

A Survey on Opportunistic Scheduling in Wireless Communications

Arash Asadi, *Student Member, IEEE*, and Vincenzo Mancuso, *Member, IEEE*

Abstract—Wireless technology advancements made opportunistic scheduling a popular topic in recent times. However, opportunistic schedulers for wireless systems have been studied since nearly twenty years, but not implemented in real systems due to their high complexity and hardly achievable requirements. In contrast, today’s popularity of opportunistic schedulers extends to implementation proposals for next generation cellular technologies. Motivated by such a novel interest towards opportunistic scheduling, we provide a taxonomy for opportunistic schedulers, which is based on scheduling design’s objectives; accordingly, we provide an extensive review of opportunistic scheduling proposals which have appeared in the literature during nearly two decades.

The huge number of papers available in the literature propose different techniques to perform opportunistic scheduling, ranging from simple heuristic algorithms to complex mathematical models. Some proposals are only designed to increase the total network capacity, while others enhance QoS objectives such as throughput and fairness. Interestingly, our survey helps to unveil two major issues: (i) the research in opportunistic is mature enough to jump from pure theory to implementation, and (ii) there are still under-explored and interesting research areas in opportunistic scheduling, e.g., opportunistic offloading of cellular traffic to 802.11-like networks, or cooperative/distributed opportunistic scheduling.

Index Terms—Resource allocation, Opportunistic Scheduling.

I. INTRODUCTION

BROADBAND data provisioning became a design principle for 3G and 4G networks as the dominant load on cellular network has switched from voice to data traffic. The emerging wireless technologies (e.g. HSPA [1], LTE [2] and WiMAX [3]) have sophisticated multiple access methods such as orthogonal frequency division multiple access (OFDMA) and code division multiple access (CDMA). In spite of capacity increments created by new technologies, mobile operators are still struggling to satisfy users demands [4], [5]. A viable alternative to bandwidth increment is to allocate the existing bandwidth optimally or at least more efficiently. Scheduling is the functionality responsible for resource allocation among users, and its job is to decide which user should transmit/receive and when, therefore it impacts on bandwidth utility efficiency.

In today’s cellular networks, the design and choice of schedulers is left to the operator. As a consequence, the significant impact of scheduling algorithms on network performance made them a popular research topic. The current

implementations of schedulers do not utilize the advanced features of physical layer and users are often scheduled regardless of their channel conditions. On one hand, fourth generation cellular technologies mostly operate on multiple subcarriers (e.g., OFDMA). On the other hand, the majority of newly proposed schedulers require user channel information to perform scheduling and it is not practical to feedback every user’s channel condition over all subcarriers.

As a consequence, commonly adopted schedulers operate regardless of user’s channel condition, e.g. : a) Round Robin, that serves users in a circular manner without any other consideration, b) earliest deadline first (EDF), which schedules the packet that will be expired the soonest, c) weighted fair queueing (WFQ), that allocates the resources with respect to the weights associated with every user [7]. In contrast, schedulers have been proposed, but not fully implemented yet, that take advantage of physical layer information, such as the user channel state. This type of schedulers are called *opportunistic schedulers*. As an example, some vendors deploy simplified opportunistic schedulers which follow the *proportional fair* strategy [6]. Proportional fair schedulers consider the current channel state of the users and the history of received throughput of each user. Unfortunately, the implementation accuracy of proportional fair schedulers is commonly limited by scarce memory and processing resources deployed at base station.

The recent literature does not offer any survey on opportunistic scheduling, discussing the state-of-the-art proposals and approaches in this field. Therefore, given the importance of opportunistic scheduling in today’s communications, our aim is to provide an extensive survey on the opportunistic scheduling algorithms emerged within the last two decades. We start the survey with a brief introduction to opportunistic scheduling. Next, the existing work is categorized based on our proposed taxonomy, namely based on *capacity*, *fairness*, *QoS* issues, and *distributed / centralized assumptions*, and the significant works under each category are reviewed. In order to provide a better understanding of the proposals, we include a comparison and evaluation of such proposals in the end of each section. A list of common techniques, approaches and evaluation methods is provided based on the reviewed literature. In addition, we discuss open research problems with some suggested solutions and possible future direction of research on opportunistic scheduling.

