re-transmissions, which cause the observed increase in the occupancy. Finally, it is worth pointing out that the performance of the other BSSs is not affected in case BSS_A applies SR.

6.3 Joint Spatial Reuse Operation

So far, we have studied the effects of applying SR at a single BSS (i.e., BSS_A). Now, we assess the potential of the joint operation by defining different situations based on the number of BSSs that apply SR. Provided that BSS_A always applies the SR operation, we propose three study cases:

- **Legacy:** all the other BSSs employ the default CCA/CS.

- **Mixed SR:** at the beginning of the simulation, each BSS randomly decides (with same probability) whether to apply the SR operation or to remain using the default configuration.
- **All SR:** all the BSSs apply the SR operation.

In order to compare the effects of applying SR in parallel with other BSSs, we define the following metrics: *i*) throughput (Γ), *ii*) percentage of time occupying the channel (ρ), and *iii*) average delay for transmitting a packet once it arrives at the queue (d). For each metric, we consider the performance improvements achieved by BSS_A (indicated with subindex A), and the average across the rest of BSSs (indicated with subindex O).

Fig. 24 shows the potential improvements achieved when applying SR in each of the proposed scenarios. While Fig. 24(a) shows the performance of BSS_A, Fig. 24(b) focuses on the performance of the others. For that purpose, the empirical cumulative distribution function (CDF) is used for each of the performance metrics. Notice that we have considered the densest scenario ($25 \times 25\text{m}$) and the highest traffic load (10,000 packets/s), thus representing the worst-case situation. As done before, we have generated 50 random scenarios for averaging purposes, and, for each of them, we have tried all the possible OBSS/PD values to be used homogeneously by the BSSs applying the SR operation. Accordingly, we have used the best value to extract the maximum average improvement of SR with respect to the legacy configuration. In every situation (*legacy*, *mixed* and *all SR*), we select the best OBSS/PD threshold from BSS_A's point of view, which is also used to assess its impact on the others. Again, the SR configuration used for the channel occupancy is the one whereby BSS_A's throughput is maximized.

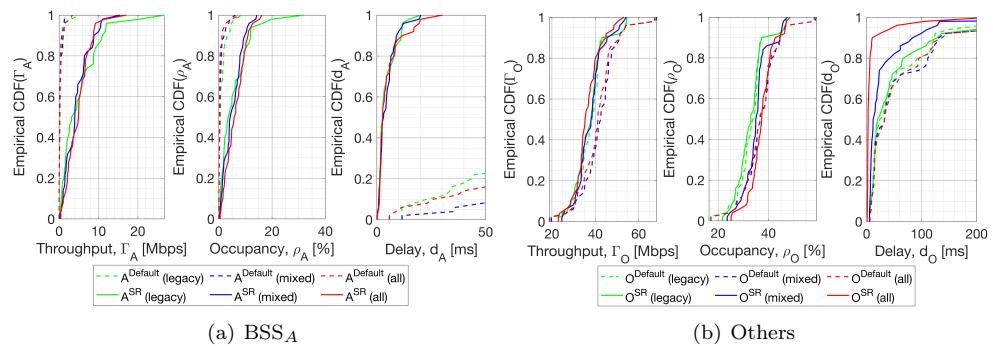


Figure 24: Mean performance improvements achieved for each SR setting by BSS_A (A) and the others (O). The results are shown for the OBSS/PD values that maximize the performance of BSS_A.

As shown in Fig. 24(a), BSS_A achieves similar performance improvements, regardless of whether the environment applies SR or not. In particular, a high

gain is noticed on the average delay. Moreover, regarding the others' performance (Fig. 24(b)), a null improvement is observed on the throughput, even for the *all SR* context. In contrast, the delay is notably reduced as the number of BSSs using SR increases.

7 Ways Forward and Research Opportunities

The IEEE 802.11ax SR operation can potentially increase spectral efficiency in dense deployments. However, it is already in a premature stage and further developments are expected to sustain progress towards next-generation wireless deployments.

7.1 Unexplored Areas within the Spatial Reuse Operation

In the context of the SR operation, the following areas have not been fully exploited yet:

- **Assignment of BSS colors:** as discussed in Sections 3.1 and 4.1, BSS coloring is key for the OBSS/PD-based SR operation since it allows differentiating between intra and inter-BSS frames. However, the way BSS colors are assigned to BSSs is not specified, thus leading to potential collisions and miss-behaviors regarding the SR operation.
- **Election of SRGs:** similarly to the BSS color, the SRG is used to sub-classify inter-BSS frames, so that different PD policies can be applied to increase spectral efficiency. However, forming SRGs is not trivial since inter-BSS interactions must be carefully captured to properly taking advantage of the SR operation. The set of policies regarding SRGs may be decided by the APs, as a result of monitoring phases.
- **Establishment of OBSS/PD thresholds:** the election of OBSS/PD thresholds for each type of frame (SRG, and non-SRG) entails a set of trade-offs. On the one hand, too low values may lead to null improvement, thus framing the legacy operation whereby the channel is shared. On the other hand, too high values may generate performance anomalies such as the hidden-terminal problem or flow starvation. A potential solution to properly establish each OBSS/PD threshold is to capture all the inter-BSS interactions on a per-STA basis.
- **Optimal transmit power:** the current transmit power restriction is useful to prevent the accentuation of unfair situations. However, the performance of the SR operation may be further increased in case of properly leveraging the transmit power according to the noticed interactions among nodes.
- **Disabling the SR operation:** there are situations in which the SR operation may be harmful to certain devices (e.g., in terms of fairness).

Therefore, a given BSS must be able to identify whether the SR operation must be disabled or not. This can be achieved by setting the OBSS/PD threshold to the default CCA/CS value. Alternatively, the SR operation can be disabled at STAs only, thus leading to an AP-only SR setting. In this regard, AP-AP interactions would be mostly targeted.

Solving most of the aforementioned problems is not straightforward and requires an in-depth analysis to offer optimal or close-to-optimal solutions. While BSS color assignment may appear to be straightforward (e.g., through graph coloring techniques), defining OBSS/PD thresholds is a very complex task that embraces many variables. In particular, inter-BSS interactions have been shown in this paper to significantly vary depending on the selected OBSS/PD values. Since the performance of IEEE 802.11 BSSs is not linear with the sensitivity and the transmission power (due to the nature of CSMA/CA), the optimal OBSS/PD threshold cannot be computed explicitly. Notice that the number of total combinations in an N-BSS scenario is $C = 21^N$, for 21 different non-SRG OBSS/PD thresholds. Therefore, the problem is intractable. If considering SRGs, the problem becomes even more complex since the number of combinations is $C = (21 \times 21)^N$ (provided that we have 21 values to be used for SRG and non-SRG OBSS/PD thresholds).

7.2 Integration of the Spatial Reuse Operation with other Techniques

In addition to problems specific to the SR operation, the integration with many other novel mechanisms remains unexplored. Among them, we highlight OFDMA [40, 41], multiple antenna systems [42], and scheduled transmissions [43]. The potential of SR goes further when combined with other techniques.

For instance, the combination of SR with directional transmissions may lead to efficient and performance maximizing communications, where SR is applied on a per-beam basis. Fig. 25 devises the potential of combining SR with directional transmissions. As illustrated, BSS_A applies the SR operation on a per-beam basis, while BSS_B remains using the default CCA/CS. In particular, collisions by hidden-node may be experienced for STA_{A3} , in case of using the inter-BSS OBSS/PD. However, channel reuse can be enhanced for transmissions to STA_{A1} and STA_{A2} , which are out of range of AP_B . Therefore, the inter-BSS OBSS/PD can be used only for transmissions involving those two STAs, while a more conservative threshold can be employed for STA_{A3} .

Similarly to the integration with directional antennas, the potential of SR can be further exploited through TB communications. In this case, users of a given BSS can be categorized into different types, so that different inter-BSS OBSS/PD values are assigned to them. Figure 26(a) shows how users can be grouped based on different OBSS/PD thresholds. As a result, transmissions within the same BSS can be scheduled in a differently, thus improving spectral efficiency. In the proposed example, STA_{A1} and STA_{A2} belong to the first group

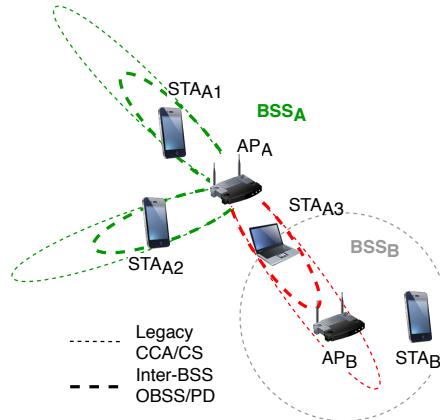


Figure 25: Potential application of SR combined with directional transmissions.

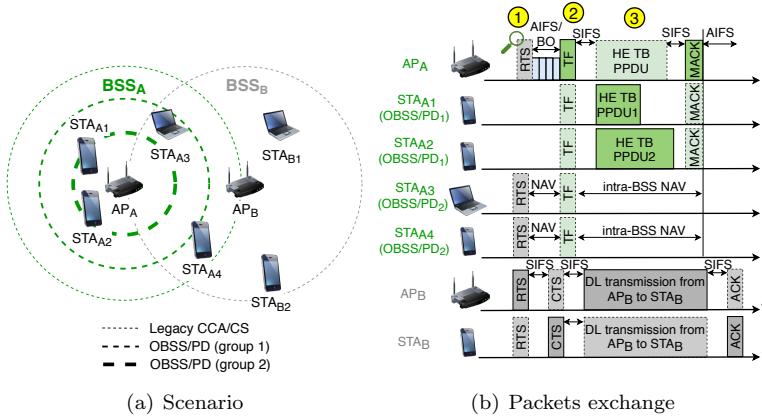


Figure 26: Potential application of SR combined with TB communications.

because of their privileged position with respect to AP_A . Therefore, a more aggressive OBSS/PD threshold is employed at the time of scheduling transmissions to these stations. The same reasoning can be applied to STA_{A3} , which, in this case, requires the usage of a more conservative OBSS/PD threshold for being scheduled in combination with SR. Finally, the legacy CCA/CS is used for STA_{A4} , in order to prevent negative interactions with respect to BSS_B . It is worth pointing out that users belonging to different groups can be scheduled together, provided that the most restrictive OBSS/PD threshold is used.

In Fig. 26(b), we show a data transmission resulting from the combination of TB communications and SR. In the yellow point #1, AP_A detects an inter-BSS transmission from AP_B , which can be ignored by using the most aggressive

OBSS/PD, i.e., the one devoted for STAs in group 1. Accordingly, it schedules an uplink transmission from STA_{A1} and STA_{A2} (yellow point #2). Finally, AP_A receives the scheduled transmissions from group 1 (yellow point #3).

7.3 Artificial Intelligence to Address Spatial Reuse Optimization

In light of the challenges posed by the 11ax SR operation, Artificial Intelligence (AI) emerges as a potential solution. In particular, WLANs are characterized by being highly varying in terms of users and channel dynamics. Moreover, we typically find decentralized deployments, at which none or little coordination is allowed. Hence, online learning stands as a suitable technique to address the optimization of SR in WLANs. In fact, many works on OBSS/PD adjustment, such as DSC [44] and COST [26], are based on iterative methods.

Machine Learning (ML), and more precisely Reinforcement Learning (RL), can contribute to improving the performance of the already existing methods. RL has been shown to properly fit with the decentralized nature of IEEE 802.11 WLANs [10, 20, 45–48]. In particular, the usage of RL allows capturing subtle information that cannot be predicted before-hand (for instance, regarding inter-BSS interactions). Such information enables conducting a learning-based procedure, which is aimed at increasing performance while reducing the number of undesired situations (e.g., poor fairness).

8 Conclusions

In this paper, we have provided an extensive tutorial of the IEEE 802.11ax SR operation, which aims to maximize the performance of next-generation WLANs by increasing the number of parallel transmissions. Our purpose has been to do so in a clear and easy-to-understand manner. Thus, significant efforts have been made in providing meaningful examples of the different specifications related to SR.

Apart from the tutorial, we have modeled the SR operation analytically using CTMNs. Through this model, we have analyzed the new kind of inter-BSS interactions that may result from applying SR in an OBSS. In particular, we have considered BSSs with a single STA, but more complex interactions are expected to happen when applying the SR operation in BSSs with multiple STAs. Apart from the analytical analysis, we have implemented the 11ax SR operation in the Komondor simulator. The potential of SR in large-scale scenarios has been evaluated through extensive simulations.

Besides significant improvements are achieved by the SR operation, other important aspects have been identified. First of all, it is important to highlight the non-intrusive characteristic of the SR operation. In particular, devices using SR can increase their performance without affecting other overlapping networks or preventing them to transmit. This is a key feature to sustain performance

growth. Moreover, the SR operation has been shown to perform better in scenarios with a high level of interference, i.e., high-density scenarios with a high traffic load. This confirms the utility of the SR for dense next-generation wireless networks.

However, finding the best SR configuration is far from trivial (it is a combinatorial problem), and remains an open problem to date. Indeed, the 11ax amendment does not provide any specification and/or guidelines on this matter. We left as future work the design of mechanisms able to find the optimal parameters within the IEEE 802.11ax SR operation. For that purpose, the usage of RL can be particularly targeted. In addition, SR can evolve and be combined with other novel techniques such as directional transmissions or distributed OFDMA, so that further performance gains can be achieved.

Appendices

A IEEE 802.11ax Frames

In this Section, we introduce the type of frames that are considered in the 11ax amendment. Such information is key to properly understand the SR operation.

A.1 HE PPDU formats

Below, we briefly describe the Physical Protocol Data Unit (PPDU) formats available in the 11ax:

- SU (Single User) HE PPDU: are meant for single user communications.
- HE Extended Range HE PPDU: are meant for single user long-range transmissions, hence only contemplate 20 MHz bandwidths in a single spatial stream.
- MU (Multi-User) HE PPDU: due to the OFDMA operation, such kind of PPDUs are meant for multiple transmissions to one or more users.
- Trigger-Based (TB) HE PPDU: in this case, MU UL transmissions are scheduled by the AP, which decides which STAs are expected to transmit during a specific elapse of time. The TB HE PPDUs can make use of OFDMA and/or MU-MIMO.

The new fields included in the abovementioned HE PPDU formats are HE Signal A Field (HE-SIG-A), HE Signal B Field (HE-SIG-B), HE Short Training Field (HE-STF), and HE Long Training Field (HE-LTF), which are shown in Fig. 27.

Among the new fields, we highlight HE-SIG-A, which includes the following elements related to the SR operation:

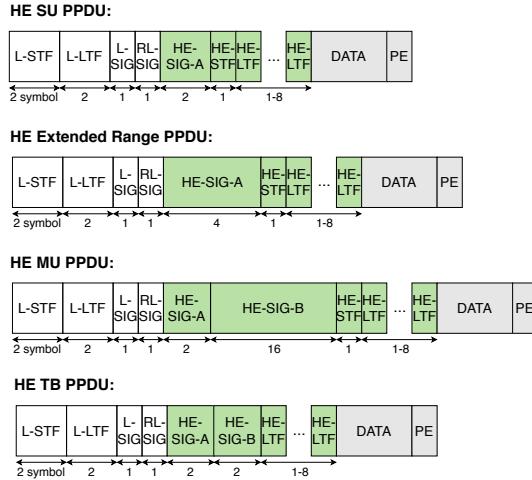


Figure 27: HE PPDU formats. New IEEE 802.11ax fields are highlighted in green.

