

Coordinated Spatial Reuse Scheduling With Machine Learning in IEEE 802.11 MAPC Networks

Maksymilian Wojnar, Wojciech Ciężobka, Artur Tomaszewski, Piotr Chołda, Krzysztof Rusek, Katarzyna Kosek-Szott, Jetmir Haxhibeqiri *Member, IEEE*, Jeroen Hoebeke, Boris Bellalta *Senior Member, IEEE*, Anatolij Zubow *Senior Member, IEEE*, Falko Dressler *Fellow, IEEE*, Szymon Szott

Abstract—The densification of Wi-Fi deployments means that fully distributed random channel access is no longer sufficient for high and predictable performance. Therefore, the upcoming IEEE 802.11bn amendment introduces multi-access point coordination (MAPC) methods. This paper addresses a variant of MAPC called coordinated spatial reuse (C-SR), where devices transmit simultaneously on the same channel, with the power adjusted to minimize interference. The C-SR scheduling problem is selecting which devices transmit concurrently and with what settings. We provide a theoretical upper bound model, optimized for either throughput or fairness, which finds the best possible transmission schedule using mixed-integer linear programming. Then, a practical, probing-based approach is proposed which uses multi-armed bandits (MABs), a type of reinforcement learning, to solve the C-SR scheduling problem. We validate both classical (flat) MAB and hierarchical MAB (H-MAB) schemes with simulations and in a testbed. Using H-MABs for C-SR improves aggregate throughput over legacy IEEE 802.11 (on average by 80% in random scenarios), without reducing the number of transmission opportunities per station. Finally, our framework is lightweight and ready for implementation in Wi-Fi devices.

Index Terms—IEEE 802.11, coordinated spatial reuse, reinforcement learning, channel access, Wi-Fi, multi-armed bandits, upper bound model, linear programming.

I. INTRODUCTION

FOR the past 25 years, channel access in IEEE 802.11 (Wi-Fi) has been based on fully distributed random access. With network densification, this ‘no-coordination’ approach no longer meets the requirements of recent applications (such as augmented reality). The first improvement in this area was the introduction of intra-basic service set (BSS) coordination with orthogonal frequency-division multiple access (OFDMA) in 802.11ax (Wi-Fi 6), where the access point (AP) coordinates both uplink (UL) and downlink (DL) transmissions. The next step is 802.11bn inter-BSS coordination, where neighboring APs coordinate their transmissions for interference management.

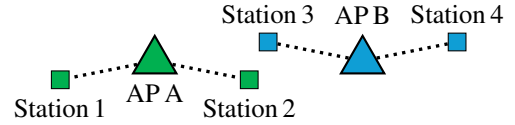
M. Wojnar (corresponding author), W. Ciężobka, P. Chołda, K. Rusek, K. Kosek-Szott, and S. Szott are with the AGH University of Kraków, Poland, e-mail: {name.surname}@agh.edu.pl.

A. Tomaszewski is with the Gdańsk University of Technology, Poland, e-mail: artur.tomaszewski@pg.edu.pl.

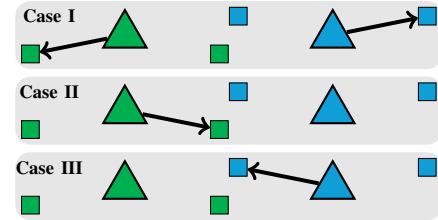
J. Haxhibeqiri and J. Hoebeke are with the University of Ghent, Belgium, e-mail: {Jetmir.Haxhibeqiri, Jeroen.Hoebeke}@UGent.be.

B. Bellalta is with the Department of Engineering, UPF Barcelona, Spain, e-mail: boris.bellalta@upf.edu.

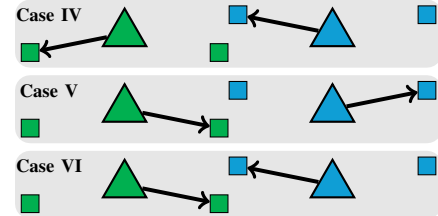
A. Zubow and F. Dressler are with the School of Electrical Engineering and Computer Science, TU Berlin, 10587 Berlin, Germany, e-mail: {zubow, dressler}@tkn.tu-berlin.de.



(a) Topology and AP-station associations.



(b) Transmission configurations that are viable.



(c) Transmission configurations leading to collisions.

Fig. 1. Example of the downlink scheduling problem in C-SR. For clarity, we omit varying the transmit power. Stations are associated to nearest APs (a). To ensure successful frame reception, APs coordinate to either transmit simultaneously to outer stations or unilaterally to inner stations (b). Without coordination, if both APs transmit simultaneously and one transmits a frame to an inner station, that frame will collide (c).

Coordination among neighboring APs (called overlapping BSS, OBSS) improves network performance in terms of higher reliability, reduced latency, increased throughput, and lower power consumption [1]. A good example of network settings where such improvements can be observed are factories and shopping malls, where the communication ranges of APs overlap. In particular, cell edge users, who suffer from interference from neighboring APs, benefit from inter-BSS coordination. Therefore, a multi-AP coordination (MAPC) feature is planned for 2028 with the release of 802.11bn [2].

The IEEE 802.11bn Task Group lists several modes of MAPC operation in its Specification Framework Document [3]: coordinated spatial reuse (C-SR), coordinated beamforming (C-BF), coordinated time division multiple access, and coordinated restricted target wake time. We focus on the first one, formally defined as “multi-AP Coordinated Spatial Reuse

(C-SR) at transmission opportunity (TXOP)-level with power control”. This definition means that Wi-Fi devices (APs or stations) transmit simultaneously, on the same channel, with adjusted transmit power to avoid interference. C-SR is not only pre-approved by the TGBn Task Group, it will also most likely be easy to implement as there are no requirements at the PHY layer. Theoretically, existing devices would require only driver updates to parse dedicated frames (for over-the-air synchronization), although the exact signaling still remains to be specified. Note that C-SR is not the same as (regular, uncoordinated) spatial reuse (SR), a feature introduced in 802.11ax. We explain the differences between SR and C-SR in Section III.

Consider a C-SR example with two APs and two associated stations per AP (Fig. 1). In this example and in the remainder of the paper, we focus on DL transmissions (since most traffic is DL, as per the Ericsson Mobility Report of November 2024). The key feature in this example are the two cell-edge stations (i.e., stations 2 and 3), which are in the interference range of both APs. If both APs are coordinated, their transmissions can be successfully decoded at the receiving stations. There are three transmission configurations that are viable: simultaneous transmission to the outer stations (case I) or a unilateral transmission (i.e., with one AP refraining from transmission) to one of the inner stations (cases II and III). Meanwhile, if both APs transmit simultaneously and one of the inner (cell-edge) stations is the intended recipient, that transmission will fail due to interference from the other AP (cases IV to VI). Given that the clear channel assessment (CCA) mechanism on both APs will not detect ongoing transmissions (as they are out of sensing range of each other), without coordination, all six cases may occur indiscriminately. Even enabling frame protection (i.e., request-to-send/clear-to-send, RTS/CTS, frames) is not a solution because APs will still not transmit simultaneously even if it is viable (e.g., case I in Fig. 1 cannot be attained with RTS/CTS). Thus, the performance of IEEE 802.11 in such topologies is greatly improved with coordinated methods such as C-SR, provided that we determine which transmission configurations are viable.

We refer to the challenge of selecting transmission configurations and transmit power settings in C-SR (i.e., distinguishing cases I to III from cases IV to VI in Fig. 1), as the *C-SR scheduling problem*. Our first contribution is a theoretical upper bound model of C-SR (Section IV), which finds the best possible transmission schedule using mixed-integer linear programming (MILP). The model can be optimized for either fairness or throughput and serves as a baseline for comparing practical C-SR scheduling solutions.

C-SR scheduling solutions typically resort to channel measurements (Section II). These measurements can be passive, e.g., based on the received signal strength (RSS) of overheard frames, or active, e.g., acquiring channel state information (CSI) from an additional channel sounding frame exchange. Then, on the basis of these measurements, a decision is made regarding which transmission configurations are viable. Unfortunately, the aforementioned measurements can be unreliable (RSS), create signaling overhead (both RSS and CSI), require inter-OBSS frame exchange (CSI), and even be unusable

and not scalable for C-SR (RSS). Consider the adjoining research area of data rate selection in Wi-Fi networks, where multiple measurement-based methods have been proposed [4]. These methods have not been adopted due to the previously mentioned issues with RSS, and today wireless drivers typically use probing-based approaches [5]. Inspired by such developments, we show how to solve the C-SR scheduling problem using a novel probing approach.

We turn to machine learning (ML) and use multi-armed bandit (MAB) methods [6], to find which transmission configurations are viable in a C-SR setting. The main benefit of MABs is that they are reinforcement learning (RL) methods which do not require *a priori* knowledge (such as channel measurements) and can learn online, which is suitable for the random and nomadic nature of Wi-Fi deployments. We aim to answer the following research questions:

- 1) How quickly can MABs learn which C-SR transmission configurations to use?
- 2) How close to optimal are the transmission configurations found?

These research questions aim to address the applicability of using MAB methods in Wi-Fi networks. We contribute to answering these questions by using two types of MABs: a classical one (flat MAB) and a hierarchical one (H-MAB), both tailored to the C-SR scheduling problem. Then, we evaluate both MAB C-SR frameworks with extensive simulations in static and dynamic topologies and in an experimental testbed. For comparison, we use both legacy IEEE 802.11 channel access (the distributed coordination function, DCF) and 802.11ax SR as lower bounds, as well as the provided upper bound model of C-SR. To encourage research in this area, all of our contributions are available as open-source¹.

We show that the proposed probing-based approach of MABs can be successfully applied to solving the C-SR scheduling problem. We show that MAB-based C-SR improves network performance over existing channel access schemes (DCF, SR) achieving high throughput while preserving the minimum number of TXOPs per each station. We also validate the proposed framework in an experimental setting, where its adaptability to the underlying network conditions is shown. Finally, we emphasize that MABs are lightweight and simple to implement and, in contrast to deep RL solutions (which require dedicated processing units), do not have any requirements that prevent them from being implemented in existing devices.

In summary, our key contributions are as follows:

- 1) An optimization model of C-SR for determining upper bound performance (defined as maximizing either throughput or fairness).
- 2) The application of a MAB probing approach to the C-SR scheduling problem coupled with a hierarchical design to tackle scalability issues².

¹<https://github.com/ml4wifi-devs/csr>

²We first applied MABs to C-SR scheduling in [7]. Here, we extend the hierarchical MABs with a third level (for transmission power selection).

- 3) An extensive simulation analysis to study the performance of the proposed approach in comparison to existing baselines.
- 4) An experimental testbed study to verify the operation of the proposed approach in real-world devices.

The remainder of this paper is organized as follows. We review the state of the art and motivate our approach in Section II. Then, we explain in detail how C-SR works, what signaling requirements it has, and how it differs from SR (Section III). Next, in Section IV, we provide a new optimization model of C-SR, which determines either the upper-bound throughput or the upper-bound fairness. Appendix A contains a scalability analysis of this model. We introduce our MAB C-SR framework in Section V, which we then evaluate in simulations (Section VI) and validate with experiments (Section VII). Finally, we discuss our findings and outline future work in Section VIII.

II. RELATED WORK

The need to coordinate transmissions in wireless networks with multiple base stations is a problem also known in the cellular domain. Hence, 3GPP proposed coordinated multi-point transmission (CoMP) in LTE Rel. 11 [8]. However, CoMP and MAPC are significantly different: CoMP operates on larger areas with a focus on cell edge users, using a different PHY/MAC stack, and being limited to strictly centralized operation. Therefore, we turn our focus to solutions for wireless local area networks.