The rest of the paper is organized as follows: in Section II, we give a short introduction on opportunistic scheduling algorithms, considering their approach through opportunism. We also provide a taxonomy for opportunistic schedulers. In Section III, we review schedulers that increase the total network capacity. In Section IV, we survey QoS-oriented

Manuscript received 25 April 2011; revised 25 September.

The authors are with Institute IMDEA Networks, Madrid, Spain and Carlos III University, Madrid (email: {arash.asadi, vincenzo.mancuso}@imdea.org).

TABLE I: The scheduling policy for MaxRate, MaxWeight and Exp rule schedulers.

Scheduler	Scheduling policy	Notations
MaxRate	$i^*(t) = \arg \max_{i \in N} R_i(t)$	N : Total number of users.
MaxWeight	$i^*(t) = \arg \max_{i \in N} R_i(t)Q_i(t)$	$i^*(t)$: The user which is scheduled at time t .
Exp rule	$i^*(t) = \arg \max_{i \in N} \gamma_i R_i(t) e^{\frac{a_i Q_i(t)}{\beta + [Q_i(t)]^\eta}}$	$R_i(t)$: Achievable rate by user i at time t . $Q_i(t)$: Queue size of user i at time t . γ_i, a_i : Arbitrary positive fixed constants. $\eta, \beta \in (0, 1)$. $Q(t) \doteq (1/N) \sum_i a_i Q_i(t)$

proposals. Section V contains proposals with fairness considerations. State of the art distributed opportunistic schedulers are discussed in Section VI. Some of the open issues on opportunistic scheduling are listed in Section VII. Finally, in Section VIII, we conclude the paper and provide a list of common techniques and evaluation approaches in opportunistic proposals. We also point out future research trends.

II. OPPORTUNISTIC SCHEDULING

Here, we provide a brief introduction on opportunistic scheduling. We point out incentives for employing opportunistic schedulers in cellular networks and provide a taxonomy for the classification of opportunistic schedulers.

A. Incentives behind opportunistic scheduling

Opportunistic schedulers take into account information such as the channel quality in terms of QoS metrics (i.e., throughput, delay, jitter) that allows the scheduler to find the proper transmission resources for each user. The notion of opportunistic scheduling was first introduced by Knopp and Humblet in [8]. They showed that using the multiuser diversity in scheduling process can significantly improve the capacity. In a pure opportunistic approach, the scheduler always chooses the user in the best channel condition to use the resources. This approach is referred to as MaxRate scheduling in the literature [9]. The gain in opportunistic scheduling depends on the multiuser diversity due to random wireless channel impairments such as fading and multipath. After [8], researchers aimed at taking advantage of diversity caused by the channel impairments instead of eliminating it. Some authors even propose techniques such as opportunistic beamforming to increase the multiuser diversity [10]–[12]. With this technique, the same signal is transmitted over multiple antennas with different transmission powers. This increases the channel diversity of users, which leads to improved opportunistic gain. MaxWeight [13] is another opportunistic scheduler that selects the user with the highest product of queue length and transmission rate. MaxWeight was considered throughput-optimal before the authors in [14] prove otherwise under flow level dynamics. However, MaxWeight is throughput-optimal with fully backlogged queues. Exp rule schedulers [15] are throughput-optimal schedulers that prioritize users based on an exponential formula using queue size and transmission rate of every user. Table I shows the scheduling policy of the aforementioned throughput-optimal schedulers.