- BSS color: it is used as an identifier of the BSS (refer to Section 3.1).
- Spatial Reuse: this field indicates whether the HE node supports the SR operation. If this is the case, the field also indicates the limit on the transmission power to be used during the SR opportunities that can potentially be detected. Notice that a single Spatial Reuse field (of length of 4 bits) is carried in HE SU/MU/ER PPDUs, while HE TB PPDUs may include up to four Spatial Reuse fields. In particular, each field is meant for the SR operation in each allowed channel width (i.e., 20 MHz, 40 MHz, 80 MHz, and 160 MHz).

Besides supporting HE PPDU formats, HE STAs are required to be compatible with legacy formats. More information regarding HE PPDU formats can be found in [49].

A.2 Management Fields for Spatial Reuse

Some operations auxiliary to SR are enabled by control frames, which are Beacon, Probe Response, and (Re)Association Response frames. Beacons are used by APs to announce the presence of a BSS and to provide details of it. In particular, an AP, by means of Beacons, may request a STA to gather information regarding the environment: information of BSSs matching a particular BSSID and/or SSID, channel-specific report, or HE Operation element of neighboring HE APs. With this information, the AP can make decisions related to the SR operation. Regarding Probe Responses, they are meant to carry the information requested by devices scanning the area through Probe Requests. Finally,

(Re)Association Response frames are sent by APs to which a STA attempts to associate.

The abovementioned kind of frames are important to the SR operation because they carry, among other fields, the following information:

- **HE Capabilities:** it is used by HE STAs to announce support for certain HE capabilities.
- **HE Operation:** it defines the operation of HE STAs. For instance, it indicates whether BSS coloring is enabled or not.
- **BSS Color Change Announcement:** it is used by HE APs to indicate the utilization of a new BSS color so that the associated STAs and the surrounding devices can be aware of the change.
- **Spatial Reuse Parameter Set (SRPS) element:** this element provides the necessary information to carry out the OBSS/PD-based SR operation, which is defined in Section 4.1. The SRPS element is further defined in Appendix A.2.1.

A.2.1 Spatial Reuse Parameter Set element

The format of the SRPS element is optionally present in Beacons, Probe Responses and (Re)Association responses. Figure 28 shows the SRPS element in detail.

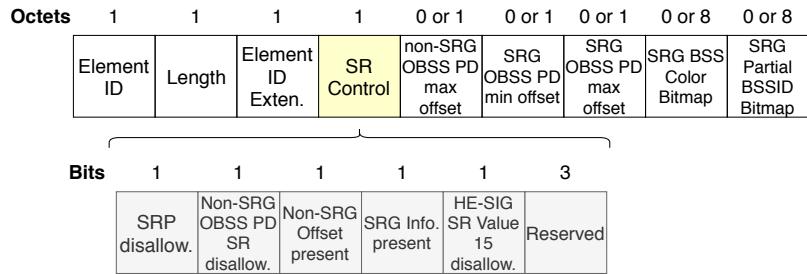


Figure 28: Spatial Reuse Parameter Set element.

Each item in the SRPS element is next described:

- Element ID: set to 255.
- Length: not defined.
- Element ID extension: set to 39.
- SR Control field: contains the following parameters:

- PSR Disallowed: indicates whether PSR transmissions are allowed or not at non-AP STAs that are associated with the AP that transmitted this element.
- Non-SRG OBSS/PD SR Disallowed: indicates whether non-SRG OBSS/PD SR transmissions are allowed or not at non-AP STAs that are associated with the AP that transmitted this element.
- Non-SRG Offset Present: indicates whether the Non-SRG OBSS/PD Max Offset subfield is present in the element.
- SRG Information Present: indicates whether the SRG OBSS/PD Min Offset, SRG OBSS/PD Max Offset, SRG BSS Color Bitmap, and SRG Partial BSSID Bitmap subfields are present in the element.
- HE-SIG-A Spatial Reuse Value 15 disallow: indicates whether non-AP STAs that are associated with the AP that transmitted this element may set the TXVECTOR parameter SPATIAL_REUSE to PSR_AND_NON-SRG-OBSS-PD_PROHIBITED to avoid PSR transmissions.
- Non-SRG OBSS/PD Max Offset: integer to generate the maximum Non-SRG OBSS/PD threshold.
- Non-SRG OBSS/PD Min Offset: integer to generate the minimum Non-SRG OBSS/PD threshold.
- SRG OBSS/PD Max Offset: integer to generate the maximum SRG OBSS/PD threshold.
- SRG BSS Color Bitmap: indicates which BSS Color values are used by the members of the SRG.
- SRG Partial BSSID Bitmap: indicates which partial BSSID values are used by members of the SRG.

Acknowledgment

This work has been partially supported by the Spanish Ministry of Economy and Competitiveness under the Maria de Maeztu Units of Excellence Programme (MDM-2015-0502), by PGC2018-099959-B-100 (MCIU/AEI/FEDER,UE), by the Catalan Government under SGR grant for research support (2017-SGR-11888), by SPOTS project (RTI2018-095438-A-I00) funded by the Spanish Ministry of Science, Innovation and Universities, and by a Gift from the Cisco University Research Program (CG#890107, Towards Deterministic Channel Access in High-Density WLANs) Fund, a corporate advised fund of Silicon Valley Community Foundation.

The authors would like to thank Dr. Malcolm Smith and Dr. Adrian Garcia for their thorough reviews and insightful comments.

References

- [1] TGax. IEEE P802.11ax/D4.0, 2019.
- [2] Boris Bellalta. IEEE 802.11 ax: High-efficiency WLANs. *IEEE Wireless Communications*, 23(1):38–46, 2016.
- [3] M Shahwaiz Afiaqui, EG Villegas, and EL Aguilera. IEEE 802.11 ax: Challenges and requirements for future high efficiency WiFi. *IEEE Wireless Communications*, 99:2–9, 2016.
- [4] Qiao Qu, Bo Li, Mao Yang, Zhongjiang Yan, Annan Yang, Jian Yu, Ming Gan, Yunbo Li, Xun Yang, Osama Aboul-Magd, et al. Survey and Performance Evaluation of the Upcoming Next Generation WLAN Standard IEEE 802.11 ax. *arXiv preprint arXiv:1806.05908*, 2018.
- [5] Evgeny Khorov, Anton Kiryanov, Andrey Lyakhov, and Giuseppe Bianchi. A Tutorial on IEEE 802.11 ax High Efficiency WLANs. *IEEE Communications Surveys & Tutorials*, 2018.
- [6] Simone Merlin and Santosh Abraham. Methods for improving medium reuse in IEEE 802.11 networks. In *Consumer Communications and Networking Conference, 2009. CCNC 2009. 6th IEEE*, pages 1–5. IEEE, 2009.
- [7] Xingang Guo, Sumit Roy, and W Steven Conner. Spatial reuse in wireless ad-hoc networks. In *2003 IEEE 58th Vehicular Technology Conference. VTC 2003-Fall (IEEE Cat. No. 03CH37484)*, volume 3, pages 1437–1442. IEEE, 2003.
- [8] Vivek P Mhatre, Konstantina Papagiannaki, and Francois Baccelli. Interference mitigation through power control in high density 802.11 wlans. In *IEEE INFOCOM 2007-26th IEEE International Conference on Computer Communications*, pages 535–543. IEEE, 2007.
- [9] Yihong Zhou and Scott M Nettles. Balancing the hidden and exposed node problems with power control in CSMA/CA-based wireless networks. In *Wireless Communications and Networking Conference, 2005 IEEE*, volume 2, pages 683–688. IEEE, 2005.
- [10] Francesc Wilhelmi, Sergio Barrachina-Muñoz, Boris Bellalta, Cristina Cano, Anders Jonsson, and Gergely Neu. Potential and pitfalls of multi-armed bandits for decentralized spatial reuse in WLANs. *Journal of Network and Computer Applications*, 127:26–42, 2019.
- [11] Masahito Mori et al. Performance Analysis of BSS Color and DSC. *Nov*, 3:11–14, 2014.
- [12] Zhao Shen, Bo Li, Mao Yang, Zhongjiang Yan, Xiaobo Li, and Yi Jin. Research and Performance Evaluation of Spatial Reuse Technology for Next Generation WLAN. In *International Wireless Internet Conference*, pages 41–51. Springer, 2018.

- [13] Sergio Barrachina-Muñoz, Francesc Wilhelmi, Ioannis Selinis, and Boris Bellalta. Komondor: a Wireless Network Simulator for Next-Generation High-Density WLANs. In *2019 Wireless Days (WD)*, pages 1–8. IEEE, 2019.
- [14] Wei Li, Yong Cui, Xiuzhen Cheng, Mznah A Al-Rodhaan, and Abdul-lah Al-Dhelaan. Achieving proportional fairness via AP power control in multi-rate WLANs. *IEEE Transactions on Wireless Communications*, 10(11):3784–3792, 2011.
- [15] Imad Jamil, Laurent Cariou, and Jean-Fran Hélard. Novel learning-based spatial reuse optimization in dense WLAN deployments.
- [16] Toshiro Nakahira, Koichi Ishihara, Yusuke Asai, Yasushi Takatori, Riichi Kudo, and Masato Mizoguchi. Centralized control of carrier sense threshold and channel bandwidth in high-density WLANs. In *Microwave Conference (APMC), 2014 Asia-Pacific*, pages 570–572. IEEE, 2014.
- [17] Pierre Chevillat, Jens Jelitto, and Hong Linh Truong. Dynamic data rate and transmit power adjustment in IEEE 802.11 wireless LANs. *International Journal of Wireless Information Networks*, 12(3):123–145, 2005.
- [18] Suhua Tang, Akio Hasegawa, Riichiro Nagareda, Akito Kitaura, Tatsuo Shibata, and Sadao Obama. Improving throughput of wireless LANs with transmit power control and slotted channel access. In *Personal Indoor and Mobile Radio Communications (PIMRC), 2011 IEEE 22nd International Symposium on*, pages 834–838. IEEE, 2011.
- [19] Chi-Kin Chau, Ivan WH Ho, Zhenhui Situ, Soung Chang Liew, and Jialiang Zhang. Effective static and adaptive carrier sensing for dense wireless CSMA networks. *IEEE Transactions on Mobile Computing*, 16(2):355–366, 2017.
- [20] Francesc Wilhelmi, Cristina Cano, Gergely Neu, Boris Bellalta, Anders Jonsson, and Sergio Barrachina-Muñoz. Collaborative spatial reuse in wireless networks via selfish multi-armed bandits. *Ad Hoc Networks*, 88:129–141, 2019.
- [21] M Shahwaiz Afiaqui, Eduard Garcia-Villegas, Elena Lopez-Aguilera, Graham Smith, and Daniel Camps. Evaluation of dynamic sensitivity control algorithm for IEEE 802.11 ax. In *Wireless Communications and Networking Conference (WCNC), 2015 IEEE*, pages 1060–1065. IEEE, 2015.
- [22] M Shahwaiz Afiaqui, Eduard Garcia-Villegas, Elena Lopez-Aguilera, and Daniel Camps-Mur. Dynamic sensitivity control of access points for IEEE 802.11 ax. In *Communications (ICC), 2016 IEEE International Conference on*, pages 1–7. IEEE, 2016.

- [23] Parag Kulkarni and Fengming Cao. Taming the densification challenge in next generation wireless LANs: An investigation into the use of dynamic sensitivity control. In *Wireless and Mobile Computing, Networking and Communications (WiMob), 2015 IEEE 11th International Conference on*, pages 860–867. IEEE, 2015.
- [24] Ioannis Selinis, Marcin Filo, Seiamak Vahid, Jonathan Rodriguez, and Rahim Tafazolli. Evaluation of the DSC algorithm and the BSS color scheme in dense cellular-like IEEE 802.11 ax deployments. In *Personal, Indoor, and Mobile Radio Communications (PIMRC), 2016 IEEE 27th Annual International Symposium on*, pages 1–7. IEEE, 2016.
- [25] Ioannis Selinis, Konstantinos Katsaros, Seiamak Vahid, and Rahim Tafazolli. Exploiting the Capture Effect on DSC and BSS Color in Dense IEEE 802.11 ax Deployments. In *Proceedings of the Workshop on ns-3*, pages 47–54. ACM, 2017.
- [26] Ioannis Selinis, Konstantinos Katsaros, Seiamak Vahid, and Rahim Tafazolli. Control OBSS/PD Sensitivity Threshold for IEEE 802.11 ax BSS Color. In *2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, pages 1–7. IEEE, 2018.
- [27] Boris Bellalta and Katarzyna Kosek-Szott. AP-initiated multi-user transmissions in IEEE 802.11 ax WLANs. *Ad Hoc Networks*, 85:145–159, 2019.
- [28] Matthew Fischer. Simulation-based evaluation of OBSS PD based SR default parameters, doc.: IEEE 802.11-16/1161r1, September 2016.
- [29] Piyush Gupta and Panganmala R Kumar. The capacity of wireless networks. *IEEE Transactions on information theory*, 46(2):388–404, 2000.
- [30] Hesham ElSawy, Ahmed Sultan-Salem, Mohamed-Slim Alouini, and Moe Z Win. Modeling and analysis of cellular networks using stochastic geometry: A tutorial. *IEEE Communications Surveys & Tutorials*, 19(1):167–203, 2016.
- [31] Xiaoguang Zhao, Xiangming Wen, Tao Lei, Zhaoming Lu, and Biao Zhang. On stochastic geometry analysis of dense wlan with dynamic carrier sense threshold and rate control. In *2016 19th International Symposium on Wireless Personal Multimedia Communications (WPMC)*, pages 211–216. IEEE, 2016.
- [32] Zhiwei Zhang, Yunzhou Li, Kaizhi Huang, and Chen Liang. On stochastic geometry modeling of wlan capacity with dynamic sensitive control. In *2015 13th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, pages 78–83. IEEE, 2015.

- [33] Motoki Iwata, Koji Yamamoto, Bo Yin, Takayuki Nishio, Masahiro Morikura, and Hirantha Abeysekera. Stochastic geometry analysis of individual carrier sense threshold adaptation in ieee 802.11 ax wlans. *IEEE Access*, 7:161916–161927, 2019.
- [34] Boris Bellalta, Alessandro Zocca, Cristina Cano, Alessandro Checco, Jaume Barcelo, and Alexey Vinel. Throughput analysis in CSMA/CA networks using continuous time Markov networks: a tutorial. In *Wireless Networking for Moving Objects*, pages 115–133. Springer, 2014.
- [35] Boris Bellalta. Throughput Analysis in High Density WLANs. *IEEE Communications Letters*, 21(3):592–595, 2017.
- [36] Sergio Barrachina-Muñoz, Francesc Wilhelmi, and Boris Bellalta. Dynamic Channel Bonding in Spatially Distributed High-Density WLANs. *IEEE Transactions on Mobile Computing*, 2019.
- [37] Sergio Barrachina-Muñoz, Francesc Wilhelmi, and Boris Bellalta. To overlap or not to overlap: Enabling channel bonding in high-density WLANs. *Computer Networks*, 152:40–53, 2019.
- [38] Francesc Wilhelmi and Sergio Barrachina-Muñoz. IEEE 802.11ax Single-Channel Spatial Reuse in the Spatial Flexible Continuous Time Markov Network (11axSR-SFCTMN). <https://github.com/sergiobarra/SFCTMN>, 2019.
- [39] S. Merlin et al. TGax Simulation Scenarios. doc. IEEE 802.11-14/0980r16, 2016.
- [40] Dmitry Bankov, Andre Didenko, Evgeny Khorov, and Andrey Lyakhov. OFDMA Uplink Scheduling in IEEE 802.11 ax Networks. In *2018 IEEE International Conference on Communications (ICC)*, pages 1–6. IEEE, 2018.
- [41] Konstantinos Dovelos and Boris Bellalta. Optimal Resource Allocation in IEEE 802.11ax Uplink OFDMA with Scheduled Access. *arXiv preprint arXiv:1811.00957*, 2019.
- [42] Ruizhi Liao, Boris Bellalta, Miquel Oliver, and Zhisheng Niu. MU-MIMO MAC protocols for wireless local area networks: A survey. *IEEE Communications Surveys & Tutorials*, 18(1):162–183, 2016.
- [43] Maddalena Nurchis and Boris Bellalta. Target wake time: scheduled access in IEEE 802.11 ax WLANs. *IEEE Wireless Communications*, 26(2):142–150, 2019.
- [44] G Smith. Dynamic sensitivity control-v2. *IEEE*, 802:802–11, 2015.
- [45] Chengnian Long, Qian Zhang, Bo Li, Huilong Yang, and Xinping Guan. Non-cooperative power control for wireless ad hoc networks with repeated games. *IEEE Journal on Selected Areas in Communications*, 25(6):1101–1112, 2007.