Several articles on the use of MAPC in Wi-Fi networks have already been published, as surveyed by Verma et al. [9]. They provide an overview of MAPC variants (e.g., C-OFDMA, C-SR, C-BF) and comment on future MAPC research directions (e.g., multi-band MAPC, proactive MAPC sounding, blockage-aware coordination, and MAPC scalability).

In terms of solutions to the C-SR scheduling problem, Nunez et al. [10] estimate the expected rate of each link and select the top results, making sure that senders and receivers are selected only once. They later extend their approach into a two-stage algorithm comprising compatible group formation and AP selection (based on various fairness criteria, e.g., most packets to send) [11]. Meanwhile, Imputato et al. [1] propose an alternative C-SR scheduling algorithm which relies on collecting RSSI information and estimating if the resulting signal to interference plus noise ratio (SINR) of coordinated transmissions allows transmitting data with a given modulation and coding scheme (MCS). Some works focus only on the power allocation problem in C-SR. Wang and Fang [12] address this with a federated learning double deep Q-network (FL-DDQN) algorithm. APs use a local DDQN for interacting with the environment and then update a global model, considering both throughput and latency requirements. Meanwhile, Talukder and Xie [13] adjust the transmit power levels by defining an appropriate interference tolerance limit using RSSI levels as input.

While most work on C-SR scheduling is simulation-based, Haxhibeqiri et al. [14] use a real-world testbed to validate their findings. Using RSSI reports and historical MCS data,

a central controller chooses C-SR groups for downlink transmissions while adjusting the MCS and transmit power values of APs. This research is extended in [15], where frame queue organization is additionally considered and AP coverage is divided into circular regions to reduce the number of potential concurrent receivers. We use the same testbed in Section VII (but with our novel, ML-based C-SR scheduling solution).

The existing works mostly use measurements (typically of SINR) as input for the C-SR scheduling problem. The most common scenario in which the proposed algorithms are tested is simplistic and assumes static topologies composed of four APs with a low number of associated stations. Typically, state-of-the-art works rely on heuristics and network simulations and focus on downlink transmissions. Furthermore, although Wilhelmi et al. [16] provide an analytic throughput model based on Markov chains, it is limited to two OBSSs. Thus, we conclude that ML-driven probing-based C-SR scheduling has not yet been tested in complex scenarios, and there are no models which estimate the upper bound C-SR performance under any given settings.

III. C-SR OPERATION

In this section, we first present the differences between centralized and distributed C-SR and explain why we focus on the former. Then, we describe how transmissions are organized in a C-SR frame exchange and show the impact of using channel measurements. We also describe the differences with respect to SR. Since C-SR has not yet been standardized and TGBn is currently working on relevant signaling aspects [17], some details of our description below may change by the final release of the 802.11bn amendment.

A. Initialization and Centralized vs Distributed Approach

A C-SR frame exchange is initialized whenever a C-SR-capable AP wins channel access according to legacy rules (DCF). This AP becomes the *sharing AP* and participates in the coordinated transmission in the upcoming TXOP. The other participants in this TXOP (the *shared APs*) are selected in centralized or distributed way.

In the centralized approach, coordination is done in a central network controller, which has a wired connection to the APs. This controller obtains and processes monitoring data from the network, decides which transmission configurations are viable, and selects an order of configurations which maximizes performance (we explain the particulars in Section IV and the required signaling in Fig. 4). This centralized approach requires that the APs are under a single administrative domain, which is reasonable in a typical C-SR setting (e.g., a factory, shopping mall, or university campus).

The alternative distributed variant assumes that the sharing AP fully controls channel access for the TXOP and invites shared APs to transmit concurrently. To execute an informed decision, the sharing AP must exchange data with neighboring APs over a wired link (if the APs are under the same administration domain) or over a wireless link (otherwise). The former case is conceptually similar to the centralized approach while the latter requires that the APs are within their

transmission ranges (e.g., if APs A and B in Fig. 1 cannot receive each other's control frames and the APs do not share a wired connection, then distributed coordination is impossible). In the following, we focus on centralized C-SR to provide a proof-of-concept of our proposed MAB C-SR framework and leave the fully distributed case, as the more demanding one, for future work due to the lack of space.

B. Signaling Issues with Channel Measurements

C-SR scheduling can theoretically be based on RSS or CSI channel measurements. The former requires two steps: (i) selecting which received frames to measure and (ii) signaling these measurements to the transmitter. Consider the example in Fig. 1, assuming APs A and B transmit only to the inner stations 2 and 3, respectively. RSS measurements need to be performed at each station, but separately for both APs. This is easy because both APs periodically transmit beacon frames. However, the next step requires additional UL transmissions to transfer the measurements from the stations to their respective APs. This signaling overhead, combined with the high variance of RSS [18], which can lead to erroneous decisions, both reduce the applicability of such measurements in C-SR.

The CSI case, while providing more reliable measurements, requires even more signaling overhead and additional processing power. To obtain CSI, dedicated channel sounding frame exchanges is required over-the-air (not only intra- but also inter-OBSS) to determine the channel quality between each pair of devices. While researchers investigate channel sounding schemes dedicated to C-SR [9], we abstain from resorting to any channel measurements (based on RSS or CSI) and adopt a new probing approach without additional inter-OBSS over-the-air signaling.

C. Transmission Organization

We assume a fully coordinated network (i.e., with only C-SR transmissions). TXOPs are initialized by the sharing AP that wins channel access according to DCF rules. Then, the network controller instructs other APs about the possibility of joining this coordinated TXOP and, if so, their recipient station and transmit power settings. C-SR transmissions can start upon reception of this message. In practice, networks will operate in a mixed mode where not all TXOPs are coordinated, but the goal of this paper is to focus solely on performance during coordinated TXOPs and leave mixed scenarios as future work.

All our assumptions and design choices (DL transmissions, central coordination, no channel measurements) simplify the structure of a C-SR TXOP, which comprises a set of parallel DL transmissions of IEEE 802.11 DATA frames, followed by a set of parallel UL transmissions of IEEE 802.11 acknowledgment (ACK) frames. These ACKs are simultaneously transmitted using OFDMA as multi-user block ACKs, which ensures that they do not collide.

D. Comparison with SR

C-SR can be confused with the SR feature of 802.11ax. In SR, a device receiving a frame from a neighboring network can

disregard it (by applying a more permissive CCA threshold) and consider the channel to be idle, provided that it performs a transmission at a reduced power level to avoid interference with the overlapping transmission. This leads to two key operational differences: in SR only the neighboring AP reduces its transmit power (while in C-SR any or all APs may reduce this power) and the transmissions are asynchronous. The outcome of this difference is that SR does not help cell-edge users such as the inner stations in Fig. 1 and therefore is not widely used. C-SR is expected to deliver higher performance gains and is the focus of our study.

IV. UPPER BOUND MODEL OF C-SR

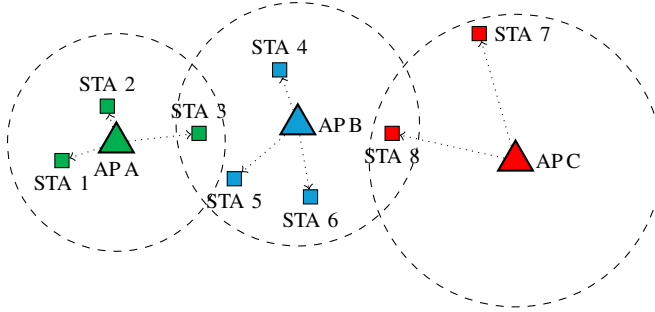
To analytically solve the C-SR scheduling problem, i.e., determine the upper-bound transmission schedule under the assumption of a full-buffer traffic model (i.e., each AP always has data to send), we develop a MILP optimization model. This model is flexible: we show that it can maximize aggregate network throughput (which we call T-Optimal) or fairness (which we call F-Optimal). Since our goal is to calculate the upper bound performance (in terms of throughput or fairness), as stated before, we do not consider APs operating in mixed-mode (with some TXOPs dedicated to non-C-SR transmissions). Furthermore, as in the remainder of the paper, we assume omnidirectional single spatial stream antennas.

Our modeling approach is derived from [19, 20]. The authors propose therein a methodology for wireless mesh networks which decouples the design of non-interfering AP-station transmissions from the design of the transmission schedule, i.e., the allocation of time periods during which various sets of non-interfering transmissions (called *transmission sets*) can occur. This methodology incorporates an optimization method known as *column generation* [21]. We adapt these concepts to MAPC C-SR as follows.

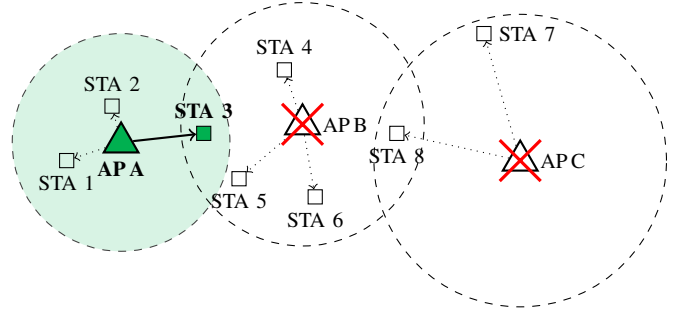
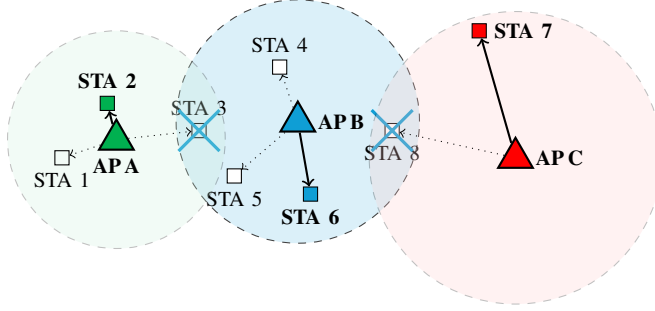
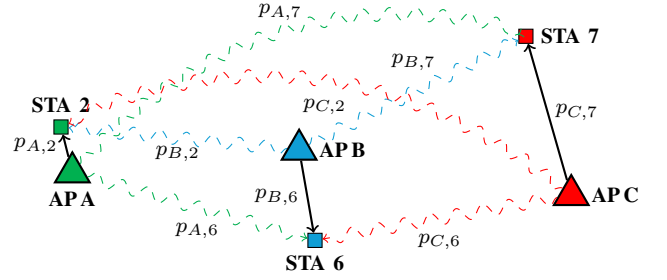
We model the network using a directed bipartite graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. The vertex set \mathcal{V} represents nodes and is partitioned into two subsets: \mathcal{A} of APs and \mathcal{S} of stations ($\mathcal{V} = \mathcal{A} \cup \mathcal{S}$ and $\mathcal{A} \cap \mathcal{S} = \emptyset$). The arc set \mathcal{E} denotes potential transmission links between APs and stations ($\mathcal{E} \subseteq \mathcal{A} \times \mathcal{S}$). For each link $e \in \mathcal{E}$, we denote its originating node as $a(e) \in \mathcal{A}$, and the terminating node as $s(e) \in \mathcal{S}$, i.e., $e = (a(e), s(e))$. We also introduce auxiliary sets of node-adjacent links: for each node $v \in \mathcal{V}$, $\delta^+(v) \equiv \{e \in \mathcal{E} : a(e) = v\}$ is the set of links originating at v , and $\delta^-(v) \equiv \{e \in \mathcal{E} : s(e) = v\}$ is the set of links terminating at v .