Opportunistic scheduling has been proposed not only to improve capacity or QoS. For instance, Wong *et al.* proposed

an opportunistic scheduling strategy to leverage multiuser diversity in an OFDM systems and do attempt to minimize the overall transmission power [16].

Although opportunistic schedulers are not widely implemented in commercial deployments yet, they have been proven to be more suitable than non-opportunistic schedulers for wireless networks. Initial proposals employed heuristics to make scheduling decision which was improvement in comparison to non-opportunistic schedulers but not optimal [17], [18]. For instance, Badia *et al.* propose a joint scheduling and resource allocation framework tailored for WiMAX and operates based on a heuristic credit-based scheduler [17]. Later, researchers obtained mathematical models for wireless channels which allowed them to study different aspect of the system (e.g., queue stability, delay, throughput) and propose optimal schedulers. The majority of recent literature is dedicated to modeling and optimizing proposals for cellular networks using different techniques which will be discussed later. To explore these proposals, we now introduce our taxonomy for opportunistic scheduling.

B. Taxonomy

The available literature on opportunistic scheduling tackles the issue of scheduling from different aspects. Most of these proposals are subclasses of four major categories: capacity, QoS, fairness, and distributed scheduling. Proposals that purely improve network capacity regardless of QoS or fairness implications are listed under the first category. We further categorize these proposals into schedulers with full/non-full channel state information and schedulers for cognitive radio networks. The second category covers proposals that aim to improve specific quality indices such as delay, jitter, average throughput, etc. The works under this category are also divided into subcategories of single QoS objective and multiple QoS objective. Works under the third category tackle the fairness issue in opportunistic scheduling. In fact, opportunistic scheduling of users can lead to highly unfair treatment toward different users. Therefore, fairness is always a concern in opportunistic scheduling because users with low channel quality can be sacrificed due to the greedy nature of pure opportunistic approaches. Eventually, most of the proposals focus on mechanisms that can be implemented at the base station of a cellular network. However, distributed mechanisms have been proposed as well. We classify these distributed opportunistic algorithms in a separate class because they usually aim at a different network configuration.

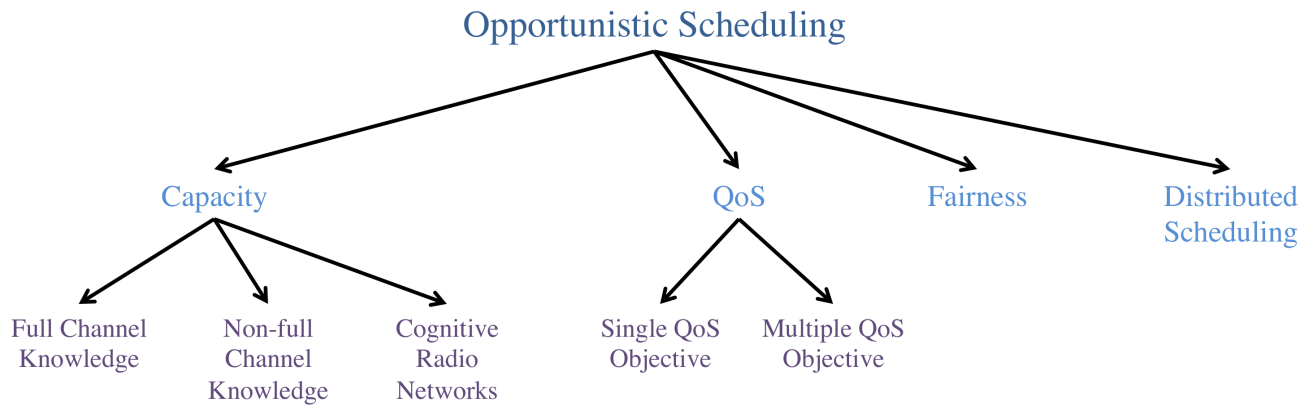


Fig. 1: Opportunistic Scheduler Classification.

In this paper, we review the existing literature based on the proposed taxonomy, which is schematically depicted in Fig. 1.