- [46] Ghasem Naddafzadeh-Shirazi, Peng-Yong Kong, and Chen-Khong Tham. Distributed reinforcement learning frameworks for cooperative retransmission in wireless networks. *IEEE Transactions on Vehicular Technology*, 59(8):4157–4162, 2010.
- [47] Pan Zhou, Yusun Chang, and John A Copeland. Reinforcement learning for repeated power control game in cognitive radio networks. *IEEE Journal on Selected Areas in Communications*, 30(1):54–69, 2011.
- [48] Euhanna Ghadimi, Francesco Davide Calabrese, Gunnar Peters, and Pablo Soldati. A reinforcement learning approach to power control and rate adaptation in cellular networks. In *2017 IEEE International Conference on Communications (ICC)*, pages 1–7. IEEE, 2017.
- [49] Lisa Ward. 1MA222: IEEE 802.11ax Technology Introduction. *Rhode & Schwarz. White paper 10.2016 - 1MA222_0e*, 2017.

On the Performance of the Spatial Reuse Operation in IEEE 802.11ax WLANs

Francesc Wilhelmi, Sergio Barrachina-Muñoz, and Boris Bellalta

Abstract

The Spatial Reuse (SR) operation included in the IEEE 802.11ax-2020 (11ax) amendment aims at increasing the number of parallel transmissions in an Overlapping Basic Service Set (OBSS). However, many unknowns exist about the performance gains that can be achieved through SR. In this paper, we provide a brief introduction to the SR operation described in the IEEE 802.11ax (draft D4.0). Then, a simulation-based implementation is provided in order to explore the performance gains of the SR operation. Our results show the potential of using SR in different scenarios covering multiple network densities and traffic loads. In particular, we observe significant improvements on the channel utilization when applying SR with respect to the default configuration, thus allowing to increase the throughput and reduce the delay. Interestingly, the highest improvements provided by the SR operation are observed in the most pessimistic situations in terms of network density and traffic load.

1 Introduction

The IEEE 802.11ax (11ax) amendment, which official publication is due to be released in June 2020, is expected to lay the groundwork of Next-Generation (NG) Wireless Local Area Networks (WLANs). One of the main goals of this amendment is to improve network efficiency by increasing the number of parallel transmissions in an Overlapping Basic Service Set (OBSS). To that purpose, the Spatial Reuse (SR) operation is introduced along with other techniques to boost the performance of NG WLANs, from which we highlight Orthogonal Frequency-Division Multiple Access (OFDMA) or Downlink/Uplink Multi-User Multiple-Input-Multiple-Output (DL/UL MU-MIMO) [1].

The SR operation is based on sensitivity adjustment together with Transmission Power Control (TPC). In particular, a specific OBSS Packet Detect (OBSS/PD) threshold is employed for the detected OBSS transmissions (also referred to as inter-BSS transmissions), so that channel utilization can be enhanced. Moreover, in order not to affect any ongoing transmission, a node applying SR must limit its transmit power as a function of the OBSS/PD threshold.

Fig. 1 depicts a use case where the SR operation could potentially improve the network efficiency of an OBSS. Notice that the dashed lines in the figure

indicate the carrier sense area of each device, provided that the transmit power of the others is fixed and that the same channel is used. As illustrated, the default Clear Channel Assessment Carrier Sense (CCA/CS) threshold would not allow simultaneous transmissions to be held between Access Points A and B (AP_A and AP_B). In that case, each device should defer its transmission when the other occupies the channel, due to the application of the Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) protocol. Nevertheless, by properly increasing the OBSS/PD threshold of any AP (e.g., as illustrated for AP_A), both devices would be able to transmit at the same time, thus improving the utilization of the channel.

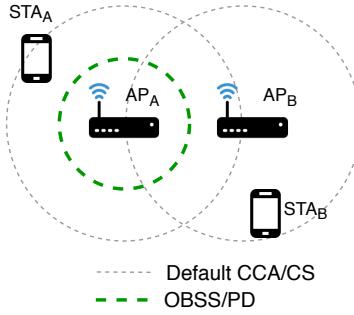


Figure 1: Example of the potential of the 11ax SR operation.

Despite SR is expected to bring significant performance gains to WLANs, its actual benefits are still unknown. First, the new kind of inter-WLAN interactions that the operation generates is unexplored, as well as their impact on the network. Moreover, the improvement achieved by applying SR on the performance (e.g., throughput, delay) is hindered by the trade-off between the number of parallel transmissions and their duration. Note that increasing the OBSS/PD threshold (i.e., using a more aggressive configuration) entails decreasing the transmission power, which may result in using a lower Modulation and Coding Scheme (MCS), or even experiencing an increased packet error rate. The effects of increasing the OBSS/PD threshold, and hence decreasing the transmission power, are summarized in Table 1. As shown, an increase in the OBSS/PD threshold entails a higher probability of accessing to the channel since the number of sensed inter-BSS transmissions can be potentially reduced (this is equivalent to reducing the exposed-node probability). In contrast, the number of hidden nodes can potentially increase as the carrier sense area is being reduced.

In this work, we shed light on the performance of the 11ax SR operation and highlight the situations in which it is worth using it. The main contributions of this paper are as follows:

- We provide a summary of the OBSS PD-based SR operation included in draft version D4.0 of the IEEE 802.11ax amendment, which is, to the date

Table 1: Effect of increasing the OBSS/PD threshold and the transmission power.

	Data rate	Channel access probability	Hidden-node probability	Exposed-node probability
OBSS/PD ↑ (Tx Power ↓)	↓	↑	↑	↓

of publishing this article, under the initial sponsor ballot phase.

- We present an implementation of the aforementioned operation in the 11ax-based Komondor simulator [2].¹
- We evaluate the performance of the SR operation through simulations, and assess its potential for next-generation wireless networks. Different network densities and traffic loads are considered for covering the analysis of multiple use cases.

2 Related Work

The SR operation has been previously surveyed and evaluated in [1, 3, 4, 6, 7]. However, these works refer to previous draft versions of the amendment (D1.0 o D2.0), which has undergone significant modifications in its current version (D4.0).

First, the Task Group ax (TGax) presented some preliminary results for cellular-type scenarios in [4]. In particular, significant gains were shown when combining BSS Coloring and Dynamic Sensitivity Control (DSC) [5]. A further analysis was then provided in [6] for office scenarios, which also showed that gains were only achieved in dense deployments. Nonetheless, the simulations conducted in that work were obtained from a customized system and link level integrated simulation platform, from which no validation was provided.

The authors in [7] provided a thorough performance evaluation of the SR operation, in addition to several other features included in the 11ax amendment. To that purpose, they proposed their own simulation platform for IEEE 802.11ax called SLISP, which mostly focuses on the MAC of the 11ax. Based on that, the SR operation was evaluated in indoor and outdoor scenarios containing multiple BSS. While important gains were shown in indoor deployments (especially for downlink traffic), a moderate gain was observed in outdoor situations.

As shown, few works attempt to provide a performance evaluation of the SR operation through simulations. The main cause lies in the novelty of the mechanism. Accordingly, there is a lack of reliable simulation platforms that include 11ax features. To the date of publishing this article, SR is still under

¹All the source code of Komondor is open and free to use (Github repository: <https://github.com/wn-upf/Komondor>), with the aim of encouraging potential collaborations with any interested researcher.

development for ns-3.² Due to the lack of simulation tools including the 11ax SR operation, in this work we provide an implementation of the SR operation in the Komondor simulator.³ Moreover, our results are gathered based on the newest draft version (D4.0).

3 Spatial Reuse Operation

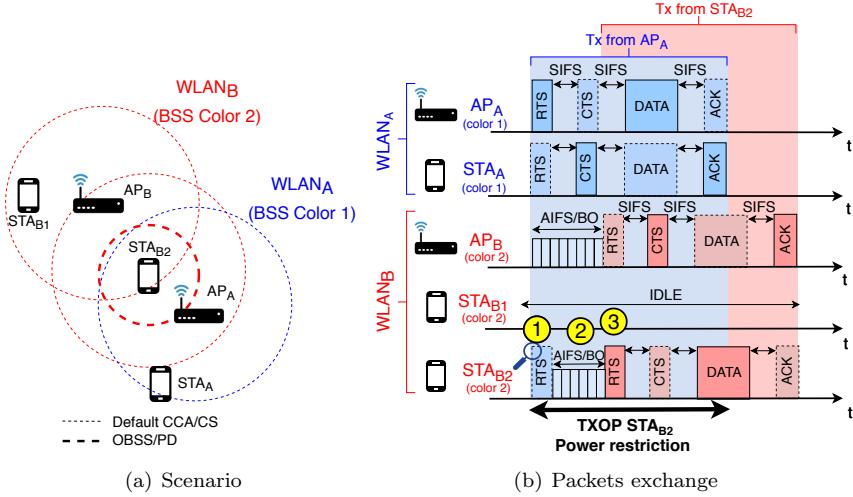


Figure 2: Example of applying the OBSS PD-based SR operation.

The 11ax SR operation is divided into two different and independent mechanisms. On the one hand, we find the OBSS PD-based SR operation, whereby 11ax devices - a.k.a High Efficiency (HE) nodes⁴ - can detect SR opportunities from inter-BSS transmissions by using a more aggressive CCA policy. On the other hand, the Spatial Reuse Parameter (SRP)-based SR operation performs similarly but only taking advantage of trigger-based communications [8]. Throughout this document, we will exclusively refer to the first mechanism (i.e., OBSS PD-based SR) because of the development cost of building trigger-based transmissions required for the SRP-based SR operation. In addition, the slow adoption of 11ax in WLANs would prevent using full scheduling transmissions schemes, in favor of CSMA/CA ones. Notwithstanding, both mechanisms are expected to lead to similar results since the procedure of adjusting the OBSS/PD is similar.

²All the new developments in ns-3 are published in the following repository: <https://gitlab.com/nsnam/ns-3-dev>

³The validation of the Komondor simulator against ns-3 can be found in [2].

⁴By 11ax node, we may refer indistinctly to an HE STA or an HE AP.

3.1 BSS Coloring and Spatial Reuse Groups

The whole SR operation is based on identifying the source of a given transmission, i.e., inter-BSS frame detection. The idea is that HE nodes can rapidly decode the MAC headers of a certain transmission and determine its origin. Then, a more aggressive OBSS/PD threshold can be employed to increase the probability of accessing to the channel.

For the fast packet source identification, two concepts are introduced, which stand for BSS Coloring and Spatial Reuse Groups (SRG). On the one hand, the BSS Color field is included in the MAC headers⁵ to uniquely identify different WLANs belonging to an OBSS. In case of detecting a color collision, the affected WLANs must change their BSS Color. On the other hand, SRGs can be formed by a set of overlapping WLANs. The SRG field is present in control frames such as Beacons, Probe responses, or (Re)Association responses.⁶ In this case, a specific OBSS/PD threshold can be used for transmissions within the same SRG.

3.2 General Constraints

The 11ax amendment includes a set of constraints on defining the OBSS/PD threshold to be used for detecting SR opportunities. In particular, the OBSS/PD value cannot exceed the following upper bound:

$$\text{OBSS/PD} \leq \max \left(\text{OBSS/PD}_{\min}, \min \left(\text{OBSS/PD}_{\max}, \text{OBSS/PD}_{\min} + (\text{TX_PWR}_{\text{ref}} - \text{TX_PWR}) \right) \right),$$

where OBSS/PD_{\min} and OBSS/PD_{\max} are set to -82 dBm and -62 dBm, respectively, the reference power $\text{TX_PWR}_{\text{ref}}$ is set to 21 or 25 dBm, according to the capabilities of the device,⁷ and TX_PWR is the transmission power in dBm.

In order to regulate the transmissions held during SR-based opportunities, the transmission power is limited according to the OBSS/PD threshold used for detecting those opportunities. In case that $\text{OBSS/PD} \leq \text{OBSS/PD}_{\min}$, the transmission power is unconstrained. Otherwise, the maximum allowed transmission power TX_PWR_{\max} is given by:

$$\text{TX_PWR}_{\max} = \text{TX_PWR}_{\text{ref}} - (\text{OBSS/PD} - \text{OBSS/PD}_{\min}) \quad (1)$$

⁵The BSS Color is carried in the HE-SIG-A field, which is present in every Physical Layer Convergence Procedure (PLCP) Protocol Data Unit (PPDU).

⁶Unlike the BSS Color, the SRG is included in the Spatial Reuse Parameter Set (SRPS) element. A bitmap is stored by each 11ax node applying SR, which maps the set of BSS Colors that belong to a certain SRG.

⁷The $\text{TX_PWR}_{\text{ref}}$ can be set to either 21 or 25 dBm, depending on the transmission capabilities of the HE node with regards to the highest supported number of spatial streams (NSS).

3.3 Example of the OBSS PD-based Spatial Reuse Operation

In order to illustrate the concepts described above, let us consider the scenario depicted in Fig. 2(a). As shown, there is a device, namely STA_{B2} , which, by using the default configuration, is prone to suffer from flow starvation as a result of the OBSS interference. Nevertheless, the OBSS PD-based SR operation allows STA to overcome the aforementioned interference, thus gaining access to the channel. This is illustrated in Fig. 2(b), where inter-BSS transmissions are ignored by STA_{B2} when using the SR operation. In marker 1 (shown in yellow), STA_{B2} inspects the Request to Send (RTS) frame sent by AP_A , which is identified as an inter-BSS transmission. Accordingly, it uses a more aggressive OBSS/PD threshold, which allows the backoff procedure to be resumed (marker 2). Finally, STA_{B2} starts its own transmission by taking advantage of the detected SR-based opportunity (marker 3). However, a power restriction is applied, thus decreasing the MCS and increasing the data transmission time, as a consequence.

4 Implementation of OBSS PD-based SR in Komondor

The Komondor simulator was conceived, among other purposes, to allow the low-complexity integration of novel mechanisms included in new IEEE 802.11 standards, which are not yet available or validated in other well-known simulators such as ns-3. Therefore, Komondor can serve as a first step towards analyzing novel features that will potentially shape future wireless networks. In this Section, we briefly introduce the implementation conceived for the SR operation.⁸

Fig. 3 shows a flowchart that summarizes the SR implementation for a given HE node in case of detecting a single inter-BSS transmission. The most important groups of functionalities (highlighted with numbers in the figure) are described in detail in the following subsections.

4.1 Initialization

First of all, any node that applies SR must indicate support for it. In addition, the parameters related to the SR operation should be set:

- **BSS Color:** identifier of the BSS to which the node belongs to. The BSS Color identifies WLANs uniquely.
- **Spatial Reuse Group (SRG):** identifier of the group to which the node belongs.

⁸The code used for the simulations of this paper can be found in pre-release v3.0 (<https://github.com/wn-upf/Komondor/releases/tag/3.0>).