Given the complex interplay of interference at receiving stations, our optimization process addresses two critical issues:

- 1) *Determining transmission sets and C-SR settings:* First, select APs that can transmit simultaneously, their recipient stations, as well as transmit power levels and MCS values to achieve successful transmissions under the resulting SINR. Then, return the family \mathcal{T} of transmission sets t : $\mathcal{T} = \{t_1, t_2, \dots\}$, where each transmission set t is a vector of throughput values (rates) achieved at the receiving stations.
- 2) *Allocating transmission sets:* Given \mathcal{T} and the available time slots, determine the number of slots to assign to each



(a) APs and their associated stations.

(b) Exemplary transmission set t_1 , where only AP A transmits and the crossed-out APs are switched off (although due to lack of interference, APC could also transmit).(c) Exemplary transmission set t_2 , where all APs are able to transmit. The crossed-out stations could not be served due to APB transmitting to STA 6.

(d) Possible interference that can decrease the transmission rates.

Fig. 2. Transmission sets in an exemplary topology. Dashed circles illustrate the potential signal reach for a successful transmission at a given transmit power (the interference range may be larger). The arrows present a potential downlink connection. The vector of throughput obtainable for transmission set t is given as $[r(t, s_1), r(t, s_2), \dots, r(t, s_8)]$, where $r(t, s_i)$ represents the throughput obtained by station i .

transmission set (i.e., how many slots are designated for transmission sets t_1, t_2 , etc.).

Identifying a transmission set is generally a complex task. While various transmission configurations are theoretically possible, only some are feasible, primarily due to the unacceptable interference from transmissions between different AP–station pairs. Since each station is associated with one AP, we identify a link by its receiver s . Moreover, from an optimization perspective, not all transmission sets are considered advantageous. This concept is illustrated with examples in Fig. 2. For example, in Fig. 2b the exemplary transmission set t_1 depends on a single transmission from APA to STA 3. In this scenario, APB cannot transmit to any of its stations due to potential harmful interference with STA 3's reception (STA 3 is in the signal range of APB). Additionally, although APC could transmit, it does not, so that there are no interfering signals for the AP A \rightarrow STA 3 transmission. Consequently, the throughput values achieved by stations in relation to this transmission set are represented as $[0, 0, r(t_1, s_3), 0, 0, 0, 0, 0]$, where the bitrate $r(t_1, s_3)$ is determined by SINR at STA 3. Conversely, another transmission set t_2 allows three simultaneous transmissions (see Fig. 2c), with station downlink throughput values $[0, r(t_2, s_2), 0, 0, 0, r(t_2, s_6), r(t_2, s_7), 0]$, where $r(t_2, s_2)$, $r(t_2, s_6)$ and $r(t_2, s_7)$ are derived from the highest feasible MCS achievable during simultaneous transmissions APA \rightarrow STA 2, APB \rightarrow STA 6, and APC \rightarrow STA 7, respectively. However, feasible MCS values should be found based on SINR including interference as shown in Fig. 2d,

which in some cases (e.g., $p_{C,2}$) may be negligible. We need to find them in the optimization model as well, which makes the task complex.

We break down the overall problem into two sub-problems. The first, *main problem* assumes that transmission sets are predefined (a family of transmission sets is given), focusing on their allocation across different time slots. By reducing the duration of time slots to short intervals, we transition from dealing with discrete time slots to working within a continuous time frame. This transition substantially eases the complexity of the problem. The second, *pricing problem* involves identifying only a single feasible transmission set. If such a new set is found, it extends the family of transmission sets to feed the main problem. The whole procedure is iterated until no new transmission sets are found (Fig. 3). A scalability analysis of the model is in Appendix A.

Below, we outline how each of these problems is formulated when maximizing *fair* access to the network (F-Optimal), i.e., when we maximize the throughput achieved by the least served station.

1) *Main Problem*: We operate under the assumption that a normalized time period $[0, 1]$ is divided into continuous intervals. In a single interval, exactly one transmission set is executed. For a given family of transmission sets, the objective of the main problem is to identify the optimal mix of time intervals that adheres to a basic fairness criterion. It aims to maximize the achievable received throughput of the station that experiences the lowest performance. For each station $s \in S$,

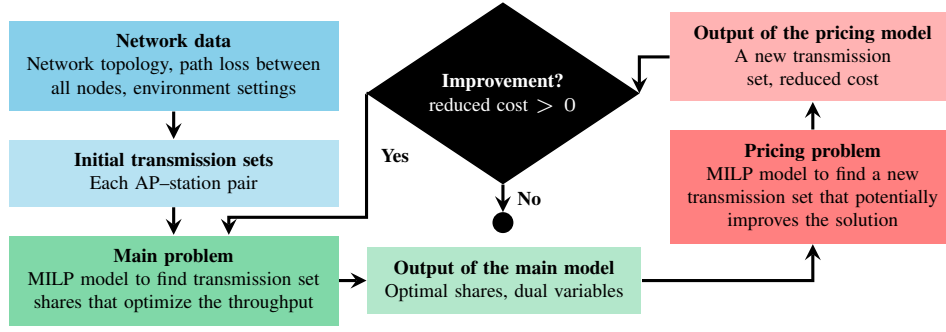


Fig. 3. Relations between the main and pricing problems.

let R_s denote the throughput achieved by station s during the normalized time period, which we seek to optimize through the transmission schedule. Consequently, the problem of transmission scheduling could be formulated as $\max \min_{s \in \mathcal{S}} R_s$. By setting a lower limit on the obtainable throughput $\underline{R} \geq 0$, we can reformulate the latter as a maximization problem: $\max\{\underline{R} : \underline{R} \leq R_s, s \in \mathcal{S}\}$, effectively seeking the highest possible minimum throughput that can be applied across all stations.

Let set \mathcal{T} denote the indices of a given family of transmission sets that have been identified so far. Initially, this family might be composed solely of all single-link transmission sets (Fig. 2b) that can transmit without co-channel interference. For each $t \in \mathcal{T}$, let the variable w_t specify the proportion of time (or the duration of a continuous interval) during which the transmission set t is activated. Then, assuming that for each $t \in \mathcal{T}$, and for every station $s \in \mathcal{S}$, $r(t, s)$ represents the throughput averaged over all slots for station s when transmission set t is activated, the cumulative throughput R_s for station s (accounting for the contributions from the shares of various transmission sets w_t) is obtained according to constraints (1c). Recall that the variable \underline{R} denotes the minimum throughput achievable by each station. Then, we formulate the main problem as the following simple *linear programming* (LP) model:

$$\max \underline{R} \quad (1a)$$

$$\text{s.t. } \sum_{t \in \mathcal{T}} w_t = 1 \quad | \quad \alpha \quad (1b)$$

$$R_s = \sum_{t \in \mathcal{T}} r(t, s) w_t \quad s \in \mathcal{S} \quad | \quad \beta_s \quad (1c)$$

$$\underline{R} \leq R_s \quad s \in \mathcal{S} \quad (1d)$$

From optimization theory, we know that the throughput value of the worst performing station can be potentially improved by expanding the set \mathcal{T} to include a transmission set for which a so-called *reduced cost* $[-\alpha + \sum_{s \in \mathcal{S}} r(t, s) \beta_s]$ is strictly positive [22, 21]. Here, α and β_s , $s \in \mathcal{S}$, represent the optimal values of the dual variables associated with problem (1). These values are accessible once an optimal solution to the main problem has been achieved.

2) *Pricing Problem*: The *pricing problem* generates (prices out) new variables for the restricted main LP problem, improving main's optimum (from the previous iteration). This

idea lies at the heart of utilizing the mentioned column generation technique, which facilitates the division of the entire optimization strategy into the two sub-problems. By adopting this method, we can discover an optimal transmission schedule through a cyclic process of solving the main-pricing problem pair. In each cycle, we introduce a new promising transmission set to the family \mathcal{T} , continuing this iterative process until the reduced cost obtained from the solution of the pricing problem reaches 0. The formulation of the pricing problem, shown below, considers potential interference while various APs transmit. Then, we are able to generate a new transmission set.

For every AP $a \in \mathcal{A}$ and station $s \in \mathcal{S}$, let $l(a, s)$ denote the loss on the transmission path from AP a to station s . Let \mathcal{M} be the set of available MCSs. For each mode $m \in \mathcal{M}$, let $r_{\mathcal{M}}(m)$ and $i(m)$ denote the transmission rate and the minimum SINR required by mode m , respectively. Let \underline{P} and \overline{P} denote the minimum and maximum transmit power of an active AP, respectively. N is used to denote the noise floor.

For each AP $a \in \mathcal{A}$, we define a binary variable x_a that is 1 if and only if (iff) AP a is selected for transmission. Similarly, for each link $e \in \mathcal{E}$, a binary variable y_e is 1 iff link e is activated for transmission. Additionally, for every $a \in \mathcal{A}$, a non-negative continuous variable p_a expresses the transmit power of AP a . For all links $e \in \mathcal{E}$ and for each mode $m \in \mathcal{M}$, a binary variable u_{em} is 1 iff transmission mode m is selected for use on link e . Furthermore, for each station $s \in \mathcal{S}$, a non-negative continuous variable r_s expresses the downlink throughput of this station. Its optimal value r_s^* will be the throughput for the new transmission set from solving the pricing problem.

To properly calculate the transmit power, we need to define the SINR condition. To illustrate the related notions, we refer to Fig. 2d, and focus on link e representing transmission APB \rightarrow STA 6. Then, we have $a(e) = \text{APB}$, $p_{a(e)} = p_{B,6}$, and the set of other APs: $c \in \mathcal{A} \setminus \{a(e)\} = \{\text{APA}, \text{APB}\}$. We use only the linear relationships related to the signal power levels as well as the SINR requirements. Therefore, we have to use constants $l(x, y)$ related to attenuation levels at y when x is transmitting. Then, on the mentioned link e , the useful signal level at station $s(e)$, when it receives a signal from AP $a(e)$, is $\frac{p_{a(e)}}{l(a(e), s(e))} = \frac{p_{B,6}}{l(B,6)}$. Continuing our example related to Fig. 2d, we should also consider interference from transmissions APA \rightarrow STA 2 and APC \rightarrow STA 7, since in this

case $c \in \{\text{APA}, \text{APB}\}$. In general, the level of interference and noise is

$$\zeta_e = \sum_{c \in \mathcal{A} \setminus \{a(e)\}} \frac{p_c}{l(c, s(e))} + N. \quad (2)$$