III. CAPACITY

In many proposals, opportunistic scheduling is employed as a solution to enhance the total capacity of wireless networks. As shown in Fig. 1, we further classify these proposals based on the assumption of *full channel state information* (CSI) availability (i.e., base station has instantaneous knowledge of all users' CSI). Hence, the scheduler always knows which user has the best channel state at every time instant [19]–[26]. However, note that CSI of mobile users is acquired via feedback in cellular networks, so that the perfect and instantaneous (*full*) knowledge of user's CSI is hardly practical in real deployments. To cope with this issue, many proposals leverage the knowledge of user's channel statistical behavior, and CSI samples, rather than the knowledge of the exact and instantaneous CSI. Significantly different is the case of opportunistic schedulers improving capacity by using cognitive radio techniques. Indeed, with cognitive radio, full or partial CSI can be available, but opportunism mainly focuses on which resources can be used and when, given that some other primary user has strictly priority access to the wireless resources. Therefore, in the following we describe first those proposals relying on full CSI availability, then proposals not using full CSI, and eventually proposals based on cognitive radio.

A. Resource allocation with full channel state information

Here, we survey the work attempting to enhance wireless network throughput under the assumption of full CSI availability at the base station.

In many wireless technologies, users can transmit over more than one carrier. This ability extends the opportunistic scheduling decision process to carrier allocation among users. Andrews and Zhang [19] tackle the problem of scheduling in a multi-carrier wireless system. Their paper is dedicated to adapt the MaxWeight algorithm for multi-carrier scenarios for which they define three objective functions that emulate the MaxWeight behavior. The first objective function simply maximizes the product of queue size and feasible rate for

each user over all subcarriers. The second and third objective functions are NP-hard problems that account for the ignorance of MaxWeight algorithm towards users with small queue and bad channel quality, as discussed in [14]. To serve this purpose, the second objective function prioritizes the users with small queues and the third objective function maximizes the negative drift of a Lyapunov function [27] (i.e., maximizes the queue length variation in every slot). The authors propose five algorithms based on the objective functions defined for MaxWeight. The algorithms which are derived from second and third objectives inherit the NP-hardness. The authors solve the NP-hard algorithms via approximations and prove their stability. The simulation results showed that the algorithms based on the second and third objective provide better performance. They also show that the algorithms which optimize the scheduling decision over all carriers instead of local carrier optimization outperform the rest.

Liu *et al.* [20] state that throughput-optimal algorithms in single channel wireless networks are not necessarily throughput-optimal in multi-channel wireless networks. Hence, they propose a joint channel-assignment and workload based scheduler (CA-WS), which is throughput-optimal in multi-channel wireless networks. In [20], flows are classified into two groups, namely, transient and resident and every flow is associated with a file transfer from source to destination. A transient flow is a flow whose file is not fully buffered at the base station, while a resident flow has fully transmitted the file to the base station. Every flow starts from a source and crosses the base station to reach the destination. Let's assume that all flow sources try to transmit a file to the base station. The scheduler keeps track of the number of allocated time slots on each channel for every flow and the CA-WS algorithm chooses the best combination of channels and slots for every user to maximize the network capacity. Under CA-WS, the transient flows are not scheduled until their status is changed to resident. This induces longer transmission delays and degrades the performance of CA-WS under low and medium traffic load. To solve this issue, the authors propose a hybrid CA-WS scheme which uses CA-WS to serve resident flows and MaxWeight to serve transient flows. In particular, in order to eliminate the full CSI assumption, the authors propose to use the learning

process that was earlier proposed in [28]. The performance of CA-WS is compared with the MaxWeight scheduler via simulation. The results confirm that CA-WS performance is poor under low and medium traffic load. However, CA-WS performs much better in presence of high traffic load. CA-WS has higher blocking probability than MaxWeight and hybrid scheduler because it waits for the files to be fully transmitted. The results also show that the hybrid scheduler can serve 20% more traffic than MaxWeight and CA-WS while maintaining a zero blocking probability. The hybrid scheduler has better delay performance than CA-WS and MaxWeight. MaxWeight has lower delay than CA-WS when the traffic is low, but its performance degrades drastically with traffic increment.