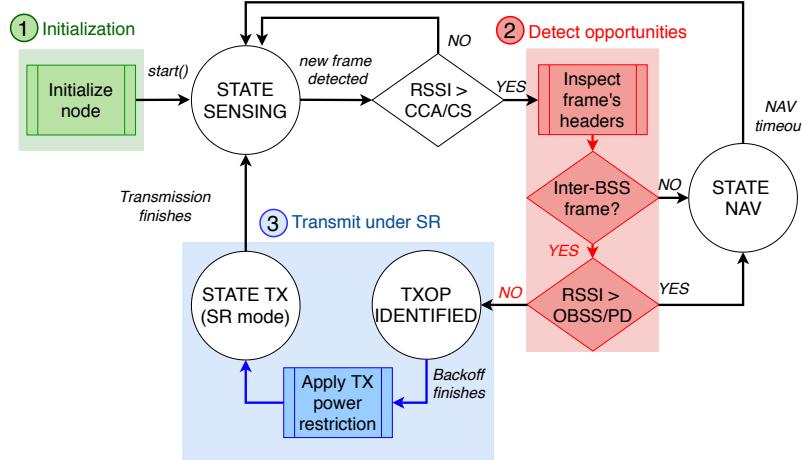


Figure 3: Flowchart of the SR implementation for a given HE node when detecting a single inter-BSS transmission.

- **Non-SRG OBSS/PD:** the sensitivity threshold to be used for generic inter-BSS transmissions.
- **SRG OBSS/PD threshold:** the sensitivity threshold to be used for inter-BSS transmissions that are originated by nodes belonging to the same SRG.

The Komondor simulator simplifies the PHY layer for the sake of efficiency, so that particular focus is put on the MAC. Moreover, operations related to management and control are simplified too. According to that, the initialization of nodes is logically performed at the beginning of the simulation, instead of simulating the actual exchange of control frames between nodes. In particular, the AP of a given WLAN is responsible for notifying the initial SR configuration and any potential change to its associated STAs.

4.2 Detection of SR-based Opportunities

Once the simulation starts and nodes begin to exchange packets, it is possible to detect SR-based opportunities from inter-BSS frames. For that, a certain HE node must first analyze the headers of any detected frame and rapidly identify its source. During this stage, the HE node will assess whether the transmitter belongs to the same WLAN (intra-BSS) or not (inter-BSS). Moreover, in case of being of kind inter-BSS, the frame is sub-categorized into SRG or non-SRG, according to the groups established during initialization.

In case of detecting an intra-BSS transmission, the default CCA/CS threshold is used. Otherwise, the corresponding OBSS/PD threshold (non-SRG or

SRG) is applied. In accordance with that, the power received P_{rx} from the incoming transmission is used to identify potential SR-based opportunities. In particular, the following two conditions must hold to identify an opportunity:

- 1) $P_{rx} \geq CCA/CS$, to guarantee the correct decoding of the MAC header, and
- 2) $P_{rx} < OBSS/PD$, to trigger the opportunity.

4.3 Transmit under the SR mode

When detecting an SR-based opportunity, an HE node detects the channel as idle, which allows decreasing the backoff. Once the backoff counter is over and the node is about to transmit, a transmit power limitation is applied (1). Finally, once the HE node finishes its SR-based transmission, it returns to the default sensing state, where channel access is scheduled according to the legacy CCA/CS threshold.

It is important to notice that several SR-based opportunities can be detected before transmitting, due to the multiple receptions of different inter-BSS frames. In that case, the most restrictive power limitation must be applied.

5 Simulation Setup

In this Section, we depict the simulation setup that has been considered for evaluating the performance of the SR operation.

5.1 Channel Model

Path-loss effects are characterized according to the TMB 5GHz indoor model for IEEE 802.11ac/11ax WLANs [9]. In particular, the path-loss PL_{TMB} between a transmitter i and a receiver j that are separated by $d_{i,j}$ meters is given by:

$$PL_{TMB}(d_{i,j}) = L_0 + 10 \cdot \gamma \cdot \log_{10}(d_{i,j}) + k \cdot \bar{W} \cdot d_{i,j},$$

where L_0 is the path-loss intercept, γ is the path-loss exponent, k is the attenuation factor that characterizes obstacles, and \bar{W} is the average number of wall obstacles per meter.

5.2 Traffic Generation and Data Rate

Only downlink transmissions are considered for the sake of capturing inter-AP interactions. Hence, a UDP traffic generator is attached to every AP. All traffic generators randomly produce packets at the same average traffic load ℓ , which varies depending on the scenario. The packets arrival process to the APs is modeled through the well-known Poisson distribution.

The rate at which data is transmitted is based on the MCS modes defined in the 11ax amendment, which are selected according to the link quality between the transmitter (the AP) and the receiver (the STA). The highest achievable data rate (135 Mbps) is achieved when using modulation 1024-QAM at a coding rate of 5/6.

5.3 Throughput Calculation and Reception Model

Nodes operate under the CSMA/CA protocol, and use the SR operation on top of that. Since Komondor simulates the actual exchange of frames between nodes in a WLAN, the throughput S experienced by it is directly obtained from:

$$S = \frac{[\text{Data bits transmitted successfully}]}{[\text{Total simulation time}]}$$

The number of data bits (or data packets) transmitted successfully depends on the varying channel conditions and sensed interference. In particular, a given transmission is considered to be successful only if the following conditions hold at the receiver:

1. The power sensed at the receiver from the frame being decoded remains above the CCA/CS.
2. The Signal-to-Interference-plus-Noise Ratio (SINR) stays above the capture effect (CE) threshold, set to 10 dB. Notice that this is an abstraction of the CE model, which is due to the simplification of the PHY in Komondor.

5.4 Scenarios for Evaluation

The 11ax SR operation is evaluated in random scenarios like the one depicted in Fig. 4. Notice that, for the sake of illustrating the potential of SR, only the WLAN in the middle (namely, WLAN_A) applies the SR operation, while the others remain using the default CCA/CS. WLAN_A is placed at the center of the scenario, so that it is normally exposed to a higher level of interference than the others. We consider that all the WLANs are operating in the same channel, as otherwise, they would not interact.

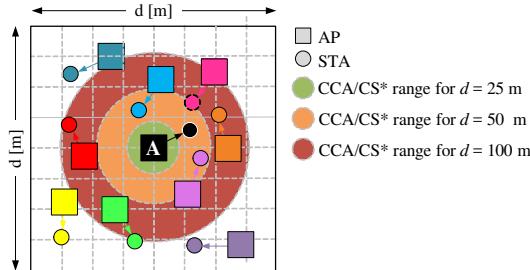


Figure 4: Random deployment with WLAN_A placed in the center. Note that we include the CCA/CS* range of WLAN_A in the map – corresponding to maximum transmission power (20 dBm) and minimum CCA (-82 dBm) – for the sake of representing the three considered map densities.

Different network densities are considered, which are useful to evaluate the gains achieved by the 11ax SR operation in different use cases. In particular,

we consider low-density (LD), medium-density (MD), and high-density (HD) scenarios. In summary, we simulate multiple scenarios accounting for $N_m = 3$ maps of sizes 25×25 , 50×50 , and 100×100 m 2 , $N_d = 50$ different random deployments (i.e., node allocation), $N_{\text{cca}} = 21$ OBSS/PD values ranging from -62 to -82 dBm for WLAN $_A$, and $N_\ell = 16$ traffic loads ranging from $\ell = 1$ to 100 Mbps. In total, $N_s = N_m \times N_d \times N_{\text{cca}} \times N_\ell = 50,400$ scenarios (or Komondor inputs) are simulated. The observation time for each simulation is $T = 10$ seconds.

The 11ax PHY and MAC parameters used in the simulations are listed in Table 2.

Table 2: Simulation parameters.

	Parameter	Value
PHY	Central frequency f_c	5 GHz
	Transmission gain G_{tx}	0 dB
	Reception gain G_{rx}	0 dB
	Capture Effect threshold CE	10 dB
	Path-loss (TMB) $PL_{\text{TMB}}(d)$	See (2)
	Path-loss intercept L_0	54.120
	Path-loss exponent γ	2.06067
	Attenuation factor k	5.25
	Average num. of walls per m W	0.1467
	Background noise level N	-95 dBm
	Legacy OFDM symbol duration σ_{leg}	4 μ s
	OFDM symbol duration (GI-32) σ_{32}	16 μ s
	Number of subcarriers (20 MHz) N_{sc}	234
	Number of spatial streams N_{ss}	1
	Transmit power levels \mathcal{T}	1 to 20 dBm (1 dBm steps)
MAC	Empty slot duration T_e	9 μ s
	SIFS duration T_{SIFS}	16 μ s
	DIFS/AIFS duration $T_{\text{DIFS/AIFS}}$	34 μ s
	PIFS duration T_{PIFS}	25 μ s
	Legacy preamble duration $T_{\text{PHY-leg}}$	20 μ s
	HE single-user field duration $T_{\text{HE-SU}}$	100 μ s
	ACK duration T_{ACK}	28 μ s
	Block ACK duration T_{BACK}	32 μ s
	Size OFDM symbol (legacy) $L_{s,l}$	24 bits
	Length of data packets L_d	12,000 bits
	Max. No. of frames in an A-MPDU N_{agg}	64
	Length of an RTS packet L_{RTS}	160 bits
	Length of a CTS packet L_{CTS}	112 bits
	Length of service field L_{SF}	16 bits
	Length of MAC header L_{MH}	320 bits
	Max. contention window (fixed) CW	15
	Allowed sensitivity levels \mathcal{S}	-82 to -62 (1 dBm steps)
Other	Traffic model Λ	Downlink
	Traffic load ℓ	1 to 100 Mbps
	Map area A	25x25, 50x50 and 100x100 m 2

6 Performance Evaluation

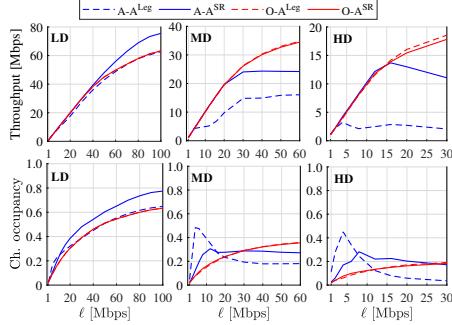


Figure 5: Throughput and channel occupancy experienced by WLAN_A (A) and the other WLANs (O) in low (LD), medium (MD) and high density (HD) deployments. Each curve is named in the legend in the format $X\text{-}A^m$, where A^m represents whether WLAN_A uses spatial reuse (SR) or not (Leg).

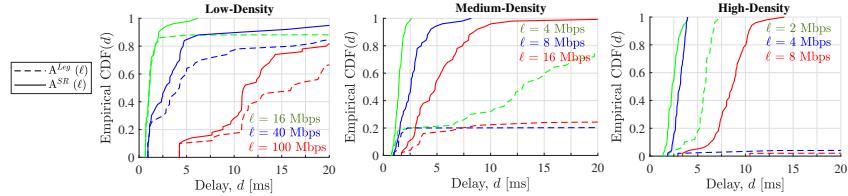


Figure 6: Empirical cumulative distribution function of the average packet delay experienced by WLAN_A . Different network densities and traffic loads are considered. Solid and dashed lines indicate whether WLAN_A uses spatial reuse (SR) or not (Leg), respectively.

Based on the simulation setup described in Section 5, we evaluate the potential of the 11ax SR operation in various situations. In particular, for each network density and traffic load, we measure the maximum performance gain that can be achieved by WLAN_A when applying SR. This upper bound is provided by the OBSS/PD value that maximizes the average throughput in each of the 50 random deployments. We also assess the impact of such an optimal SR configuration (from WLAN_A 's perspective) on the environment (i.e., the other WLANs). Specifically, we compare the throughput, channel occupancy, and delay obtained by all the other WLANs in two situations: *i*) the legacy CCA/CS is used by the entire network (including WLAN_A), and *ii*) Only WLAN_A applies the SR using the optimal OBSS/PD.

Fig. 5 shows the highest gains that can be achieved both in throughput and channel occupancy when WLAN_A implements the SR operation. As shown in

the first row, significant improvements are achieved on WLAN_A 's individual throughput, especially for the highest network density (up to 450% for medium load). Moreover, importantly, the mean throughput achieved by the rest of WLANs (on average) is barely affected by the SR operation applied by WLAN_A .

Regarding channel occupancy (second row), improvements are also noticed as traffic load increases, for every network density. Essentially, more transmission opportunities are achieved due to the lower OBSS/PD threshold used by WLAN_A . Nonetheless, an interesting phenomenon occurs for the MD and HD scenarios at low-moderate traffic loads. In these cases, the legacy approach leads to a clear increase on channel occupancy. This increase is caused by the elevated number of re-transmissions performed, which are propitiated by the higher collision probability between overlapping WLANs. We conclude that SR allows using channel resources more efficiently by reducing the contention among neighboring WLANs, and thus boosting parallel transmissions.

Finally, Fig. 6 illustrates the potential reduction on the delay experienced by WLAN_A when implementing SR. In particular, we select the empirical cumulative distribution function (CDF) to highlight the probability of experiencing small and high delays resulting from all the simulated scenarios. Similarly than before, for each scenario, we pick the average delay obtained by the best possible OBSS/PD threshold used by WLAN_A in each of the deployments (in terms of throughput) and compare it with the legacy situation. Three representative traffic load values are included for each network density. As illustrated, the probability of experiencing a high delay rapidly increases with network density and traffic load when the legacy operation is considered. Nevertheless, SR substantially improves delay performance by keeping it at moderate values in most of the simulated scenarios.

7 Conclusions

In this paper, we introduced the 11ax SR operation and evaluated its potential in a variety of scenarios covering different node densities and traffic loads. To that purpose, we provided an implementation of the SR operation in the 11ax-oriented Komondor simulator. Our results showed that significant gains can be achieved by using the SR operation, especially in high interference situations where both network density and traffic load are high. Therefore, there is a huge potential in SR when it comes to maximizing channel utilization.

As future work, the potential of SR will be studied, especially regarding the interactions that occur when more than one WLAN applies the operation. Moreover, the problem of finding the best OBSS/PD threshold will be studied. In this regard, online learning stands as a powerful and suitable solution, due to the complex inter-WLAN interactions that can be generated by using SR. Promising results have been already shown by applying Machine Learning (ML) to address the SR problem [10, 11]. Finally, the interaction of SR with other techniques included in the 11ax (e.g., directional transmissions, target wake time, OFDMA, etc.) is also worth to be studied.

Acknowledgments

This work has been partially supported by the Spanish Ministry of Economy and Competitiveness under the Maria de Maeztu Units of Excellence Programme (MDM-2015-0502), by WINDMAL PGC2018-099959-B-I00 (MCIU/AEI/FEDER,UE), by the Catalan Government under SGR grant for research support (2017-SGR-11888), and by a Gift from the Cisco University Research Program (CG#890107, Towards Deterministic Channel Access in High-Density WLANs) Fund, a corporate advised fund of Silicon Valley Community Foundation. The work by S. Barrachina-Muñoz is supported by an FI grant from the Generalitat de Catalunya.