This equation is repeated below in the full problem formulation within constraints (4i). Finally, we state that if the following inequality formulating the SINR condition holds, then we can use mode m in link e :

$$\frac{p_{a(e)}}{l(a(e), s(e))} \frac{1}{\zeta_e} \geq i(m). \quad (3)$$

The formulation of the pricing problem, then, is a complex MILP model that incorporates these variables to address the challenge of determining the most efficient transmission configuration under the constraints of the network's physical characteristics and capacities. Simultaneously, this pricing problem considers the dual values from the main problem, as they suggest which transmissions are advantageous. Therefore, this problem is formulated as follows:

$$\max -\alpha + \sum_{s \in \mathcal{S}} \beta_s r_s \quad (4a)$$

$$\text{s.t. } x_a = \sum_{e \in \delta^+(a)} y_e \quad a \in \mathcal{A} \quad (4b)$$

$$\sum_{e \in \delta^-(s)} y_e \leq 1 \quad s \in \mathcal{S} \quad (4c)$$

$$p_a \geq \underline{P} x_a \quad a \in \mathcal{A} \quad (4d)$$

$$p_a \leq \bar{P} x_a \quad a \in \mathcal{A} \quad (4e)$$

$$\sum_{m \in \mathcal{M}} u_{em} = y_e \quad e \in \mathcal{E} \quad (4f)$$

$$r_s = \sum_{e \in \delta^-(s)} \sum_{m \in \mathcal{M}} r_{\mathcal{M}}(m) u_{em} \quad s \in \mathcal{S} \quad (4g)$$

$$p_{a(e)} + M(e, m)(1 - u_{em}) \geq \quad (4h)$$

$$l(a(e), s(e)) i(m) \zeta_e \quad e \in \mathcal{E}, m \in \mathcal{M}$$

$$\zeta_e = \sum_{c \in \mathcal{A} \setminus \{a(e)\}} \frac{p_c}{l(c, s(e))} + N \quad e \in \mathcal{E} \quad (4i)$$

Constraints (4b) make the active AP transmit on exactly one link, and constraints (4c) forbid the station to receive on more than one link. Constraints (4d) and (4e) enforce the power of the transmitting AP to be in the range $[\underline{P}, \bar{P}]$. Constraints (4f) require that exactly one MCS is used on the transmitting link. Constraints (4g) determine the throughput of the station. Finally, constraints (4h) are the SINR constraints that must be satisfied if a particular MCS mode is used on the link (and thus $u_{em} = 1$). The specific form of constraints (4h) improves the numerical properties of the problem by formulating the constraint separately for each MCS. The constraint is deactivated whenever a specific transmission mode m is not used on link e (and thus $u_{em} = 0$). In this case, $M(e, m)$ on the left-hand-side makes the inequality trivially true; for all $e \in \mathcal{E}$, $m \in \mathcal{M}$, $M(e, m)$ is a constant equal to the upper bound value of the constraint's right-hand-side expression, i.e., its value when $p_c \equiv \bar{P}$, $c \in \mathcal{A} \setminus \{a(e)\}$. Note, that due to constraints (4f) only one of u_{em} for a transmitting link can be

1. Then, $M(e, m)$ is annulled, and the constraint is, in fact, the linearization of the SINR condition given by (3).

If the optimal objective function value of problem (4) is strictly positive, the optimal values r_s^* of variables r_s , $s \in \mathcal{S}$, define a new transmission set t' , which we add to the family \mathcal{T} to resolve the restricted main problem (1) (Fig. 3).

Alternatively, if we decide to optimize for aggregate throughput, then we need to introduce only minor changes. In the main problem, instead of the goal function provided by (1a) we use $\max \sum_{s \in \mathcal{S}} R_s$ and constraints (1d) become unnecessary. Additionally, in the pricing problem the goal function given by (4a) becomes $\max -\alpha + \sum_{s \in \mathcal{S}} (1 + \beta_s) r_s$.

V. MABS FOR C-SR

A classical MAB agent learns the reward distribution of each arm by exploring individual choices. In the C-SR scheduling problem, each arm is a given transmission configuration (senders, receivers, and transmit power levels) and the reward is the normalized aggregate network throughput. MAB algorithms implement a trade-off between exploration (knowledge acquisition, i.e., trying new arms) and exploitation (knowledge application, i.e., using arms giving the highest reward). Since a MAB agent can only make decisions a finite number of times, it should discover the action with the highest reward as quickly as possible. For these classical MABs, their implementation for C-SR is quite straight forward. First, we randomly uniformly select the sharing AP (which models 802.11 channel access, where each AP has a fair chance of winning the contention) and the recipient station (the destination of the head-of-line frame in that AP's transmission buffer). Then, we enumerate all available choices based on the association between APs and stations, as well as the available transmit power levels, i.e., each configuration is an independent action. Unfortunately, classical MABs are well suited only for simple topologies (as we will later show) and may fail to find a satisfactory configuration in a reasonable time, since exploration is time-consuming for systems with a large number of arms, as in the case of C-SR in dense topologies (composed of many APs and stations). Therefore, a single action should provide information about the outcomes of similar actions to improve MAB convergence time. While in the general case this could be difficult, we exploit the structure of the underlying C-SR system to construct a hierarchy of actions, which is the key design principle of H-MABs.

To explain this hierarchy, consider a toy example of scheduling pairwise communication between n entities. An agent selects who sends to whom and observes the reward of its decision. The action space consists of (sender, receiver) pairs. There are $n(n-1)/2$ such pairs and we call this a flat action space, and the agent selecting such pairs – a flat agent. Here, the action space is two-dimensional, so a single action is a sequence of two simpler sub-actions: selecting the sender and then the receiver. This decomposition is the key idea of H-MABs: agents are introduced at each level in the hierarchy where a decision is made. In our toy example, the first-level agent selects the sender and its action space has n choices. For each sender, we introduce another agent that selects the

receiver. This agent has an action space of $n - 1$ choices, but there are n such agents, so no action is lost, we only have a different representation. The key benefit of this new representation is that both agent types have smaller action spaces, while the reward propagates to similar actions. In particular, every reward obtained by the downstream agent propagates to the top agent and indirectly affects all actions corresponding to the same sender. For example, if there is a high reward for allowing particular entities to send, such an agent quickly finds these entities, while the flat agent needs to explore all possible pairs. Therefore, we design an H-MAB algorithm for C-SR that builds on this idea.

Our H-MAB C-SR framework³ (Algorithm 1) considers the network as a set of APs \mathcal{A} with associated stations. This assignment partitions the set of all stations into subsets of stations $S_i = \{S_i^1, S_i^2, \dots, S_i^m\}$ associated with each AP A_i . In addition, APs can transmit data at any power level from the set \mathcal{P} . The outcome of the algorithm is a C-SR transmission configuration, comprising the AP-station pairs \mathcal{C}^* and the transmit power levels \mathcal{P}^* . Following the approach described above, we use three steps: selecting the transmitting APs, selecting the recipient stations, and finally selecting the transmit power to use. These steps form a three-level hierarchy of agents (Fig. 4)⁴. An agent's action is denoted as a sampling operation in Algorithm 1 because the agent is probabilistic and the action is sampled from a distribution.

The algorithm operates on consecutive C-SR TXOPs and begins by randomly uniformly selecting the initial (A_k, S_k^l) pair (lines 2-3). Selecting A_k models 802.11 channel access, while selecting S_k^l – the head-of-line frame in that AP's buffer. The next decision is whether to add any parallel C-SR transmissions to this pair.

The first-level agent α^I (whose parameters depend on the initially selected (A_k, S_k^l) pair) selects the subset of other transmitting APs from all possible subsets $2^{\mathcal{A} \setminus \{A_k\}}$ (line 6). For all APs that transmit simultaneously \mathcal{F} , except the initial one A_k , a receiver is selected (line 9). This is the second-level action and it depends on the previous-level action within the constraints posed by the associations. Thus, we have the explicit parameters of the second-level agent $\alpha^{II} [A_i, \mathcal{F}]$. Finally, for all C-SR pairs, we select the transmit power level. This action is taken by a third-level agent $\alpha^{III} [S_i^j, \mathcal{F}]$. The described procedure generates a transmission configuration, which is then executed by performing parallel transmissions according to \mathcal{C}^* with transmit powers set to \mathcal{P}^* and results in an effective data rate r (bytes received across all parallel transmissions during a TXOP).

Since the goal of the proposed H-MAB algorithm for

³An early version of this framework, with only two levels of agents, appeared in [7].

⁴We formally define our H-MAB framework as an independent, fully cooperative multi-agent RL problem. Although the independent learning framework may not be precise for some problems, such as transmit power selection, which formally requires joint action, it is often used in practice and yields satisfactory results [23]. Furthermore, this method significantly decreases the number of actions (from $|\mathcal{P}|^\nu$ to $\nu|\mathcal{P}|$, where \mathcal{P} is the set of available transmit power levels and ν is the number of transmissions), facilitating faster convergence and improving performance in real-world applications.

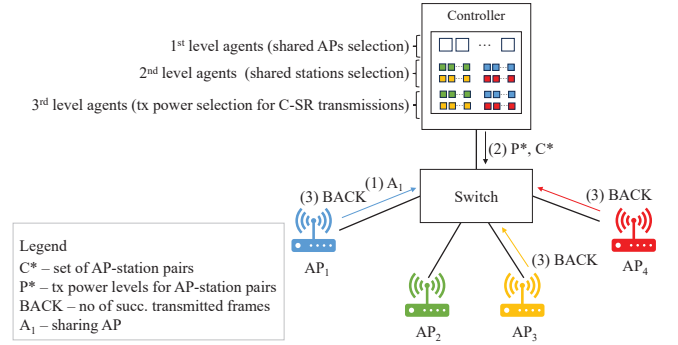


Fig. 4. Example of C-SR operation with H-MABs: (1) AP 1 notifies the controller that it is the sharing AP, (2) the controller provides the set \mathcal{C}^* of AP-station pairs to transmit and the transmit power levels \mathcal{P}^* of the transmitters in the next TXOP (which include APs 3 and 4), (3) after the TXOP, all transmitting APs inform the controller how many frames (bytes) were received successfully.

C-SR group selection is to maximize r , we use it as the reward and apply a MAB-dependent learning rule denoted as an update. This update is done in reverse order to the action selection: the third level agent is updated first, then the second level agent, and finally the first level agent. The reward expressed as the aggregate throughput of all transmissions propagates through the unchanged hierarchy. This enables the agent, especially its third level, to learn more efficiently by introducing information on the behavior of other agents. However, the higher the agent's level, the more noise (due to low-level mistakes) in the reward it observes, which is important from the learning algorithm point of view. For this reason, the third level should achieve faster convergence and undertake less exploration. Conversely, the first level should focus on increased exploration and ideally reach convergence subsequent to the third and second levels. In addition, we assume that MAB agents at each of the three levels are of the same type, although each level has its own configuration (hyperparameter set θ).

We tested multiple MAB types (ϵ -greedy, Softmax, Upper Confidence Bound – UCB, Thompson sampling – TS) [6] for flat MABs and H-MABs and fine-tuned their hyperparameters in random scenarios using Optuna⁵. From this comprehensive analysis, we selected the best MAB types for further analysis: Softmax for flat MABs and UCB for H-MABs.

As explained in Section III-A, we consider a centralized C-SR approach. Fig. 4 presents the H-MAB deployment (in case of a flat MAB deployment, only the logic in the controller should be changed). The signaling overhead is small and limited to data transferred over wired links. All computational costs are moved to the controller. Importantly, H-MABs can also be used in a fully distributed manner, with independent agents deployed at each of the APs (one per AP). However, this would require defining additional signaling between APs, which would introduce overhead. Therefore, the applicability of such a distributed solution is left for future work.