Al-Zubaidy *et al.* in [21] target scheduling in high-speed downlink packet access (HSDPA) networks. Modeling the behavior of HSDPA network is challenging because it uses both time division multiplexing and code division multiplexing. A 2-state finite state Markov chain (FSMC) [29] is used to model channel state transition probabilities. The HSDPA is modeled by a 5-tuple $(T, S, A, P_{ss'}(a), R(s, a))$, where T is time, S and A represent the system and action space. $P_{ss'}(a)$ and $R(s, a)$ are the system state and reward functions when the system is taking action a in state s . The reward function can be used to tune between throughput and fairness. The authors introduce the optimal scheduling policy based on their proposed model and solve the model with the dynamic programming technique [30]. Next, they study the behavior of the optimal policy using a Markov decision process (MDP), based on which they propose a heuristic scheduling policy with near optimal performance but lower complexity. The performance of the optimal policy and the one of the heuristic are shown to be comparable using simulation. Round Robin scheduler is also included in the simulation as a benchmark. System throughput under light traffic load is the same for Round Robin, optimal, and heuristic policies. The optimal policy performs 10% and 25% better than Round Robin under medium and high traffic load, respectively. Al-Zubaidy *et al.* [21] investigate the impact of the number of codes per user, i.e., the code chunk granularity, on the scheduling policy. It is observed that higher code chunk granularity increases the average queue delay. Hence, policies with finer granularity are more favorable.

B. Resource allocation without full channel state information

Here, we describe the proposals considering that base stations do not have direct and instantaneous access to user's CSI, but they can periodically acquire CSI from mobile user's reports, see Fig. 1. These *non-full-CSI* proposals are inherently different for the ones surveyed in the previous subsection, and therefore lead to substantially different mechanisms.

In [28], Liu *et al.* propose a throughput-optimal scheduler that does not require any prior knowledge of channel state and user demands. This can be achieved using the so called workload-base scheduling with learning (WSL). The authors define the flows that continuously inject traffic as *long-lived* and those with finite number of bits as *short-lived*. In order to find the maximum possible data rate of short-lived flows,

their data rate is monitored for a learning period. The authors of [28] also provide the necessary conditions for stability of a scheduler which is: a) the service allocated to each users should not be less than what was requested if the service were supportable at all. b) the total airtime allocated to short-lived and long-lived flows should less than or equal than the total available time. The authors prove that WSL is throughput-optimal.

In the same paper, Liu *et al.* discuss the basic problem of MaxWeight, e.g., a flow with small backlog may never be served. A solution for this problem is to use the product of the head of line delay (delay-based scheduling). However, in [28] it is shown that the delay-based scheduler is also not stable in presence of short-lived flows. The authors conclude the paper with a set of simulations to evaluate the performance of WSL, MaxWeight, and delay-based scheduler. It is shown that WSL can sustain a zero blocking probability while admitting almost 20% more traffic. WSL also shows better delay performance.

The authors of [20] and [28] benchmark their proposals against the MaxWeight algorithm. In Table II, we can see how these proposals affect the blocking probabilities with reference to MaxWeight. Since CA-WS increases the waiting time of the transient flows, it increases the call blocking probability that results in lower performance than MaxWeight (i.e., negative gain). The rest of the proposed schedulers improve the call blocking probability over MaxWeight.

TABLE II: The effect of proposed algorithms on call blocking probability normalized over MaxWeight achieved rate under high traffic load.