References

- [1] Bellalta, Boris, “IEEE 802.11 ax: High-efficiency WLANs,” IEEE Wireless Communications, vol. 23, num. 1, pp. 38–46, 2016.
- [2] Barrachina-Muñoz, S., Wilhelmi, F., Selinis, I., & Bellalta, B. (2019, April). Komondor: a Wireless Network Simulator for Next-Generation High-Density WLANs. In 2019 Wireless Days (WD) (pp. 1-8). IEEE.
- [3] Khorov, E., Kiryanov, A., Lyakhov, A., & Bianchi, G. (2018). “A tutorial on IEEE 802.11 ax high efficiency WLANs.” IEEE Communications Surveys & Tutorials, 21(1), 197-216.
- [4] Itagaki, T., Morioka, Y., Mori, M., Ishihara, K., Shinohara, S., & Inoue, Y. “Performance Analysis of BSS Color and DSC.” IEEE 802.11 ax, doc. IEEE802. 11-14/1403r0.
- [5] Smith, G. K. (2017). “Dynamic sensitivity control for wireless devices.” U.S. Patent No. 9,596,702. Washington, DC: U.S. Patent and Trademark Office.
- [6] Shen, Z., Li, B., Yang, M., Yan, Z., Li, X., & Jin, Y. (2018, October). “Research and Performance Evaluation of Spatial Reuse Technology for Next Generation WLAN.” In International Wireless Internet Conference (pp. 41-51). Springer, Cham.
- [7] Qu, Q., et al., & Au, E. (2018). “Survey and Performance Evaluation of the Upcoming Next Generation WLAN Standard-IEEE 802.11 ax.” arXiv preprint arXiv:1806.05908, unpublished.
- [8] Bellalta, B., & Kosek-Szott, K. (2019). “AP-initiated multi-user transmissions in IEEE 802.11 ax WLANs.” Ad Hoc Networks, 85, 145-159.
- [9] Adame, T., Carrascosa, M., & Bellalta, B. (2019, April). The TMB path loss model for 5 GHz indoor WiFi scenarios: On the empirical relationship between RSSI, MCS, and spatial streams. In 2019 Wireless Days (WD) (pp. 1-8). IEEE.

- [10] Wilhelm, F., Cano, C., Neu, G., Bellalta, B., Jonsson, A., & Barrachina-Muñoz, S. (2019). Collaborative spatial reuse in wireless networks via selfish multi-armed bandits. *Ad Hoc Networks*, 88, 129-141.
- [11] Wilhelm, F., Barrachina-Muñoz, S., Bellalta, B., Cano, C., Jonsson, A., & Neu, G. (2019). Potential and pitfalls of multi-armed bandits for decentralized spatial reuse in wlans. *Journal of Network and Computer Applications*, 127, 26-42.

A Flexible Machine Learning-Aware Architecture for Future WLANs

Francesc Wilhelmi, Sergio Barrachina-Muñoz, Boris Bellalta,
Cristina Cano, Anders Jonsson, and Vishnu Ram

Abstract

Lots of hopes have been placed on Machine Learning (ML) as a key enabler of future wireless networks. By taking advantage of large volumes of data, ML is expected to deal with the ever-increasing complexity of networking problems. Unfortunately, current networks are not yet prepared to support the ensuing requirements of ML-based applications in terms of data collection, processing, and output distribution. This article points out the architectural requirements that are needed to pervasively include ML as part of future wireless networks operation. Specifically, we look into Wireless Local Area Networks (WLANs), which, due to their nature can be found in multiple forms, ranging from cloud-based to edge-computing-like deployments. In particular, we propose to adopt the International Telecommunications Union (ITU) unified architecture for 5G and beyond. Based on ITU's architecture, we provide insights on the main requirements and the major challenges of introducing ML to the multiple modalities of WLANs. Finally, we showcase the superiority of the architecture through an ML-enabled use case for future networks.

1 Introduction

Wireless communications have reached a point where a paradigm shift is required to satisfy the increasing needs of next-generation applications [1]. Based on the current trend, Artificial Intelligence (AI), and more precisely Machine Learning (ML), is expected to conduct a revolution, especially regarding the network planning, operation, and management of the 5th and 6th generations (5G/6G) of mobile communications.

ML is meant to empower a computational system for learning automatically, based on experience, so that future situations can be properly managed without having been programmed explicitly. Concerning wireless communications, there is a huge amount of unexploited data generated at both infrastructure and user levels, which could be extremely useful for learning complex patterns, thus improving network performance. For instance, the behavior of users in a network-oriented service can be predicted through ML given the information

Table 1: Machine learning methods, algorithms, potential networking applications, and examples of input data.

ML method	Algorithms	Potential applications	Examples of input data
SL	Linear classifiers, regression methods such as Autoregressive Integrated Moving Average (ARIMA), neural networks, hidden Markov models, random forest, support vector machines, k-nearest neighbors, principal component analysis	Traffic forecasting, mobility pattern prediction, flow classification, routing, anomaly detection, spectrum management, failure detection, QoE prediction	IP traffic matrices, temporal user location, availability of routing paths, application data, channel measurements, performance metrics
UL	Clustering, mixture models, generative models, non-negative matrix factorization, evolutionary algorithms	Traffic classification, virtual topology design, path computation, intruder detection, signal separation	IP traffic matrices, historical end-to-end bit-rate, received symbols
RL	Monte Carlo, Q-learning, State-Action-Reward-State-Action (SARSA), deep Q network, actor-critic, multi-armed bandits, learning automaton, Markov decision processes	Power control, rate adaptation, routing, channel selection, spatial reuse, smart caching, traffic offloading, cognitive channel access, energy harvesting, energy efficiency	Channel measurements, link status, performance metrics (e.g., throughput, delay), server occupation, power consumption

from previous sessions. Based on these predictions, network resources can be appropriately accommodated in future sessions.

Unfortunately, the potential benefits of ML for real networks are currently limited by the existing infrastructure, which is not yet prepared to accommodate ML-oriented tasks such as data collection, processing, and output distribution. Instead, current networking systems are typically meant for delivering content, without taking into account the underlying characteristics of the processes that generate it.

The first steps towards AI-enabled networking are currently being made in 5G through Network Function Virtualization (NFV). Unlike traditional hardware-based networks, NFV allows rapid elasticity and fast reconfiguration on assigning network resources. This is particularly useful to enable verticals such as autonomous driving in the automotive sector or smart manufacturing in Industry 4.0. Besides, network virtualization is useful to boost inter-operator coordination and bringing the ML operation to a macro-scale level, counting with vast information and computation resources.

To conduct the evolution towards ML-aware networks, standardization is key to reach consensus between operators and manufacturers. In this regard, we find many initiatives held by standardization organizations, from which we highlight the Focus Group on Machine Learning for Future Networks including 5G (FG-ML5G), which belongs to the International Telecommunication Union Telecommunication Standardization Sector (ITU-T). The FG-ML5G aims to enable the convergence of future communications with ML technologies. To that end, the focus group has released a specification on a *Unified architecture for 5G and beyond*, recently turned into an ITU Recommendation [2]. Re-

markably, ITU’s standardized architecture provides a common nomenclature for ML-related mechanisms so that interoperability with other networking systems is achieved.

Apart from the ITU-T initiatives, other important standardization bodies such as the 3rd Generation Partnership Project (3GPP) or the European Telecommunications Standards Institute (ETSI) are currently working on the integration of data analytics to network functions. The 3GPP contemplates AI as one of the priority topics for shaping its upcoming release (Release 17) and architectural requirements are currently under study [3]. Furthermore, we highlight the ETSI groups on Experiential Networked Intelligence (ENI) and Zero-touch network and Service Management (ZSM), which actively study the integration of AI to networks [4]. Unlike the ITU’s unified architecture, most of the work held by the 3GPP and the ETSI focuses on centralized data collection and data analytics solutions. Nevertheless, we understand that the works in [2–4] are complementary to each other.

To contribute to the evolution of wireless communications towards AI-based systems, we provide a realization of the ITU’s architecture for IEEE 802.11 Wireless Local Area Networks (WLANs), which will be an essential part of the 5G/6G ecosystem in the unlicensed bands. Unlike for cellular networks, WLANs have received much less attention when designing AI-aware architectural solutions. The fact is that cellular-based deployments fit in perfectly with big data analytics, due to the vast amount of data and high computation resources available for mobile network operators. In contrast, WLANs pose a set of specific challenges resulting from their multiple deployment modes (e.g., campus network, residential usage) and their typical decentralized nature. Despite WLANs can count with plenty of data to be used by ML methods in large and planned deployments, we find other residential-type scenarios that lack of powerful centralized equipment. In these cases, huge computing and processing resources cannot be endowed to the ML operation.

To enable the integration of ML-based approaches into the different modalities of WLANs, the module-based ITU’s architecture allows adapting to the problem instance and the set of available resources, thus providing flexibility in terms of deployment heterogeneity. For instance, despite deep learning is a powerful solution that may improve the performance in multiple scenarios, it entails a set of computation, storage and communication requirements that may not be fulfilled in other deployments, or parts of the network.

The main contributions of this paper are as follows:

- We devise and discuss the potential of ML-enabled future communications. Then, we focus on IEEE 802.11 WLANs and provide a set of use cases.
- We provide an overview of the ITU’s ML-aware architecture for 5G networks and beyond.
- We adopt the module-based ITU’s architecture and provide a realization for IEEE 802.11 WLANs, thus pointing out the major technical challenges and opportunities.

- We depict the potential advantages of ML-based approaches enabled by the architecture through numerical results in a particular use case.

2 Machine Learning as Enabler of Future Wireless Networks

In this section, we discuss the role of ML for sustaining the progress of future wireless networks. Then, we motivate the application of ML to IEEE 802.11 WLANs through a set of illustrative ML-driven use cases.

2.1 Machine Learning in Communications

The proliferation of new communication-based applications is defining the shape of future networks through a set of strict requirements [5]. Some examples are Vehicle to Everything (V2X), Industry 4.0, and Virtual Reality / Augmented Reality (VR/AR). These applications are really challenging in terms of bandwidth (10-20 Gbps), latency (<5 ms), reliability (1 packet lost for every 10^5 packets sent), and scalability (1,000,000 devices/km²), as well as require a flexible network response to cope with the high heterogeneity of devices and contents.

In 5G, the previous concepts are referred to as Enhanced Mobile Broadband (eMBB), Massive Machine to Machine and Internet of Things (IoT) Communication (mMTC), and Ultra-Reliable and Low Latency Communication (uRLLC), respectively. Similarly, IEEE 802.11 groups are also considering these aspects in the design of next-generation amendments, such as High Efficiency (HE) IEEE 802.11ax and Extreme High Throughput (EHT) IEEE 802.11be.

To meet the abovementioned strict requirements, not only a technological innovation is required (e.g., use of more spectrum or improve multiple-antennas technologies), but a paradigm shift is necessary when designing novel solutions for network planning, operation, and management. In particular, intelligent wireless networks need to be empowered with cognitive and context-aware capabilities, which may require additional infrastructure such as environmental sensors and cameras. To that end, ML is expected to be important during the lifetime of 5G and will become pervasive - as included from the beginning in their conception - in 6G networks.

The actual utility of ML lies in those problems that are hard to solve by hand-programming due to their underlying complex patterns (e.g., network traffic prediction). Formally, a machine is said to learn if it improves the performance \mathbb{P} obtained from undertaking task \mathbb{T} , based on the gathered experience \mathbb{E} [6]. Different ML techniques have been categorized in multiple ways, but the most common taxonomy differentiates between supervised learning (labeled data is used for training), unsupervised learning (no labels are used on input data), and reinforcement learning (exploration-exploitation trade-off with label/unlabeled data). Table 1 provides a list of algorithms and potential networking applications for each type of ML techniques, as well as some examples of input data to

be used by these methods. For further details, we address the interested reader to [7–9], which survey a plethora of ML-based applications for networking.

Apart from the specific ML solutions to problems in communications, some efforts have been made towards enabling AI-aware networking in more general terms. In particular, several architectural proposals have been provided so far [10–12]. Most of the referenced works agree in the necessary steps for enabling big data analytics in cellular deployments: (1) data collection, (2) data preparation, (3) data analysis, and (4) decision making. Nevertheless, none of these works provide architectural guidelines to introduce ML to wireless networks. In this regard, the ITU’s architecture looks deeper into the ML operation and targets the actual procedures involving information gathering, processing, and communication. Besides, the ITU-T provides a data handling framework for ML-aware networks [13], which defines processes concerning data collection, processing, and output distribution.

2.2 Machine Learning-Enabled Use Cases in WLANs

To showcase the potential of applying AI in IEEE 802.11 WLANs, we next describe a set of use cases where ML allows improving the network operation.

2.2.1 OFDMA-Based Smart Network Slicing

Network slicing is one of the hottest topics in 5G because it allows virtually separating network resources to meet diverse application requirements. In next-generation WLANs, network slicing can be realized through the allocation of radio resources via Orthogonal Frequency-Division Multiple Access (OFDMA). However, the heterogeneity of applications and devices, and their subsequent elasticity prevent allocating frequency resources easily. To solve this, ML can be used to make predictions on the user requirements so that the access network can be optimized.

As an example, Fig. 1 shows a scenario where multiple users operate under different requirements, based on the applications they use. While the central controller can make predictions on user behavior, the local schedulers may consider information such as the user profile, the current performance, and the environmental circumstances. Accordingly, the Access Point (AP) can allocate the most suitable OFDMA resources to each device, based on the predicted needs and network status.

2.2.2 Cloud-Based User Association and Handover

Most of the current user association and handover procedures held in WLANs typically rely on the Strongest Signal First (SSF) mechanism. This might be problematic in terms of load balancing and can potentially lead to severe performance degradation in dense Basic Service Sets (BSSs). By introducing ML, it is possible to handle contextual information such as the traffic load, which can be useful for decision-making. Furthermore, mobility pattern prediction and

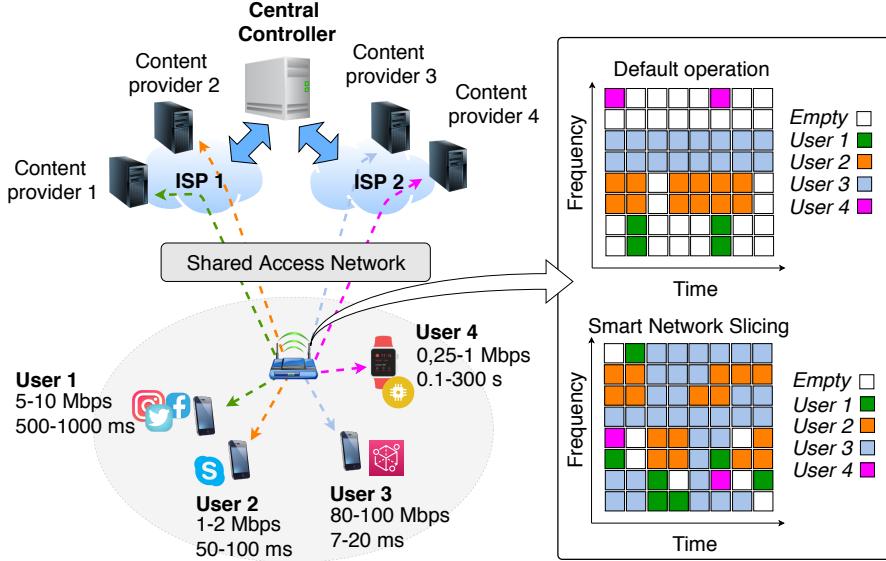


Figure 1: OFDMA-based smart network slicing.

user requirements forecasts can be included in the system, thus empowering the association and handover mechanisms with insightful information.

2.2.3 Inference-Based Coordinated Scheduling

Contrary to traditional cellular-type networks, WLAN deployments can be chaotic, especially in residential scenarios where anyone can set-up an AP and create a wireless network. This typically leads to complex scenarios where inter-BSS interactions prevent the existing scheduling approaches to ensuring a minimum quality of service. Fortunately, ML can be used to infer these interactions and provide a solution accordingly. In particular, through coordinated ML-assisted scheduling, different APs can trigger uplink/downlink transmissions from/to the appropriate stations (STAs), thus increasing the network throughput whilst reducing the number of packet collisions.