⁵<https://optuna.readthedocs.io/en/stable/>

Input:

$\mathcal{A} = \{A_1, A_2, \dots, A_N\}$ – set of APs
 $\mathcal{S}_i = \{S_i^1, S_i^2, \dots, S_i^m\}$ – stations associated with A_i
 $\mathcal{S} = \{S_1, S_2, \dots, S_N\}$ – structure of APs and stations
 $\mathcal{P} = \{P_1, P_2, \dots, P_W\}$ – available tx power levels
 $\theta_1, \theta_2, \theta_3$ – agents' hyperparameters

Initialize:

$\alpha^I \leftarrow$ the first-level agents based on \mathcal{A}, \mathcal{S} with θ_1
 $\alpha^{II} \leftarrow$ the second-level agents based on \mathcal{A}, \mathcal{S} with θ_2
 $\alpha^{III} \leftarrow$ the third-level agents based on $\mathcal{A}, \mathcal{S}, \mathcal{P}$ with θ_3

Algorithm:

```

1: for each TXOP do
2:   Randomly and uniformly select  $A_k \in \mathcal{A}$ 
3:   Randomly and uniformly select  $S_k^l \in \mathcal{S}_k$ 
4:    $\mathcal{C}^* \leftarrow \{(A_k, S_k^l)\}$ 
5:    $\mathcal{P}^* \leftarrow \{\}$ 
6:    $\mathcal{F}_k^l \in 2^{\mathcal{A} \setminus \{A_k\}} \leftarrow \text{sample } \alpha^I [S_k^l]$ 
7:    $\mathcal{F} \leftarrow \mathcal{F}_k^l \cup \{A_k\}$ 
8:   for  $A_i \in \mathcal{F}_k^l$  do
9:      $S_i^j \in \mathcal{S}_i \leftarrow \text{sample } \alpha^{II} [A_i, \mathcal{F}]$ 
10:     $\mathcal{C}^* \leftarrow \mathcal{C}^* \cup \{(A_i, S_i^j)\}$ 
11:   end for
12:   for  $A_i, S_i^j \in \mathcal{C}^*$  do
13:      $P^j \in \mathcal{P} \leftarrow \text{sample } \alpha^{III} [S_i^j, \mathcal{F}]$ 
14:      $\mathcal{P}^* \leftarrow \mathcal{P}^* \cup \{P^j\}$ 
15:   end for
16:   Perform simultaneous transmissions according to  $\mathcal{C}^*$ 
   with transmit power levels  $\mathcal{P}^*$ .
17:    $r \leftarrow$  effective data rate in the TXOP
18:   for  $A_i, S_i^j \in \mathcal{C}^*$  do
19:     Update agent  $\alpha^{III} [S_i^j, \mathcal{F}]$  with  $r$ 
20:     if  $A_i \neq A_k$  then
21:       Update agent  $\alpha^{II} [A_i, \mathcal{F}]$  with  $r$ 
22:     end if
23:   end for
24:   Update agent  $\alpha^I [S_k^l]$  with  $r$ 
25: end for

```

Algorithm 1: H-MAB C-SR framework.

VI. PERFORMANCE EVALUATION

We compare the performance of our proposed MAB C-SR schemes (flat MAB, H-MAB) with the following baselines:

- **DCF** – the basic IEEE 802.11 channel access mechanism,
- **SR** – the OBSS Power Detect SR variant of 802.11ax,
- **F-Optimal** and **T-Optimal** – the upper bound model (Section IV) optimized for either fairness or throughput.

We use our custom-built open source simulators, with all design details provided online⁶.

We focus on two types of overlapping topologies: multi-room and open space (Fig. 5). The *multi-room* topologies have a varying floor size (from 2x2 up to 4x4 square rooms) and varying room sizes (separated by walls). Each room contains a single AP and four stations, all uniformly distributed. In the *open space* topologies, the positions of the APs are drawn

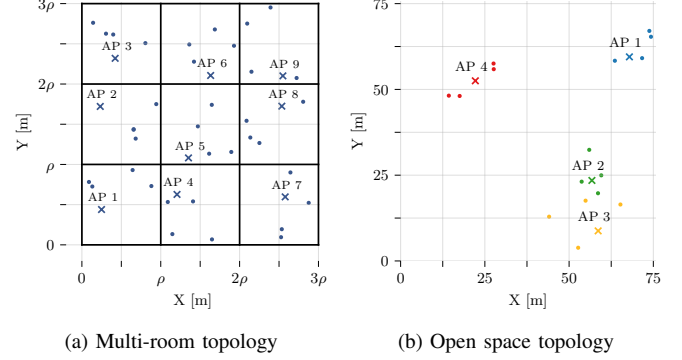


Fig. 5. Exemplary topologies: (a) multi-room and (b) open space. Crosses denote APs, dots – stations, thick lines – walls. In (a), an AP and four stations are all randomly placed in each of the $\rho \times \rho$ m rooms. In (b), the APs are placed randomly, while the stations are placed randomly near their APs.

TABLE I
SIMULATION PARAMETER SETTINGS

Parameter	Value
Band	5 GHz
PHY	IEEE 802.11ax
Channel width	20 MHz
Spatial streams	1, SISO
Guard interval	800 ns
Frame aggregation	Enabled
TXOP duration	5.484 ms
Path loss model	TGax enterprise [24]
Multipath fading	Additive white Gaussian noise
Maximum transmit power	16 dBm
Noise floor	-94 dBm
OBSS_PD threshold	-72 dBm
Wall penetration loss	7 dB
Per-station traffic load	Downlink, full buffer
Frame size	1500 B
Independent repetitions	10

randomly within a square 75 m \times 75 m area without walls. Additionally, station positions are randomly selected from a normal distribution with a mean equal to the AP's position and a variable standard deviation. In each open-space tested topology there are 2–5 APs and each AP has 3–5 associated stations. Further details on all scenarios are available in the aforementioned repository.

The simulation parameter settings are given in Table I. We assume, both in these simulations and in our testbed (Section VII), that all APs operate on the same 20 MHz channel (channel allocation being an orthogonal problem) and use single-stream transmissions. Wider channels and multiple spatial streams would impact the presented results only quantitatively. For data rate selection, we match the MCS used with the expected SINR at the receiver. Each result presented is the average of 10 independent runs. The bands around the lines and the data points denote the 95% confidence intervals of the mean.

Furthermore, we introduce a heuristic to detect the convergence of MAB agents based on Holt's exponential smoothing [25]. We fit the model to the mean value of all runs, then extract the trend, normalize, and smooth. When we detect that the trend is lower than a scenario-dependent threshold for

⁶<https://github.com/ml4wifi-devs/csr>

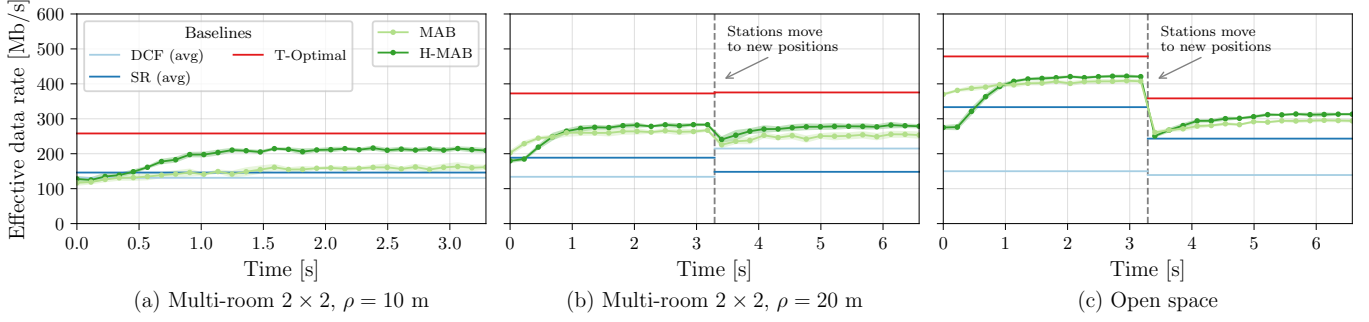


Fig. 6. Aggregate effective data rate for two multi-room and one open space topology.

a certain number of steps (i.e., that the throughput achieved by the MAB-driven network has stabilized), we consider this moment as the convergence point.

A. Transient Performance

We first compare the transient performance of MABs and H-MABs with three baselines (DCF, SR, and T-Optimal). In Fig. 6a, when all stations are static, both our approaches converge equally fast although H-MABs ultimately achieve a higher effective data rate. In Fig. 6b, we introduce nomadic mobility (all nodes change position within their rooms halfway through the simulation), but both MAB architectures quickly adjust to the new conditions. Additionally, in both cases, they outperform DCF and SR and approach the T-Optimal value. In the open space topology with a random position change (Fig. 6c), the convergence time for H-MABs is higher than for flat MABs, although their overall performance is similar to each other. These results also show that MABs and H-MABs can learn a new best C-SR scheduling configuration without resetting their state. Again, both outperform DCF and SR while remaining close to T-Optimal. As we show later, the observed similar behavior of (flat) MABs and H-MABs significantly depends on the underlying topology.

We see (both here and in later results) that SR does not always outperform DCF. The benefits of SR appear only when the separation between APs is large enough to enable SR-based simultaneous transmissions. Otherwise, decreased performance is a known effect [26].

B. Steady-State Performance in Multi-Room Topologies

Next, we focus on steady-state performance. For the multi-room topology, we analyze varying floor configurations (different number of rooms, different room sizes). We present results for $\rho = 20$ m (Fig. 7); results for other ρ values are qualitatively similar and available in our online repository. H-MABs outperform the baselines (DCF, SR) as well as flat MABs in most of the room configurations tested. T-Optimal achieves better rates than H-MABs since it does not account for any minimum fairness guaranteed by our approach (lines 2-3 of Algorithm 1), while F-Optimal achieves worse rates because it prioritizes fairness. This is visible in Fig. 8, which presents a cumulative distribution function (CDF) of the stations' effective data rates in a 2x3 multi-room configuration

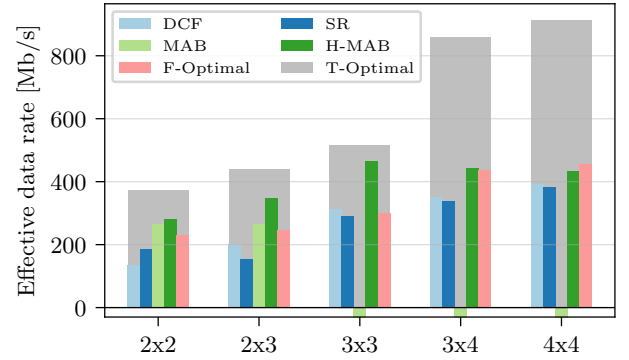


Fig. 7. Performance in multi-room topologies with $\rho = 20$ m. Negative bars indicate unrealistic computational requirements. Confidence intervals are not shown, as they are too small to be visually distinguishable.

TABLE II
AVERAGE STEPS NEEDED TO CONVERGE IN MULTI-ROOM TOPOLOGIES WITH $\rho = 20$ m AND THEIR CORRESPONDING CONVERGENCE TIMES FOR SHORT (0.1 ms), MEDIUM (1 ms), AND LONG (5.484 ms) TXOP DURATIONS. *inf* INDICATES THAT NO ADVANTAGEOUS CONFIGURATIONS WERE FOUND IN A LIMIT OF 200 000 STEPS.

		Scenario size				
		2x2	2x3	3x3	3x4	4x4
Steps	MAB	540	1320	N/A	N/A	N/A
	H-MAB	690	1680	14 400	120 000	inf
Short TXOP	MAB	0.05 s	0.13 s	N/A	N/A	N/A
	H-MAB	0.07 s	0.17 s	1.44 s	12.00 s	inf
Medium TXOP	MAB	0.54 s	1.32 s	N/A	N/A	N/A
	H-MAB	0.69 s	1.68 s	14.40 s	120.00 s	inf
Long TXOP	MAB	2.96 s	7.24 s	N/A	N/A	N/A
	H-MAB	3.78 s	9.21 s	78.97 s	658.08 s	inf

for $\rho = 20$ m. Importantly, in all cases shown in Fig. 7, H-MABs outperform DCF and SR. The lower gain of C-SR compared to DCF and SR in the 3x4 and 4x4 topologies results from the suboptimal performance of H-MABs for such complex scenarios. H-MABs would require more time to converge to a better solution.