Algorithm	Call blocking probability normalized to MaxWeight's blocking probability (%)
WSL	≈ 71
Delay-Based	≈ 14
CA-WSU	≈ -100
Hybrid CA-WSU	≈ 60

Ouyang *et al.* propose a scheduling algorithm with near-optimal performance in [31]. The paper tackles the issue discussed earlier: in practice the scheduler will know a user's CSI after it received a transmission feedback from the user, not before. The classic *exploration vs. exploitation* problem comes to picture under this situation: the scheduler has to decide whether to wait for more information hoping to make a better decision (future gain) or to transmit based on the current information (immediate gain). Ouyang *et al.* model the fading channel via a two-state Markov chain, high h_i and low l_i states representing the maximum allowable data rates in the lower and higher states, respectively, and propose a joint estimation and scheduling process using a partially observable Markov decision process (POMDP). We can observe the scheduling process in the Fig. 2.

As shown in Fig. 2, a belief value π_i is defined for every user i that states the probability that channel C_i is in state h_i . The belief value π_i is updated via the feedback from user i at the end of each transmission period. The scheduler picks users in a manner to maximize the total throughput of the system. The authors of [31] state that this scheduling problem is a

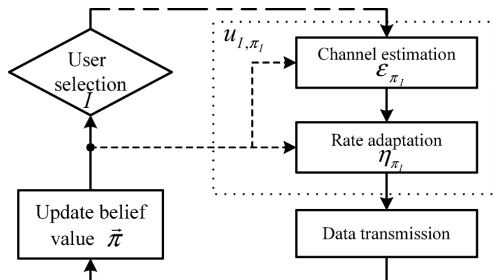


Fig. 2: Opportunistic scheduling with joint channel estimation and scheduling [31].

restless multi-armed bandit process (RMBP) [32], [33]. They use the Whittle's index policy to solve the RMBP problem. If the Whittle's index policy exists, the scheduler has near optimal performance. Using numerical evaluation, they show that the indexing policy shows near optimal performance and provide more than 90% gain over a randomized policy that schedules users randomly with uniform distribution.

In a realistic scenario, in order to make opportunistic decisions, the base station needs to acquire the channel information from users. In 3G, each user has a dedicated channel to feedback this information. On one hand, we can save the costly wireless bandwidth by reducing the amount of feedback channels. On the other hand, the opportunistic decisions are made based on the channel information of users that is obtained via these feedbacks. Therefore, again we encounter the classic *exploitation vs. exploration* trade-off.

Chaporkar *et al.* [34] address this trade-off and propose a scheduler with joint probing and transmission strategy. They show that throughput-optimal schedulers under full CSI assumption are also optimal when CSI is obtained via probing users. They used a Markov chain to model the queues, and the Lyapunov drift technique [35] to proof stability. The choice of optimum number of feedback requests is a generalized optimal stopping time problem [36] with an additional decision on choosing the users that should send their feedback first. The authors state that optimal solution to this problem can be found under special cases, hence they propose an approximate scheduling policy and probing and transmission strategy. The results of numerical simulations show the maximum system throughput using full CSI and probing under different probing time β . It can be observed that probing channel has significant impact on throughput. It is also observed that, if the number of users is large enough, there exists a threshold above which probing extra users is unnecessary.

Jacko studies the value of information in opportunistic scheduling in [37]. He models the wireless network using the Gilbert-Elliot model considering flow dynamics of the network, and proposes an opportunistic scheduler based on an optimal job (flow) indexing policy. Jacko shows that the channel information becomes valuable when queues are not fully backlogged. He also concludes that user's channel steady-state information is enough to make near optimal scheduling decisions.