2.2.4 Reinforcement Learning-Based Spatial Reuse

Spatial reuse aims to improve channel utilization through sensitivity adjustment mechanisms. However, selecting the best sensitivity threshold is not trivial given the complex spatial interactions that occur among devices. As a potential solution, reinforcement learning can be applied locally to improve spectral efficiency in a decentralized manner.

3 ITU Unified Architecture for Future Networks

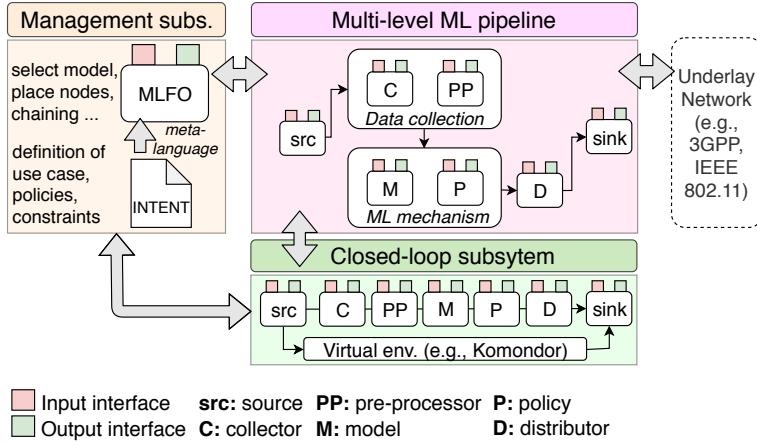


Figure 2: ITU’s logical architecture for future networks [2]. Entities contain input/output interfaces for communication, while the ML intent is a declarative file with information related to the use case.

The FG-ML5G was created in November 2017 by its parent group, the ITU-T Study Group 13, to study the integration of ML mechanisms into future networks. This includes the definition of interfaces, protocols, data formats, and architectures. During its lifetime, the FG-ML5G has released several reports and contributions. Among them, we highlight the ITU’s logical interoperable architecture for future networks [2], which defines an ML overlay that operates on the top of any unspecified underlay network technology (e.g., 3GPP, EdgeX, IEEE 802.11). The ITU’s architecture aims to fulfill a set of technology-agnostic high-level requirements to support ML. For instance, the architecture must be able to support multiple types of data, thus taking advantage of heterogeneous data sources.

Figure 2 shows the elements that compose the ML overlay (management subsystem, multi-level ML pipeline, and closed-loop subsystem). These elements are further described in the following subsections. Based on this standard overlay, ML applications can be instantiated in the logical entities (represented by white boxes).

3.1 Management Subsystem

The management subsystem is in charge of the deployment and the orchestration of the ML services that operate in the underlying network. To that purpose, the Machine Learning Function Orchestrator (MLFO) entity is defined. The MLFO is first instantiated by a declarative intent that uses a meta language. It specifies the ML use case to be applied, including initialization, policies, and

constraints. Then, the MLFO initializes the elements of the ML pipeline and monitors their operation during execution.

3.2 Multi-Level Machine Learning Pipeline

The multi-level ML pipeline performs the actual ML operation in a given network underlay and it is in charge of the data collection, model application, and output distribution. The following logical entities compose the ML pipeline:

- **Source (src):** generates data to be used by the ML mechanism.
- **Collector (C):** collects the data generated by sources.
- **Pre-processor (PP):** prepares the data collected for its utilization by the ML mechanism.
- **Model (M):** applies the ML model specified by the intent.
- **Policy (P):** provides a set of constraints and/or guidelines that delimit the behavior of the model.
- **Distributor (D):** spreads the ML output across all the corresponding targets (or sinks).
- **Sink (sink):** applies the ML output that is received from the distributor.

3.3 Closed-loop Subsystem

In order to address network dynamics, the ML operation is assisted by a closed-loop subsystem, which can provide information to the system beforehand. As for the ML pipeline, the closed-loop subsystem is orchestrated by the management subsystem. In particular, a sandbox can be formed of real devices (pre-production internal network) or even be virtual (simulator/emulator). Network simulators such as ns-3 and Komondor [14] are examples of closed-loop subsystems and can serve two purposes: *i*) generate synthetic data for training, and *ii*) run simulations to validate potential solutions before being applied in production.

4 Machine Learning-Aware Architecture for IEEE 802.11 WLANs

Based on their independence degree in terms of management and operation, WLAN deployments can be divided into two main families:

- **Enterprise:** a set of BSSs can be jointly operated from the edge and/or the cloud, thus providing management and orchestration functionalities such as centralized authentication, or channel allocation. Enterprise-like deployments are realized through Extended Service Sets (ESS) and can be

typically found in environments controlled by a single network operator, like university campuses, offices, stadiums, etc.

- **Residential:** each BSS is responsible for its own management and operation. In the context of residential scenarios (but not limited to), peer-to-peer deployments are gaining popularity for infrastructureless communications (e.g., Wi-Fi direct).

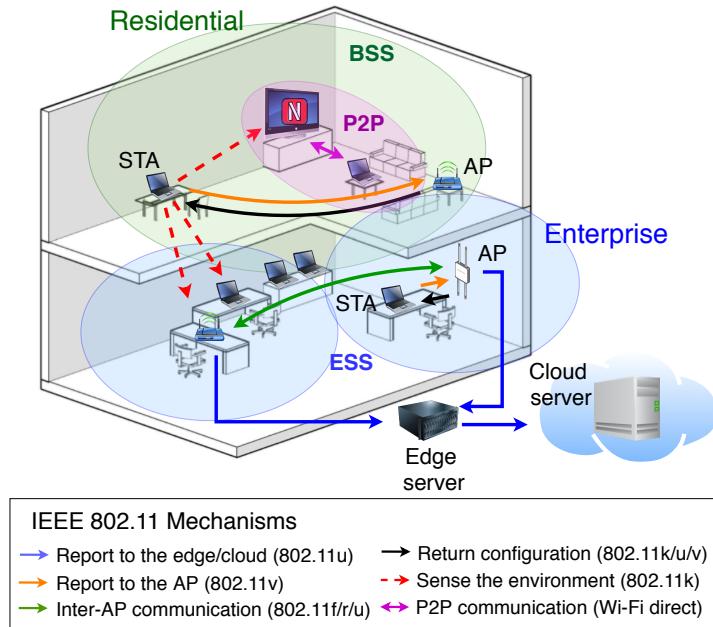


Figure 3: Enterprise and residential-like deployments and complementary IEEE 802.11 mechanisms to enable the utilization of ML.

Figure 3 illustrates the enterprise and residential-like deployments as well as a set of mechanisms that can facilitate the adoption of the ML-based architecture in WLANs. The following functionalities are provided:

- **Information gathering (802.11k/r/v):** ML mechanisms can use information about the network topology and RF measurements to infer the behavior of other devices, or to extract important environmental characteristics.
- **Interoperability (802.11f/u):** Interoperability enables coordinated operations (e.g., scheduling, resource allocation), thus allowing to apply centralized/coordinated mechanisms such as in federated learning.
- **Security (802.11w):** ML mechanisms can use management frames that are protected so that a higher level of security is granted.

- **Validation (802.11t):** Performance evaluation in WLANs through test metrics can be useful to define optimization goals within the ML operation.

4.1 Challenges in IEEE 802.11 WLANs

The application of ML methods in WLANs is tightly tied to the technological challenges posed by these types of networks. The major challenges encountered in wireless communications stand for fast data expiry and lack of resources for data handling (e.g., storage, computation, and information exchange). Regarding Wi-Fi networks, we find the following challenges:

1. **Non-stationarity:** channel fluctuations due to multipath fading, mobility of users and varying traffic needs entail a big challenge to ML applications. As a result of network dynamics, adaptability should be granted by continuously retraining ML models.
2. **Limited communication resources:** since Wi-Fi operates under unlicensed bands, resources are scarce and shared. Hence, any potential communication required by a certain ML mechanism (as for distributed learning) may fail or be delayed if the medium is congested. As a result, the ML operation must be robust and resilient enough to react to potential communication issues.
3. **Limited computation and storage resources:** computation and storage resources may also be scarce in WLANs, especially in residential-like deployments. Therefore, the ML operation should include computation-efficient procedures. Another implication of limited resources lies in the availability of information to be used by ML algorithms, especially for online learning methods.
4. **Adversarial environment:** in many cases, Wi-Fi deployments are chaotic in the sense that many overlapping BSSs coexist without cooperation. This is a particularly interesting challenge for ML methods, where competition among agents may lead to an adversarial setting. Moreover, multi-vendor devices may implement different ML mechanisms, leading to clashing interests.
5. **Legacy devices:** BSSs may coexist with other legacy devices that do not perform any intelligent operation. It is then required for ML methods to be aware of those devices, so that unfair situations are avoided.

Apart from the previous WLAN-specific challenges, other inter-domain issues should be considered. For instance, end-to-end security is required since ML mechanisms store and/or exchange sensitive data that may be exposed. Besides, interoperability should be tackled when deploying ML solutions to different underlay networks. In this regard, the standardized ITU ML pipeline stands up as a promising solution.

4.2 Computation Paradigms in IEEE 802.11 WLANs

The various types of WLAN deployments and their computation and communication capabilities are closely linked to the type of ML solutions that can be applied to them: *cloud* or *edge-oriented*.

Cloud-oriented ML applications are characterized by bearing high computational and storage resources, thus allowing them to collect various types of data from multiple sources, and to provide global and long-term solutions. The major challenge for cloud-oriented methods lies in the management of data and the corresponding synchronization, availability, and heterogeneity issues.

In edge-oriented mechanisms, the ML operation is mainly ruled by edge devices (e.g., APs and/or STAs), which, contrary to the cloud approach, typically lack powerful computation and storage resources. In consequence, edge-oriented mechanisms may only allow using simple and lightweight computing ML algorithms. Nevertheless, edge servers can be added to deploy more powerful solutions promptly. The edge-oriented approach is useful for real-time ML applications that manage local (and even highly-varying) information.

Apart from cloud and edge-oriented settings, we may distinguish between methods based on their cooperation degree. In cooperative approaches, nodes interact among them for the sake of jointly conducting the learning operation (e.g., sharing a reward). However, reliable and timely connections among learners are often required. In this regard, [15] showed the role of communications on speeding up a distributed training procedure over a set of nodes in a network. Alternatively, for the non-cooperative case, the learning operation may lead to adversarial settings, especially since BSSs share resources such as the spectrum.

4.3 Realization of the ML-Aware Architecture for WLANs

To showcase the adoption of the architecture, let us retake the AP (re)association and handover example (see Fig. 4). We now consider a hybrid solution where two main ML-based processes are held: training (learn from data) and placement (apply the learned knowledge).

While the training procedure is carried out at the cloud (collect data from multiple sources), the placement operation is done at the edge (provide timely responses to new cases). Notice that the system can also be re-trained during the placement phase, based on newly acquired local data.

Specifically, the training procedure consists of the following steps (shown in red):

1. **Data collection:** the cloud server collects information of different kinds from APs and STAs, such as user information (e.g., location), performance (e.g., delay), application data (e.g., traffic load), or channel status reports (e.g., sensed interference). This information can be used either for training or feeding auxiliary algorithms that help the main AP association procedure (e.g., predict user behavior).
2. **Pre-processing:** the data collected at the cloud is pre-processed so that

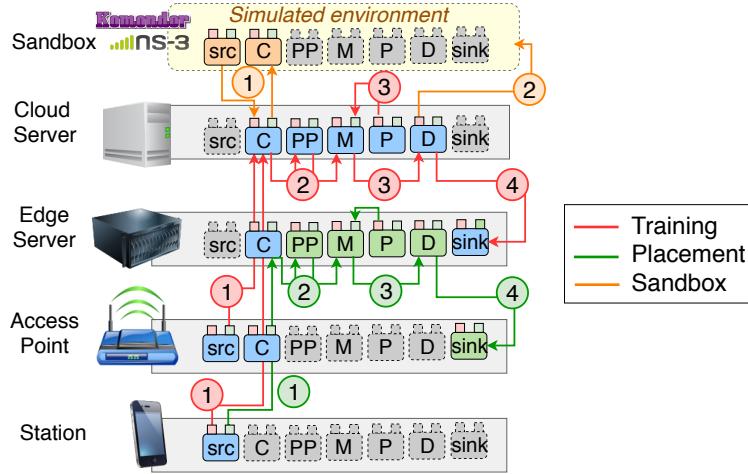


Figure 4: Realization of the ITU’s ML architecture for IEEE 802.11 WLANs through a hybrid ML-based solution for AP (re)association and handover.

the ML method can properly manage it. For instance, in case of applying a multiple linear regression, the input information needs to be converted into normalized features (i.e., convert the rate given in Mbps into a scalar between 0 and 1).

3. **Model generation:** when generating the ML model, certain policies need to be considered. For instance, an AP may set a maximum number of associated STAs. The policies are strongly tied to the capabilities of the devices or the existing regulations (e.g., maximum regulated transmission power).
4. **Output distribution:** once the ML method in the cloud generates the output (i.e., the predicted function for new (re)associations), it is distributed throughout the sink edge servers, which are then ready to give quick response to new cases.

In the placement phase (shown in green), we find:

1. **Handle new requests:** new (re)association requests or potential handovers are detected based on newly acquired information from STAs. This information is collected by the edge server.
2. **Pre-processing:** the acquired information is then processed by the edge server, just like for the training phase.
3. **Run the ML solution:** the ML method provided by the cloud is applied locally at the edge server, which provides an output for the new request.

4. **Apply the ML solution:** the (re)association decision is distributed to the corresponding AP.

Finally, it is worth pointing out the role of the sandbox, which can be mainly twofold (shown in orange):

1. **Generate data for training:** the sandbox can act as a source in the ML pipeline by generating synthetic data for training purposes. Nevertheless, the data provided by the sandbox is limited to several factors such as the accuracy of the simulation model or the degree of similarity between the sandbox and the real network.
2. **Preliminary model testing:** alternatively, the sandbox can be used to validate the output of the ML method before being applied to the real network.

To showcase the potential of the ML-based architecture through numerical results,¹ we compare the performance of classical AP association procedure (SSF) against a novel ML-based approach (based on vanilla neural networks). In particular, the neural network predicts the throughput that an STA will obtain after associating to a given AP based on a set of features or characteristics (e.g., current load, received signal strength). Figure 5 shows the throughput received by each STA versus the load it generates, for different deployment densities. We observe that the ML approach improves the average performance and balances the results obtained by all the STAs. This is because the ML function can capture complex patterns from dense deployments, thus guaranteeing minimum throughput requirements to STAs (at the expense of missing the maximum performance peaks).

5 Concluding Remarks

Current networks are not yet prepared for the pervasive adoption of ML-based operation. Hence, disruptive architectural changes are required. For the sake of moving forward in this field, this article introduced the ITU’s unified architecture for future networks and provided a realization for IEEE 802.11 WLANs. The different forms of Wi-Fi networks allow uplifting the flexibility characteristic of the ITU’s architecture, thus enabling from edge to cloud-oriented solutions, including hybrid approaches.

To conclude, future wireless networks are envisioned to share a common flexible architecture that allows a fast adaptation of resources to accommodate

¹Given the novelty of the technologies studied in this paper, our results have been obtained from well-known standard-compliant models, hence their accuracy is tied to them. Nevertheless, this is a first step to understand the potential benefits of using an ML-based architecture in next-generation wireless networks. For the sake of reproducibility and disclosure, all the source code is open and publicly available at https://github.com/fwilhelmi/machine_learning_aware_architecture_wlans, accessed on Jan. 31, 2020.