In fact, the convergence times of MABs and H-MABs (Fig. 7) increase with topology size and TXOP duration (Table II). Flat MABs have too demanding computational requirements if the number of APs exceeds six due to the

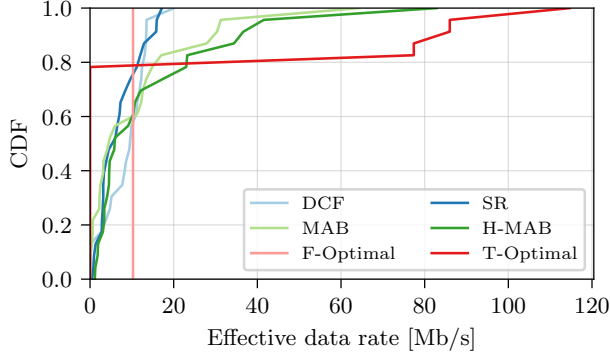


Fig. 8. CDFs for the 2x3 multi-room topology with $\rho = 20$ m. The effective data rate for each station is the average of 10 independent runs.

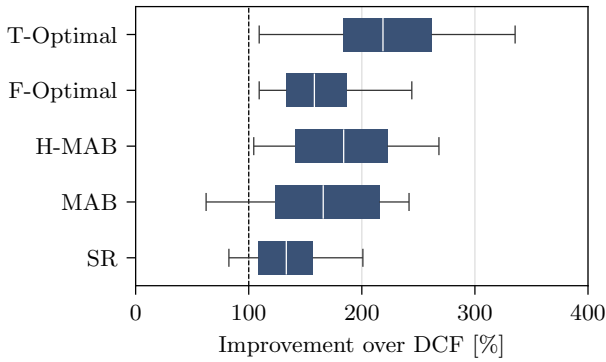


Fig. 9. Effective data rate improvement averaged over 24 random open space topologies with a random topology change halfway through each simulation.

extremely large action space, which is at least partially solved by the proposed H-MABs (up to 12 APs with four associated stations each). These results clearly show that in large and highly congested deployments, another level of C-SR operation is required, i.e., the creation of independent C-SR clusters managed by separate MAB/H-MAB agents, as we show in Section VI-D. This clustering procedure can reduce the size of the action space for a particular sharing AP (e.g., the number of considered shared APs) as well as the convergence time. Additionally, recent 802.11bn advancements [27] suggest that C-SR agreements can be created for long-term operation, and then the convergence time is less critical considering the large possible gains (e.g., the ones observed for the 2x2, 2x3, and 3x3 topologies).

C. Performance in Random Open Space Environments

We next compare different channel access methods in open space topologies in terms of the average effective data rate (Fig. 9). We introduce a random topology change halfway through each simulation, where all nodes are displaced to new positions. This sudden change forces C-SR schemes to adapt to completely different network topologies. These results confirm that (i) both MAB-based approaches outperform not only DCF (on average by 80%) but also legacy SR, (ii) only H-MABs consistently outperform DCF (there are no outliers

TABLE III
AVERAGE STEPS NEEDED BY THE CLUSTERED C-SR TO CONVERGE IN MULTI-ROOM TOPOLOGIES WITH $\rho = 20$ m AND THEIR CORRESPONDING CONVERGENCE TIMES FOR SHORT (0.1 ms), MEDIUM (1 ms), AND LONG (5.484 ms) TXOP DURATIONS.

		Scenario size	
		3x4	4x4
Steps	MAB	1080	760
	H-MAB	960	1040
Short TXOP	MAB	0.11 s	0.08 s
	H-MAB	0.10 s	0.10 s
Medium TXOP	MAB	1.08 s	0.76 s
	H-MAB	0.98 s	1.04 s
Large TXOP	MAB	5.92 s	4.16 s
	H-MAB	5.26 s	5.70 s

below the 100% line), and (iii) H-MABs are on average better than flat MABs. Additionally, T-Optimal and F-Optimal give results similar to those for the multi-room topologies.

D. Clusterization for Larger Topologies

For the highly congested multi-room topologies (3x4, 4x4) analyzed in Section VI-B for which the convergence times (Table II) were particularly large, we implement a C-SR clusterization scheme (shown in the left part of Fig. 10). Each cluster is governed by a single independent MAB or H-MAB agent, so this remains a centralized approach. Recall that in all the cases analyzed, the APs share the same wireless channel. Therefore, this setting shows the operation of clustered C-SR under strenuous conditions (i.e., the contention level does not change compared to the cases analyzed in Section VI-B).

The effective data rate values presented in Fig. 10 are much higher for clustered C-SR, not only in comparison to the baselines (DCF and SR) but also in comparison to the clusterless C-SR schemes which were unable to find a satisfactory solution in a reasonable time. Both effective data rates improve and convergence times drastically decrease (Table III). The changes are especially visible for the (flat) MAB-based C-SR. Without clusterization, it was unable to converge as a result of the extremely large action space. With the clustered approach, it converges quickly and results in data rates comparable to those of H-MABs.

Next, we move to more favorable, symmetrical large topologies, inspired by the IEEE 802.11ax enterprise scenario [24], where each BSS consists of an AP with full buffers and four associated stations, placed at a distance of 2 m (Fig. 11). Neighboring APs are placed at a distance of $\rho = 30$ m from each other. This topology is favorable for C-SR, because it allows parallel transmissions for all APs during the same TXOP (if proper transmission power levels are configured).

For each topology, we implement a clustered C-SR scheme as follows. For the 3x4 and 4x4 cases, the clusterization is the same as in the 3x4 and 4x4 multi-room topologies (Fig. 10). Furthermore, in the 1x4 and 2x4 cases, two clusters are created (each composed of two and four APs, respectively). Each cluster is governed by an independent MAB/H-MAB agent. In such a regular topology, the performance of DCF is the

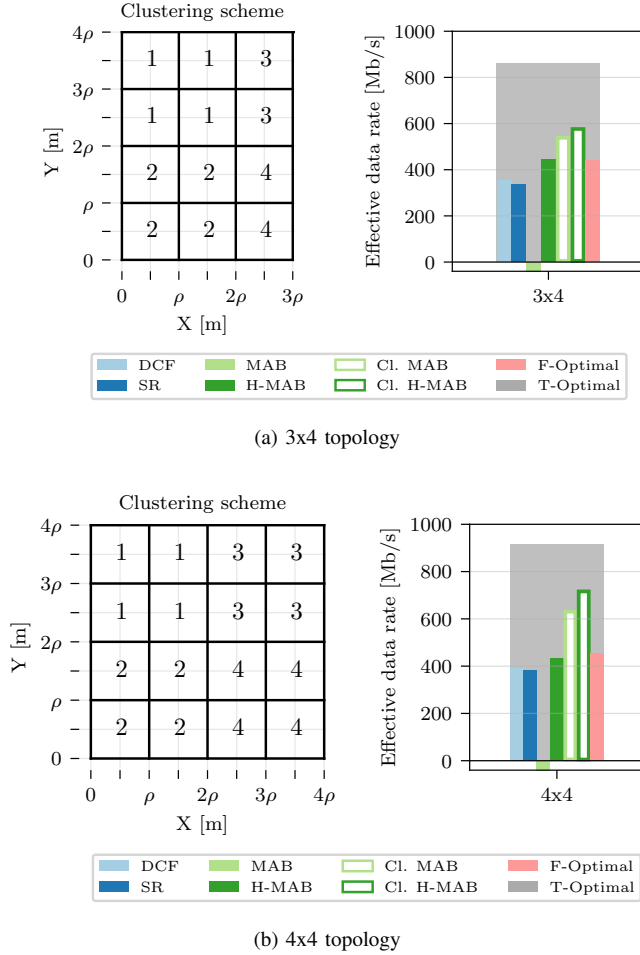


Fig. 10. Performance of clustered MAB and H-MAB variants in the (a) 3x3 and (b) 4x4 multi-room topologies (for $\rho = 20$ m). The numbers in each room identify the assigned cluster.

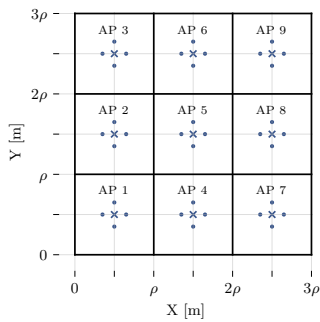


Fig. 11. Exemplary symmetrical topology. Crosses denote APs, dots – stations, thick lines – walls. Each BSS is composed of an AP and four stations are symmetrically placed at a distance of 2 m.

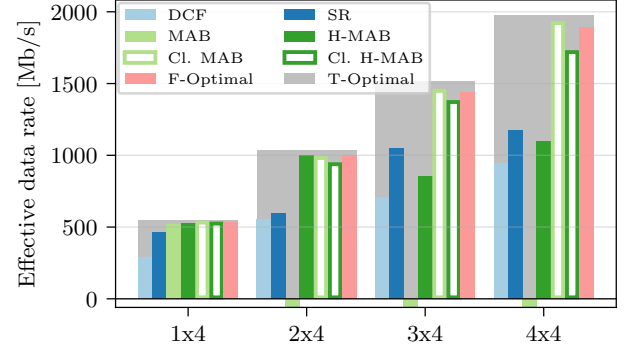


Fig. 12. Performance in the symmetrical enterprise scenario. Confidence intervals are not shown, as they are too small to be visually distinguishable.

worst (because of its strict CCA threshold) and is only partially improved by SR, which relaxes this constraint (Fig. 12). However, the performance of clustered C-SR with MAB and H-MAB independent agents is much better in terms of the effective data rate, which comes close to that of F-Optimal and T-Optimal (which are similar on account of the favorable conditions mentioned before). Importantly, the performance of both clusterless MABs and H-MAB C-SR is worse than that of SR, because of the significant time that they require to converge to the optimum.

A full implementation of a clustered C-SR approach (i.e., with dynamic cluster creation) requires modifications in the presented C-SR scheme (at the network controller level) and is left for future study. However, this subsection shows the potential of clustered C-SR operation and signals an open research challenge.

E. Impact of Legacy Devices

So far, we have analyzed the performance of the proposed C-SR scheme under no external interference (i.e., an isolated case). Here, we measure the impact of additional legacy APs on C-SR operation in an exemplary 2x2 multi-room topology (Fig. 5a). We proceed to uniformly add legacy (non-C-SR) APs so that each of these extra APs contends with exactly one of the four C-SR APs. For example, if there are seven non-C-SR APs, then three of the C-SR APs contend with two extra APs, while the fourth one contends with one extra AP. As reference levels, we use the performance of DCF and SR, both operating in a network topology without the extra APs. Therefore, in Fig. 13, the presented performance of DCF and SR as well as the data point for H-MAB and zero legacy APs, corresponds to the results of the 2x2 room in Fig. 7. We analyze the effective data rate achieved under the condition that one of the C-SR APs successfully wins the channel contention. These results show that when the number of legacy APs increases, the operation of C-SR steadily deteriorates due to an increased probability of collisions and fewer opportunities for successful C-SR transmissions. However, even with multiple additional APs the performance of C-SR is better than that of SR and DCF working in a less crowded environment. This suggests that C-SR can bring important gains and improve

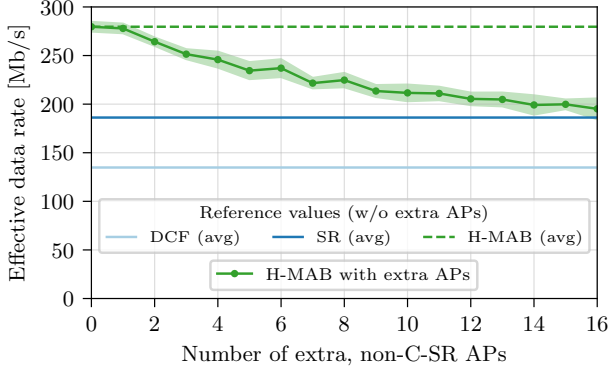


Fig. 13. Impact of additional legacy APs on C-SR operation compared to baseline DCF and SR algorithms operation in less crowded environment (i.e., without additional legacy APs) in the 2x2 multi-room topology with $\rho = 20$ m.

overall channel utilization, not only for homogeneous, but also heterogeneous deployments.