Emerging cellular technologies such as LTE and WiMAX operate with multiple channels and multiple users. If every

user sends feedbacks for every available channel at the base station, a big chunk of bandwidth is dedicated to the feedbacks which is an unwanted load for the operators infrastructures. As mentioned earlier, these feedbacks are crucial for the base station to make opportunistic scheduling decisions that result in significant capacity gain in the network. Ouyang *et al.* [38] propose an algorithm to maintain the opportunistic gain while reducing the feedback overhead. They propose an opportunistic weight-based feedback allocation (WBF) scheme that provides a fraction of full throughput region under the MaxWeight scheduling, while limiting the number of feedbacks on every channel. WBF prioritizes users according to their queue size and transmission rates and it allocates more feedbacks to probe the channels of users with larger queues and higher transmission rates. It is shown via simulations that WBF can achieve almost identical results as compared to a scenario where full CSI is send to the base station.

C. Cognitive radio networks

This subsection covers the opportunistic schedulers proposed for cognitive radio networks under the capacity category. In a cognitive radio network, secondary users have transmission opportunity only if primary users are not transmitting. It is desirable to design a scheduling scheme that improves the service received by secondary users while minimizing the collision possibility between primary and secondary users. The cognitive network control scheduling algorithm (CNC) [25] maximizes the throughput of secondary users while bounding the maximum number of collisions with primary users. The authors of [25] consider a network with primary and secondary users. There is a dedicated channel for every primary user. CNC consists of a flow control policy and a resource allocation policy. The flow control policy takes into account the current backlog of secondary users to decide whether a new packet must be admitted to the queue or not. A virtual collision queue is introduced that monitors how much a primary user experiences collisions more than a predefined threshold. Using Lyapunov drift and Lyapunov optimization [35], the authors prove that CNC maintains a worst case bound for both backlog queue and number of collisions. The authors also propose to use CNC in a distributed manner by using greedy maximal match scheduling [39], [40]. The distributed implementation supports any rate within 50% of the capacity region. The simulations results [25] show that CNC can bound the total average congestion while delivering almost all the traffic.

In [26], Khalil *et al.* propose a cooperative scheduling scheme for cognitive radio networks. In a classic cognitive network, secondary users utilize the slots which are not used by primary users. In contrast, Khalil *et al.* consider a scenario in which secondary users in good channel state help primary users in bad channel to increase the channel capacity. The secondary users can be rewarded immediately or in the long term. Assume a primary user sends its data to a secondary user using a $(1 - \alpha)$ fraction of the time slot, $0 \leq \alpha \leq 1$. If the secondary user transmits the primary user's data using a fraction β of the time slot, $\alpha(1 - \beta)$ is the remaining time that is saved by relaying data to secondary user in good channel,

$0 \leq \beta \leq 1$. Fig. 3 is a depiction of the cooperative cognitive network operation proposed in [26]. With the immediate reward scheme, the secondary user transmits its own data during the remainder of the time slot. With the long term scheme, the scheduler guarantees to allocate a portion of resources saved by means of cooperation to secondary users. The stability of the proposed algorithm is proved using the Lyapunov drift and optimization techniques. The proposed scheduler is not optimal but it can be pushed towards optimality at the cost of longer average queues. Numerical results illustrate that Khalil *et al.*'s proposal improves the total network utility up to 5%, while providing secondary users with non-zero utility.

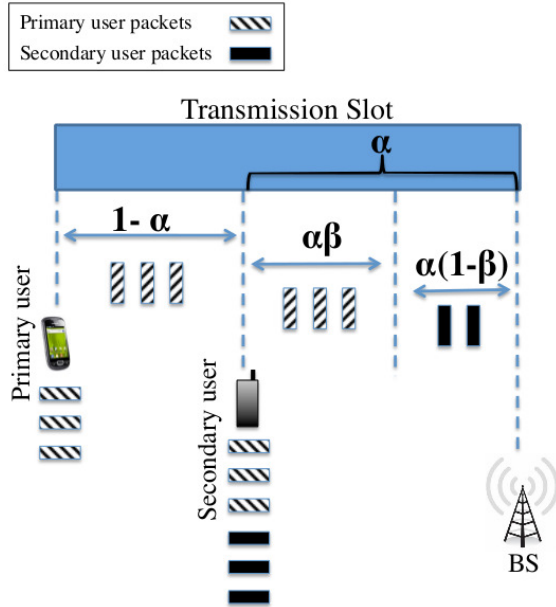


Fig. 3: Cooperative cognitive radio transmission slot structure. Primary user sends its data to secondary user in a fraction of $1 - \alpha$. Secondary user forwards this traffic in $\alpha\beta$ and in a fraction of $\alpha(1 - \beta)$, the secondary user sends its own traffic.