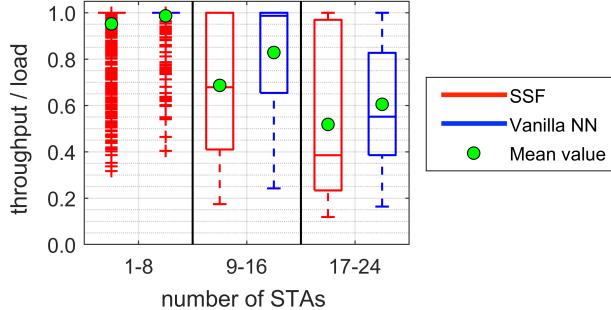


Figure 5: Performance evaluation of the AP association problem in WLANs: SSF versus Neural Network (NN). The mean performance of each mechanism is represented by a green dot.

a plethora of ML-enabled verticals. Nevertheless, a lot of effort is still required before reaching fully intelligent wireless networks. Among several open issues, we highlight the ones related to data handling (*how/where to store data? how to assess the expiry of data?*), orchestration (*how to distribute the ML operation? how to deal with heterogeneity?*), and robustness of the ML methods (*how to deal with uncertainty? how to prevent unprecedented events?*).

Acknowledgment

This work has been partially supported by grants MDM-2015-0502, 2017-SGR-11888, by WINDMAL PGC2018-099959-B-I00 (MCIU/AEI/FEDER,UE), by a Gift from the Cisco University Research Program (CG#890107) Fund, and by SPOTS project (RTI2018-095438-A-I00) funded by the Spanish Ministry of Science, Innovation and Universities. The work by Sergio Barrachina-Muñoz is supported by an FI grant from Generalitat de Catalunya.

References

- [1] Afif Osseiran *et al.*, “Scenarios for 5G mobile and wireless communications: the vision of the METIS project,” *IEEE Comm. Magazine*, vol. 52, no. 5, pp. 26–35, 2014.
- [2] ITU-T Rec. Y.3172, “Architectural framework for machine learning in future networks including IMT-2020,” 2019.
- [3] 3GPP TR 23.791 V16.2.0 (2019-06), “Study of Enablers for Network Automation for 5G,” 2019.
- [4] ETSI GS ZSM 002 V0.13.5 (2019-07), “Draft Zero-touch network and Service Management (ZSM); Reference Architecture,” 2019.

- [5] ITU-T Supp. Y.Supp55, “Machine learning in future networks including IMT-2020: use cases,” 2019.
- [6] Mitchell, Tom M and others, “Machine learning,” *Burr Ridge, IL: McGraw Hill*, vol. 45, no. 37, pp. 870–877, 1997.
- [7] Chunxiao Jiang *et al.*, “Machine learning paradigms for next-generation wireless networks,” *IEEE Wireless Comm.*, vol. 24, no. 2, pp. 98–105, 2016.
- [8] Zhang, Chaoyun and Patras, Paul and Haddadi, Hamed, “Deep learning in mobile and wireless networking: A survey,” *IEEE Comm. Surveys & Tutorials*, vol. 21, no. 3, pp. 2224–2287, 2019.
- [9] Muhammad Usama *et al.*, “Unsupervised machine learning for networking: Techniques, applications and research challenges,” *IEEE Access*, vol. 7, pp. 65 579–65 615, 2019.
- [10] Suzhi Bi *et al.*, “Wireless communications in the era of big data,” *IEEE Comm. Magazine*, vol. 53, no. 10, pp. 190–199, 2015.
- [11] I Chih-Lin *et al.*, “The Big-Data-Driven Intelligent Wireless Network: Architecture, Use Cases, Solutions, and Future Trends,” *IEEE Vehic. Tech. Magazine*, vol. 12, no. 4, pp. 20–29, 2017.
- [12] Mowei Wang *et al.*, “Machine learning for networking: Workflow, advances and opportunities,” *IEEE Network*, vol. 32, no. 2, pp. 92–99, 2018.
- [13] ITU-T Rec. Y.3174, “Framework for data handling to enable machine learning in future networks including IMT-2020,” 2019.
- [14] Sergio Barrachina-Muñoz *et al.*, “Komondor: a Wireless Network Simulator for Next-Generation High-Density WLANs,” in *Wireless Days*. IEEE, 2019, pp. 1–8.
- [15] Yujun Lin *et al.*, “Deep Gradient Compression: Reducing the Communication Bandwidth for Distributed Training,” in *ICLR*, 2018; <https://openreview.net/forum?id=SkhQHMW0W>, accessed on Jan. 31, 2020.

Usage of Network Simulators in Machine-Learning-Assisted 5G/6G Networks

Francesc Wilhelmi, Marc Carrascosa, Cristina Cano,
Anders Jonsson, Vishnu Ram, and Boris Bellalta

Abstract

Without any doubt, Machine Learning (ML) will be an important driver of future communications due to its foreseen performance in front of complex problems. However, the application of ML to networking systems raises concerns among network operators and other stakeholders, especially regarding trustworthiness and reliability. In this paper, we devise the role of network simulators for bridging the gap between ML and communications systems. Network simulators can facilitate the adoption of ML-based solutions by means of training, testing, and validating ML models before being applied to an operative network. Finally, we showcase the potential benefits of integrating network simulators into ML-assisted communications through a proof-of-concept testbed implementation of a residential Wi-Fi network.

1 Introduction

Beyond the fifth-generation (5G) of mobile communications systems, namely the sixth generation (6G), Artificial Intelligence (AI), and more precisely Machine Learning (ML), are expected to be pervasively included as part of the network operation, which would entail a huge leap towards optimization, automation, and self-healing. This is possible thanks to the paradigm shift driven by the softwarization of networks – achieved through Software Defined Networks (SDN) and Network Function Virtualization (NFV) – which provides the necessary flexibility to empower data-driven approaches.

The integration of ML to communications has started to be considered for the upcoming versions of 5G. This fact is supported by the content already approved by the 3rd Generation Partnership Project (3GPP) for Release 16 (2020) and Release 17 (2021) [1], which aim to continue improving the efficiency of 5G systems in many domains such as interference mitigation, Self-Organizing Networks (SON) and Big Data, power consumption, and user mobility, to name a few. Besides, we find of high relevance the contributions made by the International Telecommunication Union (ITU) Focus Group on Machine Learning for 5G and Beyond (FG-ML5G) and the Study Group 13 (SG13), which have published specifications on an ML-aware architecture [2, 3].

Through the exploitation of the rich amount of available data, ML can overcome the systemic complexity inherited from novel use cases like Vehicle to Everything (V2X) communications, Machine Type Communications (mMTC), and extended reality and high-quality video content delivery. These use cases comprise heterogeneous scenarios with mobility, a huge number of devices, and high-bandwidth and low-latency requirements. In particular, ML may offer substantial performance gains due to the inherent flexibility of automatically learning diverse situations, thus allowing to solve problems related to interference management, improving spatial reuse, or efficient resource allocation.

While ML promises significant productivity gains, it also raises serious challenges and concerns. First of all, the successful application of ML depends on the quality of the training data provided. These data, by nature, can often be limited or noisy, and draw insightful conclusions might be challenging for many problems. Apart from that, dealing with non-stationary data is still an open challenge, which casts doubts on the validity of potentially learned models. A prominent example is that of IEEE 802.11 Wireless Local Area Networks (WLANs). The typical decentralized nature of WLANs (e.g., residential deployments) affects data collection and also leads to complex and highly non-stationary environments.

These challenges put into question the worthiness of introducing ML to networking systems. In particular, network operators and other stakeholders may have concerns regarding architectural (e.g., how to train and transfer ML models across a network) and operational aspects (e.g., how to provide trustworthy ML optimizations). While significant efforts have been put towards designing ML-based network architectures [1–4], only a small number of works have been devoted to study and address the side effects that ML can produce when applied to networks.

In this paper, we devise the usage of network simulators to enable the paradigm shift towards ML-assisted communications. Network simulators play a crucial role both in academia and industry. By prototyping complex problems and systems, simulators are key to provide insights on the potential gains of new features and technologies, thus boosting innovation. In this regard, we believe that network simulators can contribute to providing reliable and robust ML mechanisms for communications. To the best of our knowledge, this is the first work on addressing this emerging issue. The main contributions of this paper are as follows:

- We discuss the main aspects related to the reliability of ML for future communications.
- We devise the usage of simulators for training, testing, and evaluating the performance of ML models for communications.
- We showcase the potential integration of network simulators within the ITU ML-aware architecture, which is an adaptable and interoperable framework for realizing specific ML-based network functionalities.

- We provide insights on practical aspects for their integration to ML-assisted communication systems.
- We illustrate the potential advantages of using simulators into ML-assisted networks by applying the outcome of an ML-driven simulation to a residential WLAN testbed.

2 Reliability of Artificial Intelligence for Communications

ML has shown great potential for improving a plethora of applications in communications (see, for instance, the surveys in [5–9] and the references therein). Much of the credit resides in the extraction of useful information from large amounts of data. For instance, the authors in [7] show that autonomous Unmanned Aerial Vehicles (UAV) can be empowered by Artificial Neural Networks (ANN). In particular, on-time decisions such as the flying direction can be optimized based on the collected data (e.g., users’ location, available resources, or wireless environment). These data, which may come from multiple sources, can be exploited and comprehended by the ANN for the sake of optimization.

Despite the abovementioned efforts towards designing ML-based solutions, less attention has been paid to overcome the potential negative impact of ML in communications. The fact is that many ML approaches are seen as black boxes due to the non-linearity of their output (e.g., a prediction), especially when dealing with high dimensional spaces. This is accentuated in Deep Learning (DL), where neurons at multiple hidden layers may have different behaviors. Despite it is possible to obtain a certain intuition on the way a neural network operates (e.g., through visualization tools), the logic behind some processes remains unknown.

The uncertainty associated with ML methods can lead to performance degradation when applied to networks. For instance, an online learning mechanism that is driven by exploration-exploitation may fail to comply with Service Level Agreements (SLAs). The fact is that exploration triggers configuration settings which may lead to undesired performance. This is a critical aspect to take into consideration since many applications rely on certain minimum requirements to operate, and not meeting them could be even dangerous (for instance, consider networking applications for autonomous driving). As a result, the application of ML can raise concerns and lead to mistrust when applied to networks.

To address the lack of confidence that ML may generate, network simulators can be used for training, testing, and evaluating the effect of ML models before being applied to operative networks. In particular, simulators can provide diverse functionalities to enhance the confidence level of future ML-assisted networks:

1. **Validate the output of ML models:** a simulator can be used to test and evaluate the output of a certain ML optimization before being applied to a production environment.

2. **Assess the impact of ML models on networks:** apart from evaluating the performance of a given ML model on specific networking functionalities, it is important to study the effect that ML has on the rest of the network. The whole procedure can be simulated together if the simulator includes ML functionalities, which is the case, for instance, of ns-3 and Komondor.
3. **Generate training data:** sometimes, training data extracted from network devices can be sparse, limited, incomplete, or incoherent. To address this, simulators can generate synthetic data, which would broaden the available training data sets. However, assessing the quality of synthetic data sets can be challenging for operators, especially concerning complex problems that cannot be modeled accurately. For that reason, it is important to monitor the effects of applying ML models trained with synthetic data on operative networks.
4. **Train ML models:** with a strong connection to the two previous points, ML models can be trained in a simulation environment. As an example, consider the case where online learning is performed during the simulation.
5. **Complement ML models:** simulators can also contribute to filling the intersection between model-based and data-driven approaches. The fact is that simulators can act as *experts* to assist the operation of ML algorithms. As an example, random initialization is typically employed for ML methods, which sometimes leads to converging to suboptimal saddle points. By adding additional knowledge from simulations, the learning procedure can be improved.

Apart from the utilization of network simulators, we find other ways to enhance the reliability of AI mechanisms such as *explainable AI* [10] and *safe Reinforcement Learning (sRL)* [11]. Explainable AI is based on the interpretation of AI-based decisions, which is useful to devise the impact of potential optimizations and predict misbehavior. However, explainable AI is not mature enough, and the existing techniques are mainly based on visualization, so they are subjective and may lead to misinterpretation. For that reason, explainable AI currently lacks applicability for enhancing the reliability of ML-assisted communications.

Regarding sRL, it aims to minimize the negative effects that unconstrained exploration methods can incur during a learning procedure. This can be achieved either by adding extra information to the exploration procedure (e.g., external advice), or by applying certain risk-aware criteria (e.g., exploration based on water-filling methods). While sRL is useful to mitigate the randomness of exploration, its application may provide moderate improvements and lead to slow optimization when applied to networks, which can be worsened in non-stationary systems. Besides, sRL is restricted only to RL mechanisms, thus leaving open the challenges posed by other kinds of mechanisms such as DL.

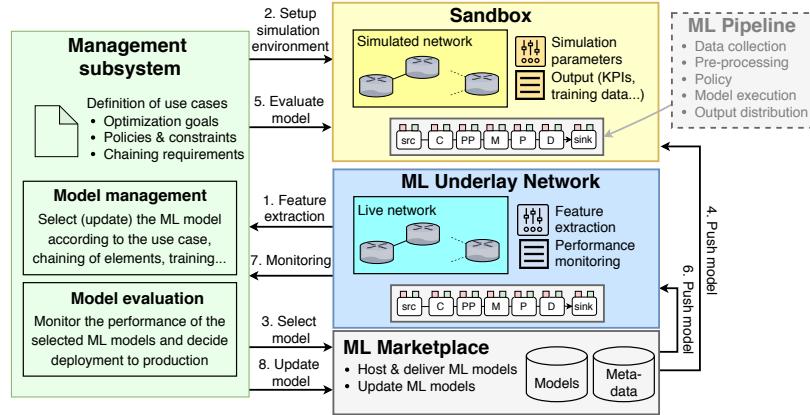


Figure 1: Architectural elements and procedures for evaluating the output of ML models.

3 Network Simulators to Enable Artificial Intelligence in Communications

In this Section, we describe the architectural aspects of integrating network simulators to ML-assisted communications. Besides, we analyze the key features and requirements for simulators to be included in an ML-based networking architecture.

3.1 Architectural Integration

Most of the existing simulation platforms have no relation with AI/ML techniques, nor have any integrated module for evaluating and training ML models. Moreover, current simulated network functionalities are typically too specific (e.g., simulate the effect of multiple antennas on the PHY layer performance), and seldom support open interfaces, as a result of being developed by focused academic or industrial organizations. To enable the next generation of ML-based communication systems, it is imperative to design interoperable mechanisms for simulated networks and ML mechanisms. For that purpose, we find of high relevance the ITU ML architecture defined in [2].

The ITU ML architecture defines a set of logical components, interfaces, and procedures to realize ML-assisted communications. For a complete overview of ITU architecture, we refer the interested reader to the work in [12], which proposes a realization for future IEEE 802.11 WLANs, an important part of the 5G/6G ecosystem in unlicensed bands. In particular, the ML-aware architecture is composed of the following elements:

- **Management subsystem:** this element is responsible for the management and orchestration of the ML operation in a network. The respon-

sibilities of this module range from data collection to model deployment and monitoring.

- **ML underlay network:** network at which the ML optimization is applied.
- **Sandbox:** evaluation domain that includes the usage of network simulators.
- **ML marketplace:** container of ML models that are applied to the ML underlay networks.
- **ML pipeline:** set of elements that interact with underlay networks to perform the ML optimization.

To integrate simulators in the loop of ML-assisted networks, standardization of elements, interfaces, and data handling procedures is key. This is captured by ITU architecture through the sandbox subsystem. The sandbox is an isolated domain for reproducing the behavior and operation of live networking systems, which is useful to evaluate the performance of ML models before being deployed in production environments. Network simulators can be included in the sandbox and used to evaluate and train ML models. To that end, interoperability allows building end-to-end ML pipelines in simulated network underlays.