VII. EXPERIMENTAL VALIDATION

To validate our simulation findings, we evaluate our proposed MABs in experiments. In the following, we describe the testbed composition and the results obtained.

A. Testbed Description

We use the Industrial IoT Lab⁷ from Ghent University to perform a real-world validation of our approach. All Wi-Fi devices use openwifi⁸, which is the first open-source implementation of Wi-Fi based on software defined radio (SDR). The design includes an open-source implementation of the physical layer, the lower and higher MAC layers, and is fully compatible with off-the-shelf Wi-Fi nodes.

In addition to the normal Wi-Fi features and flexible parameter configuration, openwifi supports accurate time synchronization with microsecond accuracy [28] as well as a traffic scheduling mechanism that assigns dedicated time slots to hardware queues on a periodic basis [29]. In addition, openwifi allows selecting MCS and transmit power values on a per-time slot basis and remotely configuring the SDR devices using a dedicated controller. The duration of these time slots is fully configurable. With these main enablers, the openwifi-based testbed is suitable to perform tests of concurrent downlink transmissions from multiple APs (C-SR) [14] and, therefore, is an ideal candidate to test our proposed solution in practice. We optimize MAB and H-MAB hyperparameters in a digital-twin of the testbed recreated in our simulation environment. Then, we install (flat) MABs and H-MABs on the network controller, which communicates the decisions on the transmission configurations and transmit power settings to APs using Ethernet connections, as shown in Fig. 4. The controller also collects information whether the performed transmissions were successful or not, which then

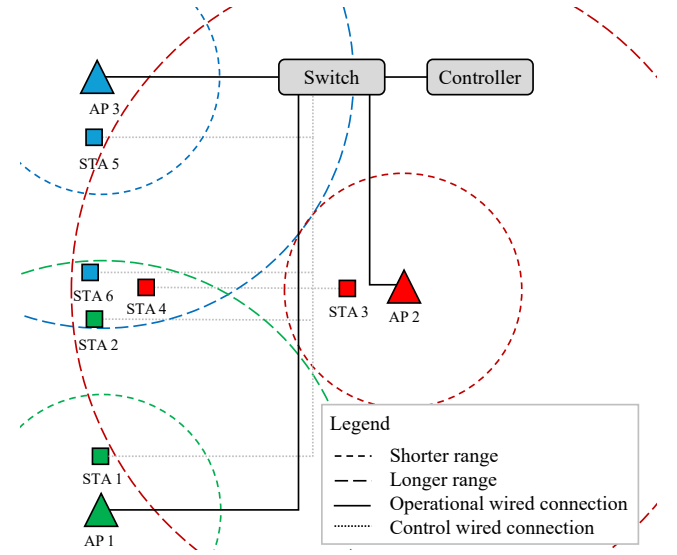


Fig. 14. Schematic illustration of the testbed topology.

serves as a reward for the implemented agents. Additionally, a dedicated switch is used to synchronize all of the APs in the network, while the stations are synchronized over the air.

B. Experimental Setup

The network topology comprises three APs and six stations (Fig. 14). All devices use IEEE 802.11n and operate on the same 20 MHz channel centered at 3.686 GHz (to avoid external interference). Each AP serves two stations: one placed near its serving AP (1 m away) and the other in the boundary coverage zone between all APs (6 m to 8 m away from its AP). However, due to different station placement (i.e., different propagation and channel conditions) their transmission success probability varies under a given MCS (Fig. 15, left Y axis, shows the success probability of different stations for the MCS index set to 6) which constitutes a complex real scenario.

In this setting, concurrent C-SR transmissions are possible by adjusting the transmit power levels of APs. Three discrete power levels are allowed in the testbed: the default value (17 dBm for openwifi APs), the default decreased by 6 dB and the default decreased by 12 dB. All APs are interconnected to a switch and further to a network controller via a wired backbone. This controller collects rewards and determines C-SR scheduling according to Section V.

We configure openwifi so that parallel transmissions can start synchronously. A transmission slot length of $307 \mu\text{s}$ restricts the time slot duration to exactly one 1000 B frame transmitted with MCS 6 (Table IV). We increase the CCA threshold significantly to force nodes to perform transmissions. We update the configuration selected by the agent every 10 frame transmissions; we call this a *step*. To accurately measure the number of received frames, we implement a traffic generation and monitoring framework, which reports the number of sent and received frames over the wired link.

⁷<https://www.ugent.be/ea/idlab/en/research/research-infrastructure>

⁸<https://github.com/open-sdr/openwifi>

TABLE IV
TESTBED PARAMETER SETTINGS

Parameter	Value
Band	3.7 GHz
PHY	IEEE 802.11n
Channel width	20 MHz
Spatial streams	1, SISO
MCS index	6
Tx power attenuation levels	{0, -6, -12} dB
Frame size	1000 B
CCA threshold	-52 dBm
Transmission slot length	307 μ s
Packets per slot	1
Slots per configuration	10
Number of repetitions	10

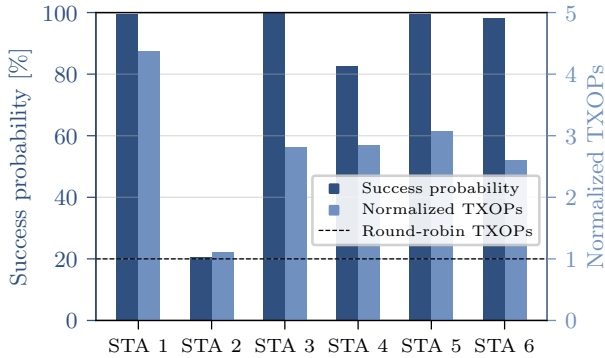


Fig. 15. Testbed results: success probability under single concurrent transmissions and number of TXOPs given by H-MABs to each station normalized to a perfect round robin scheme (i.e., the long term performance of DCF).

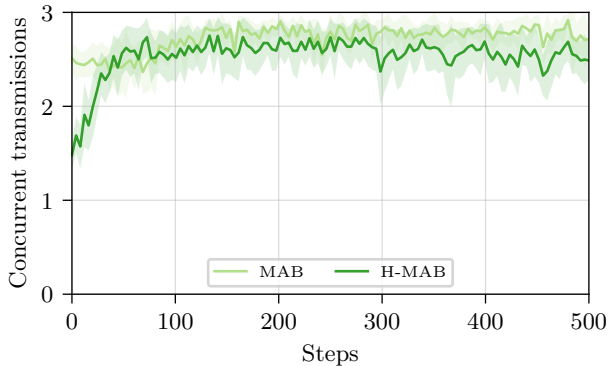


Fig. 16. Comparison of MAB and H-MAB C-SR operation in the testbed: average number of concurrent transmissions.

C. Results

Our results show that, in this C-SR setting, both flat MABs and H-MABs quickly find appropriate AP-station pairs and transmit power settings that enable three concurrent transmissions (Fig. 16). Since the tested topology is relatively small, flat MABs slightly outperform H-MABs. Nonetheless, the results validate that the proposed C-SR approach based on MABs is feasible and works well in a real environment.

We also show actions taken by H-MABs gathered from the last 200 steps of all runs (Fig. 17). More than 80% of actions

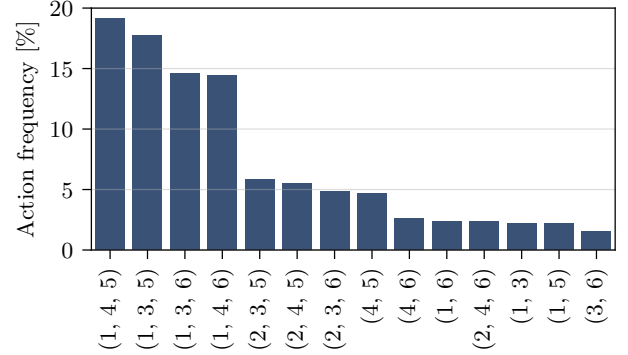


Fig. 17. Actions taken by H-MAB in the last 200 steps of the experiments. The X axis displays the set of stations selected for parallel transmissions. Only actions with a frequency above 0.5% are shown.

taken involve three parallel transmissions. As expected, C-SR configurations including station 2 are less frequent because they result in lower throughput. To explain this, we examine the underlying success probability of transmissions to each station (Fig. 15, left Y axis), where station 2 has noticeably worse radio conditions. As shown in Fig. 15 (right Y axis), station 2 preserves its TXOPs as compared to DCF. This behavior confirms that H-MABs not only configure parallel transmissions but also adjust to the observed environmental conditions. Furthermore, this confirms that fairness among stations is not a straightforward goal for C-SR, as some stations may be unable to benefit from C-SR (but should be granted no fewer TXOPs than under DCF, which we do achieve with the proposed framework).

VIII. DISCUSSION AND FUTURE WORK

Our first contribution is the theoretical upper bound model of C-SR. We show that it can be optimized for fairness or throughput and provide an additional scalability analysis (Appendix A). We expect new C-SR scheduling solutions to emerge as work on IEEE 802.11bn progresses, and our model provides the research community with an upper bound baseline. As future work in this area, we plan to address other optimization goals (such as proportional fairness or assigning each station a fair number of transmission opportunities, ideally no less than under DCF) and add lexicographic optimization. With these updates, the model should provide more realistic (less extreme) results than those presented here.

Our second contribution, inspired by the probing-based algorithms for data rate selection algorithms in Wi-Fi networks, is the application of MABs (both flat and hierarchical) for solving the C-SR scheduling problem. Our results show that MABs are a highly promising learning mechanism. In addition to good performance (high throughput, low convergence time for small and medium-sized networks) and resilience to abrupt topology changes, an important advantage is their simplicity, which makes the solution easy to deploy in hardware. For large-scale networks, we evaluate clustered MABs, which also demonstrate strong performance and rapid convergence. Furthermore, we assess the impact of legacy APs on the operation of our proposal and find that H-MABs are reasonably resilient.

A classical MAB is defined for a constant number of arms (choices), and this is in contrast to the dynamic behavior of users who join and leave the network. Although this may seem a blocking characteristic, we argue that this limitation is tied only to the initial implementation presented. Consider a highly realistic scenario where APs operate with non-full buffers while stations arrive and depart freely. In these cases, it is reasonable to preallocate agents up to some maximal number of APs and stations, respectively. Some of the preallocated agents or arms are not used and are called coldspares. When new stations arrive, we can allocate their adequate coldspares. When APs have empty buffers, we can move them to coldspares. The update operation remains unchanged, as we update only the agents that are in use. We modify the sampling to ensure that the coldspare is never selected. For this, we propose rejection sampling that repeats the sample operation when the coldspare is selected. This procedure fishes for probabilistic agents, and we have an undergoing effort to implement this dynamic based on a proof-of-concept implementation that can be found in *Reinforced-lib* [30].

Furthermore, we foresee the following other future research items. Although the presented MABs use aggregate network throughput as a reward, other optimization goals (e.g., latency or fairness) can be considered. To simplify the presentation, we have idealized the MCS selection, but it can and should be jointly optimized along with the transmission pairs and transmit power. Our H-MAB framework can operate in a distributed manner (having independent agents on different APs and synchronizing updates), but this would complicate the signaling and introduce additional overhead, so we do not describe it in this work. However, a distributed learning scheme with independent agents (i.e., assuming little or no communication) would be interesting to study (especially to see if convergence can be achieved). Finally, large-scale scenarios with increased mobility, external interference and other dynamics, as well as uplink traffic should be considered in the future.

APPENDIX A

SCALABILITY ANALYSIS OF THE OPTIMIZATION MODEL

The practical applicability of the proposed optimization model (Section IV) hinges on its ability to efficiently handle real-world network scenarios, which often involve a larger number of APs and stations. Therefore, it is crucial to assess the scalability of the model, i.e., how its computational requirements grow as the network size increases. This analysis will provide valuable insights into the model's feasibility for deployment in real-world sized Wi-Fi networks.

The performance and scalability of the optimization model are significantly influenced by the underlying solver used to find the optimal solution. In this study, we compare two solvers: COIN-OR branch and cut (CBC) [31] and IBM ILOG CPLEX Optimization Studio (CPLEX) [32]. CBC is an open-source MILP solver readily available in the PuLP library in Python. Its accessibility and ease of integration make it a popular choice for prototyping and experimentation. However, being a general-purpose solver, it is not optimized

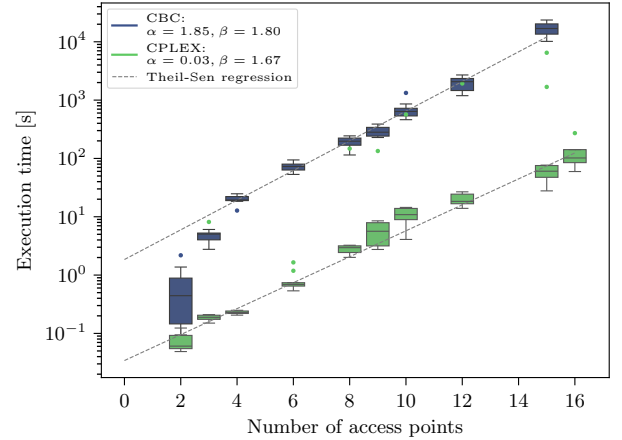


Fig. 18. Scalability of the optimization solution depending on the solver used.

for the specific structure of our C-SR scheduling problem. CPLEX is a commercial solver widely recognized for its high performance and advanced algorithms tailored for solving large-scale linear and MILP problems. While it requires a separate license, its computational efficiency and scalability make it a strong candidate for handling complex optimization tasks, such as C-SR scheduling.

We assess the scalability of the T-Optimal model in the multi-room topology. We vary the number of APs within this grid to simulate networks of increasing size. The execution time of both solvers (CBC and CPLEX) is measured for each network size 10 times, providing a direct comparison of their scalability.

Fig. 18 presents the execution time of both solvers on a logarithmic scale as a function of the number of APs in the network. The Theil-Sen regression lines [33], chosen for their robustness to outliers, are fitted for both solvers in log space to determine the quantitative measure of the scalability of each solver. The results clearly indicate that the execution time for both solvers increases exponentially with network size, as expected. The regression lines capture this exponential relationship, with their slopes representing the exponents (β) of the time complexity. CPLEX demonstrates superior scalability ($\beta = 1.67$) compared to CBC ($\beta = 1.80$), which implies that CPLEX's execution time grows at a slower rate as the network size increases, making it computationally more efficient, especially for larger networks.

While CPLEX generally exhibits lower execution times, it also shows greater variation, with some outliers having execution times comparable to those of the CBC solver. This suggests that CPLEX's performance may be more sensitive to specific problem instances or network configurations, whereas CBC's performance tends to be more consistent, albeit slower overall.

ACKNOWLEDGMENT

This research was funded by the National Science Centre, Poland (2020/39/I/ST7/01457) and by the German Research Foundation (DFG DR 639/28-1). This research was

also partially funded by the Flemish FWO SBO S003921N VERI-END.com project. Additionally, the work of Boris Bellalta was supported by Wi-XR PID2021123995NB-I00 (MCIU/AEI/FEDER,UE).

REFERENCES

- [1] P. Imputato et al., “Beyond Wi-Fi 7: Spatial reuse through multi-AP coordination,” *Computer Networks*, vol. 239, p. 110 160, 2024.
- [2] L. G. Giordano et al., “What Will Wi-Fi 8 Be? A Primer on IEEE 802.11bn Ultra High Reliability,” *preprint arXiv:2303.10442*, 2023.
- [3] R. J. Yu, *Specification Framework for TGbn*, Accessed on Oct 10, 2024. [Online]. Available: <https://mentor.ieee.org/802.11/dcn/24/11-24-0209-05-00bn-specification-framework-for-tgbn.docx>.
- [4] W. Yin et al., “MAC-layer rate control for 802.11 networks: A survey,” *Wireless Networks*, vol. 26, pp. 3793–3830, 2020.
- [5] R. Grünblatt, I. Guérin-Lassous, and O. Simonin, “Simulation and performance evaluation of the Intel rate adaptation algorithm,” in *Proc. of ACM MSWiM*, 2019, pp. 27–34.
- [6] A. Slivkins et al., “Introduction to multi-armed bandits,” *Foundations and Trends® in Machine Learning*, vol. 12, no. 1-2, pp. 1–286, 2019.
- [7] M. Wojnar et al., “IEEE 802.11bn Multi-AP Coordinated Spatial Reuse with Hierarchical Multi-Armed Bandits,” *IEEE Communications Letters*, vol. 29, no. 3, pp. 428–432, 2025.
- [8] J. Lee et al., “Coordinated multipoint transmission and reception in LTE-advanced systems,” *IEEE Comm. Mag.*, vol. 50, no. 11, 2012.
- [9] S. Verma et al., “A Survey on Multi-AP Coordination Approaches Over Emerging WLANs: Future Directions and Open Challenges,” *IEEE Communications Surveys & Tutorials*, vol. 26, no. 2, pp. 858–889, 2024.
- [10] D. Nunez et al., “TXOP sharing with coordinated spatial reuse in multi-AP cooperative IEEE 802.11be WLANs,” in *Proc. of IEEE CCNC*, 2022.
- [11] D. Nunez, M. Smith, and B. Bellalta, “Multi-AP Coordinated Spatial Reuse for Wi-Fi 8: Group Creation and Scheduling,” *preprint arXiv:2305.04846*, 2023.
- [12] J. Wang and X. Fang, “Research on Next-Generation Wi-Fi Spatial Reuse Power Control Based on Federated Reinforcement Learning,” in *2024 IEEE VTC2024-Spring*, IEEE, 2024.
- [13] M. Talukder and J. Xie, “Enhanced Coordinated Spatial Reuse: Bidirectional Multiple AP Coordination for IEEE 802.11be,” in *IEEE ICC 2023*, IEEE, 2023, pp. 660–665.
- [14] J. Haxhibeqiri et al., “Coordinated Spatial Reuse for WiFi Networks: A Centralized Approach,” in *2024 IEEE 20th International Conference on Factory Communication Systems (WFCS)*, 2024.
- [15] J. Haxhibeqiri et al., “Coordinated SR and Restricted TWT for Time Sensitive Applications in WiFi 7 Networks,” *IEEE Communications Magazine*, vol. 62, no. 8, pp. 118–124, 2024.
- [16] F. Wilhelmi et al., “Throughput Analysis of IEEE 802.11bn Coordinated Spatial Reuse,” *preprint arXiv:2309.09169*, 2023.
- [17] K. Lu et al., *Multi-AP Coordination Setup Scheme*, doc.: IEEE 802.11-24/1217r02, Aug. 2024.
- [18] W. Ciezobka et al., “Using ranging for collision-immune IEEE 802.11 rate selection with statistical learning,” *Computer Communications*, vol. 225, pp. 10–26, 2024.
- [19] A. Capone, G. Carello, I. Filippini, S. Gualandi, and F. Malucelli, “Routing, scheduling and channel assignment in wireless mesh networks: Optimization models and algorithms,” *Ad Hoc Networks*, vol. 8, no. 6, pp. 545–563, 2010.
- [20] M. Pióro et al., “On max–min fair flow optimization in wireless mesh networks,” *Ad Hoc Networks*, vol. 13, pp. 134–152, 2014.
- [21] R. Spliet, “A simple perspective on simultaneous column and row generation,” *Operations Research Forum*, vol. 5, no. 69, Aug. 2024.
- [22] J. Desrosiers and M. E. Lübbecke, “A primer in column generation,” in *Column Generation*, G. Desaulniers, J. Desrosiers, and M. M. Solomon, Eds. Boston, MA: Springer US, 2005, pp. 1–32.
- [23] G. Papoudakis, F. Christianos, L. Schäfer, and S. V. Albrecht, “Benchmarking multi-agent deep reinforcement learning algorithms in cooperative tasks,” in *NeurIPS Datasets and Benchmarks*, 2020.
- [24] S. Merlin et al., *TGax Simulation Scenarios*, doc.: IEEE 802.11-14/0980r16, Nov. 2015.
- [25] C. C. Holt, “Forecasting seasonals and trends by exponentially weighted moving averages,” *International Journal of Forecasting*, vol. 20, no. 1, pp. 5–10, 2004.
- [26] M. Natkaniec and N. Bieryt, “An Analysis of the Mixed IEEE 802.11ax Wireless Networks in the 5 GHz Band,” *Sensors*, vol. 23, no. 10, 2023.
- [27] J. C. et al., *Details on the unified MAPC framework*, Accessed on April 2, 2025. [Online]. Available: <https://mentor.ieee.org/802.11/dcn/25/11-25-0502-00-00bn-details-on-the-unified-mapc-framework.pptx>.
- [28] M. Aslam et al., “Hardware efficient clock synchronization across Wi-Fi and Ethernet-based network using PTP,” *IEEE Transactions on Industrial Informatics*, vol. 18, no. 6, pp. 3808–3819, 2021.
- [29] J. Haxhibeqiri et al., “Bringing time-sensitive networking to wireless professional private networks: Filling gaps and bridging the innovation,” *Wireless Personal Communications*, vol. 121, pp. 1255–1271, 2021.
- [30] M. Wojnar et al., “Reinforced-lib: Rapid prototyping of reinforcement learning solutions,” *SoftwareX*, vol. 26, 2024.
- [31] J. Forrest et al., *Coin-or/cbc: Release releases/2.10.12*, version releases/2.10.12, Aug. 2024.
- [32] IBM, *IBM ILOG CPLEX Optimization Studio CPLEX user’s manual*, 2023. [Online]. Available: <https://www.ibm.com/docs/en/icos/20.1.0?topic=cplex-users-manual>.
- [33] X. Dang, H. Peng, X. Wang, and H. Zhang, “Theil-sen estimators in a multiple linear regression model,” *Olemiss Edu*, vol. 2, 2008.