To illustrate the integration of network simulators within the high-level ITU architecture, Fig. 1 depicts an example where the output of an ML model is evaluated at the sandbox before being applied to the operative network. The involved procedures are as follows:

1. The management subsystem extracts features from the ML underlay network.
2. Based on the characteristics of the ML underlay network, the simulation environment is prepared.
3. The management subsystem selects the ML model from the marketplace, according to the meta-data describing the use case, the optimization goals, and the available ML models.
4. The ML model is pushed to the sandbox to be applied to the simulated network.
5. The ML model is evaluated in the simulator. Evaluation of other ML models may be considered upon unsuccessful results.
6. Once the evaluation is successfully done, the ML model is pushed to the operative network, where the ML optimization takes place.
7. The network performance is monitored, as well as new data is gathered.
8. The information obtained from monitoring is used to update the ML models and/or metadata in the marketplace.

The ML output evaluation procedure allows devising the potential benefits and drawbacks of using a certain ML-based optimization in a network. The fact is that ML outputs can sometimes look surprising from the perspective of a network operator, and their effect on the network may be unknown a priori. This is accentuated in complex problems for which ML is entailed to outperform legacy solutions because the knowledge on the problem is limited.

3.2 Practical Integration Aspects

To simulate multiple types of scenarios, technologies, and network functionalities, we find a plethora of proprietary and open-source network simulators (e.g., ns-3, OMNET++, OPNET, NetSim, Komondor). Apart from network technologies, we must take into account the capability of the different simulators to capture other specific phenomena in detail when required. This is the case, for instance, of Simulation of Urban MObility (SUMO) and UnderWater simulator (UWsim), which simulate vehicular urban mobility and underwater physical effects, respectively, and can be used along with OPNET and ns-3 simulators.

When it comes to integrating simulators into ML-assisted networks, a set of challenges arise with respect to execution, interoperability, and portability aspects:

1. **Execution:** to test, train, and evaluate the performance of ML methods in simulators, it is important to reproduce the behavior of the target operative network. For the proper integration of simulators into the ML-aware architecture, it is required to transfer simulation-related meta-data to the elements of the ML pipeline. This includes supported technologies and network functionalities, maturity of simulation blocks (e.g., beta release), and the potential number of domains the simulators can span (e.g., from core to access network).

Concerning pluggable ML functionalities, built-in ML modules can boost the procedures for simulating the behavior of ML mechanisms or training ML models in the sandbox. A few existing simulators support ML functionalities, but we find the framework connecting ns-3 with OpenAI Gym [13], and the agent-based implementation in Komondor.

Apart from supported capabilities, short execution and configuration times can serve to empower ML-driven real-time applications. First, we consider the time it takes for the simulator to generate a given output, which may indicate the tractability of simulating large-scale scenarios. Second, fast reconfiguration of network simulators would allow following potential changes on the operative network (e.g., user demands, available resources, policies, etc.). For instance, an update of policies should be reflected in the simulation domain, so that operators' requirements can be fulfilled.

2. **Interoperability:** an important requirement lies in the degree of flexibility of simulators for interacting with the components of the ML-aware

architecture. Interoperability is therefore meant to enable a seamless integration of intelligent network functionalities in the communication network. For that, it is imperative that the simulated network functionalities are managed using the same operation and maintenance mechanisms as for the network functionalities in the ML underlay. This can be achieved through standard Application Programming Interfaces (APIs). Features that may facilitate the interoperability of out-of-the-box simulators are the support for Command-Line Interface (CLI) execution mode, the level of monitoring supported (real-time, batch, model-based, etc.), and automation of data collection and in applying the ML output in the simulator (e.g., reading from log files vs. API-based interface with ML functions).

3. **Portability:** network simulators are written in multiple programming languages (e.g., C/C++, Java) and supported by different specific platforms. Thus, portability is another important requirement for simulators. In this regard, containerization (e.g., via Docker) can be of great utility and allow network operators to deploy simulators in a flexible manner. Apart from that, parallelization is important to determine, for instance, the number of ML pipeline nodes and simulated network functionalities that the simulator can support at any instant.

3.3 Accuracy of Network Simulators

The degree of reliability of a network simulator depends on its accuracy on reproducing the actual real phenomena. In other words, simulations must be as close as possible to reality. This topic was previously addressed in [14], where the authors defended that simulators do not really fit the actual behavior of networks, based on experimental results in a MANETs testbed. Nevertheless, it was also shown that simulation results can serve as a good upper-bound for testbed setups.

In general, network simulators accurately reproduce the behavior of protocols in higher levels of the TCP/IP stack. However, they can fail at characterizing complex physical phenomena such as radio propagation, antenna radiation, or energy consumption. As a result, network simulators typically provide accurate qualitative performance results and help to predict the behavior of real networks under certain circumstances. In contrast, results may lack quantitative precision, thus deviating from the exact performance that will be then experienced in real networking systems. Alternatively, hybrid approaches can be employed for simulating certain layers (e.g., MAC) while taking advantage of the actual interactions that occur in real implementations. Unfortunately, and to the best of our knowledge, there is little literature on this topic.

4 Use-case: Power Control in Residential WLANs

To illustrate the potential of integrating simulators to ML-assisted networks, we provide a testbed implementation of an IEEE 802.11 WLAN that suffers from

starvation due to the high sensed interference of a residential environment. To address this problem, a joint ML-based solution is simulated and then provided to the testbed devices.

4.1 From Testbed to Simulation Domain

The considered testbed implementation comprises two overlapping Basic Service Sets (BSSs) in a residential environment, which are characterized by being highly dense and uncoordinated. The decentralized nature of WLAN deployments in a neighborhood may lead to high interference, which can be extremely variable due to the heterogeneous usage of the network and the complex physical phenomena that can occur. The non-stationarity of residential environments is, therefore, one of the critical aspects to be considered when designing dynamic solutions for improving network performance. Hence, the usage of network simulators can contribute to reducing the performance losses originated by transitory phases (e.g., exploration in online learning).

Our proposed testbed-simulator integration is illustrated in Fig. 2, where the ML solution is provided by a simulated version of the testbed. Two identical BSSs are deployed in a high-density residential scenario. However, since they are positioned at different locations, they are subject to different interference conditions, and so offer different performance. The characterization of the WLAN testbed is done with the IEEE 802.11ax-oriented Komondor simulator, which includes the operation of agents for simulating the behavior of ML mechanisms when plugged into wireless nodes.¹

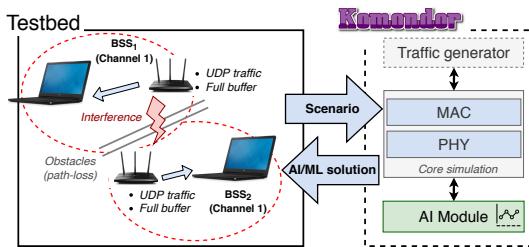


Figure 2: Use case application of the Komondor simulator to apply ML in a testbed WLAN.

Through the procedures that have been previously illustrated in Fig. 1, the testbed scenario is first characterized in the simulator by gathering parameters such as the location of nodes, path-loss effects, or the traffic load. As an example of the characterization of the testbed in the simulator, consider the path-loss model selected, which is chosen based on the degree of similarity with respect to testbed measurements. After preparing the simulation environment, the ML

¹All the details of the experimental part and source code are open and available at the following repository: https://github.com/fwilhelmi/usage_of_simulators_in_future_networks, accessed on May 15, 2020.

model is applied in the simulator for the sake of improving a certain performance metric. Finally, the optimized ML-based configuration is passed and applied to the real devices, in which performance is expected to be enhanced.

4.2 Machine-Learning-based Transmit Power Control

To improve the performance of the target WLAN, we simulate a Multi-Armed Bandits (MABs) application for Transmit Power Control (TPC), as previously done in [15]. We take an online learning approach to address the complexity of spatial interactions in WLANs, where the effect of tuning the transmit power can be hindered. Accordingly, the MABs framework is useful to reduce the complexity of the problem and effectively improving the performance at a low computational cost.

This use case is particularly revealing since the transmit power is a critical parameter to be freely adjusted, and trying several configurations before finding the best performance may lead to unpredictable effects during the transitory regime. Moreover, commercial equipment typically offers a high delay when changing the transmit power or other parameters such as the primary channel. As a result, network simulators can play a crucial role in palliating the negative impact that exploration can have in communications.

Figure 3 illustrates the temporal throughput obtained by each BSS when simulating the MABs approach for tuning the transmit power. Also, the performance that would be obtained by both BSS when using the default configuration is illustrated. As shown, both BSSs experience an unstable transitory regime before reaching a stable state whereby performance is improved. Among a set of input transmit power levels, the most popular one to be used by both BSSs is 7 dBm, which, based on simulation results, is expected to improve the average throughput by 88.48%.

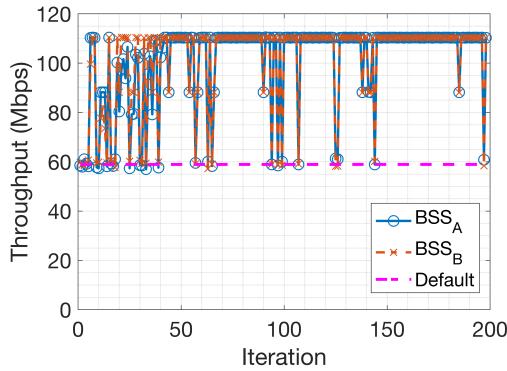


Figure 3: Simulated throughput evolution after applying MABs for tuning the transmit power in an OBSS. Each learning iteration corresponds to 5 seconds in the simulation.

Finally, we give some insights on the time it takes the simulator to bring up results for the testbed. To include the operation of simulators in future networks (especially for real-time applications), it is very important to find an equilibrium between the stability of the output and the time it takes to generate it. Figure 4 shows the variability obtained on the simulation results, for different simulation time values. The execution time is also displayed. As observed, the higher the simulation time, the higher the stability is. However, this is paid with execution time, which varies for different network simulators.

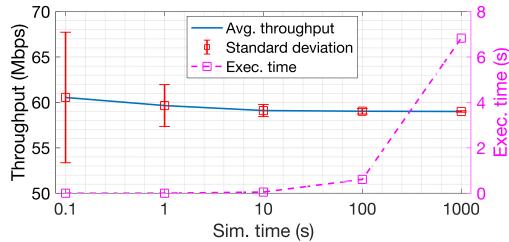


Figure 4: Execution time versus variability of the results in Komondor simulator.

4.3 Testbed results

Now, we show the results of applying the configuration suggested by the simulator on the testbed. Figure 5 compares the performance of applying the ML-based configuration (both BSSs use a transmit power equal to 7 dBm) with that used by default (i.e., 23 dBm).

As shown, both BSSs improve their throughput significantly by using the configuration suggested by the simulator. While BSS₁ improves its throughput by 76.16 %, BSS₂ experiences a 93.98 % improvement. Besides, based on the lower number of observed outliers, we notice higher stability in terms of throughput variability (especially for BSS₁). Note, as well, that BSS₂ suffers drops for some throughput values, which are originated by the high channel variability found at the residential environment the tests were performed.

5 Concluding Remarks

Future communications are expected to evolve towards automated systems enabled by ML. However, the application of ML to networking systems can generate instability and degrade KPIs. To address that, we envision the integration of sandbox environments for ML-assisted networks. In particular, we find network simulators of great utility for training, testing, and evaluating the performance of ML models before being deployed to production environments. In this article, we devised the potential usage of network simulators for future ML-based communications and provided insights on integration aspects. Our testbed re-

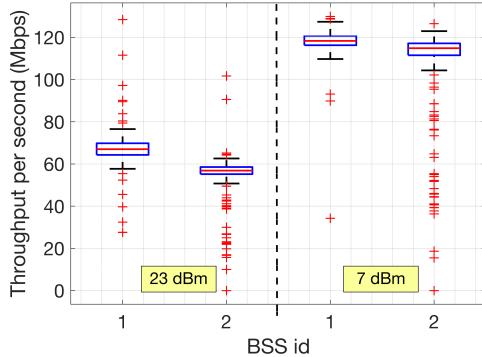


Figure 5: Performance comparison of default (23 dBm) and ML-based (7 dBm) configurations at the testbed WLAN.

sults in a residential IEEE 802.11 WLAN showed how network simulators allow mitigating the negative effects of directly applying ML in the operative network.

Network simulators are expected to contribute to filling the gap between AI and communications. Nevertheless, a lot of effort is still needed with regards to the architectural integration of simulators into ML-assisted networks. The most important challenges lie in the definition and implementation of standardized interfaces.

Acknowledgment

This work has been partially supported by grants MDM-2015-0502, WIND-MAL PGC2018-099959-B-I00 (MCIU/AEI/FEDER,UE), 2017-SGR-11888, and by SPOTS project (RTI2018-095438-A-I00) funded by the Spanish Ministry of Science, Innovation and Universities.

References

- [1] 3GPP TR 23.791 V16.2.0 (2019-06), “Study of Enablers for Network Automation for 5G,” 2019.
- [2] ITU-T Rec. Y.3172, “Architectural framework for machine learning in future networks including IMT-2020,” 2019.
- [3] ITU-T Rec. Y.3174, “Framework for data handling to enable machine learning in future networks including IMT-2020,” 2019.
- [4] ETSI GS ZSM 002 V0.13.5 (2019-07), “Draft Zero-touch network and Service Management (ZSM); Reference Architecture,” 2019.

- [5] Klaine, P. V. et al., “A survey of machine learning techniques applied to self-organizing cellular networks,” *IEEE Comm. Surveys & Tutorials*, vol. 19, no. 4, pp. 2392–2431, 2017.
- [6] Zhang, C., Patras, P. and Haddadi, H., “Deep learning in mobile and wireless networking: A survey,” *IEEE Comm. Surveys & Tutorials*, vol. 21, no. 3, pp. 2224–2287, 2019.
- [7] Chen, M. et al., “Artificial neural networks-based machine learning for wireless networks: A tutorial,” *IEEE Comm. Surveys & Tutorials*, vol. 21, no. 4, pp. 3039–3071, 2019.
- [8] Liang, L. and Ye, H. and Li, G. Y., “Toward intelligent vehicular networks: A machine learning framework,” *IEEE IoT Journal*, vol. 6, no. 1, pp. 124–135, 2018.
- [9] O’Shea, T. and Hoydis, J., “An introduction to deep learning for the physical layer,” *IEEE ToCCN*, vol. 3, no. 4, pp. 563–575, 2017.
- [10] Samek, W., *Explainable AI: interpreting, explaining and visualizing deep learning*. Springer Nature, 2019, vol. 11700.
- [11] García, J. and Fernández, F., “A comprehensive survey on safe reinforcement learning,” *JMLR*, vol. 16, no. 1, pp. 1437–1480, 2015.
- [12] Wilhelmi, F. et al., “A flexible machine-learning-aware architecture for future wlans,” *IEEE Comm. Mag.*, vol. 58, no. 3, pp. 25–31, 2020.
- [13] P. Gawłowicz and A. Zubow, “ns-3 meets OpenAI Gym: The Playground for Machine Learning in Networking Research,” in *Proceedings of the 22nd International ACM Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems*, 2019, pp. 113–120.
- [14] Cavin, D., Sasson, Y., and Schiper, André, “On the accuracy of manet simulators,” in *Proceedings of the second ACM international workshop on Principles of mobile computing*, 2002, pp. 38–43.
- [15] Wilhelmi, F. et al., “Potential and pitfalls of multi-armed bandits for decentralized spatial reuse in WLANs,” *JNCA*, vol. 127, pp. 26–42, 2019.