D. Summary

The authors of [19]–[21] focus on throughput optimal scheduler in presence of full CSI. The proposals in [19], [20] tries to adapt MaxWeight to multi-carrier scenarios. The algorithms proposed in [19] are based on generalization of MaxWeight with optimization problems for multi-carrier networks. Due to hardness of these optimization problems, the authors investigate on bounded approximations. CA-WS and hybrid CA-WS [20] are shown to be both throughput optimal, which resolve the weakness of MaxWeight in presence of flow dynamics.

In [28], [31], [34], the authors consider the effect of feedback on system performance. For instance users' CSI is predicted by a short observation period in WSL [28] or by using POMDP as in [31]. Chaporkar *et al.* propose a technique which reduces the number of feedback while maintaining throughput optimality.

In [25], authors exploit the concept of virtual collision queues to reduce the number collisions between secondary and primary users. Khalil *et al.* proposes a scheduling scheme that

increases the total system throughput by exploiting secondary users with good channel conditions [26]. Table III shows each proposal mainly focusing on capacity with details regarding the assumptions taken by the authors, analytical tools used for the proposals, the scenario in which the proposal is applicable, and other considerations taken into account besides capacity improvement.

IV. QUALITY OF SERVICE

With the recent advent of applications such as VoIP and video conferencing, quality of service (QoS) gained popularity in both research and industry. There are several QoS objectives defined such as throughput, delay, jitter, packet loss, error rate, latency and so on. Among the QoS metrics, the opportunistic scheduling proposals pay more attention to delay and throughput as depicted in Fig. 1. Some of the proposed schedulers may adopt a single QoS objective [41]–[43] while others use multiple objectives [44], [45]. In the following, we review the proposals with single QoS objective. Then we continue with proposals taking multiple QoS objectives into account.

A. Single QoS objective

In this subsection, we focus on proposals with one single QoS objective. Among available QoS objectives, delay and average throughput are the most common used.

In [41], Kim and de Veciana investigate the performance of opportunistic scheduling with heterogenous traffic (i.e., QoS and best effort (BE) flows). They show that traffic integration—i.e., the coexistence of QoS and BE traffic in the same network—deteriorates the performance of the system in terms of capacity, stability, and delay. This *performance anomaly* was previously dealt with at packet-level in [46]–[48].

Kim and de Veciana [41] studied the interaction of QoS and BE traffic at flow-level for the first time. They find necessary and sufficient stability conditions for the traffic integration models that was previously provided using a 2-dimensional Markov chain in [49]–[51]. The proposed opportunistic scheduler is designed in a way that QoS flows receive a fixed average throughput per slot. Other QoS objectives such as delay or jitter are not considered. The authors argue that allocating an average throughput to QoS flows in every slot reduces the chance of starvation in the long period. BE traffic is modeled as finite file transfers using HTTP or FTP and its performance is evaluated through the average time needed to complete a file transfer.

Additionally, the authors of [41] propose an opportunistic scheduling scheme that monitors the number of QoS and BE flows. In order to be able to guarantee the fixed average throughput in every slot, the maximum number of QoS flows is limited such that the total promised bandwidth remains less than the total available bandwidth. It should be noted that average channel quality of users affects the total capacity of the network and the maximum number of QoS flows. Kim and de Veciana also propose a bandwidth borrowing/lending scheme that allows QoS services to borrow bandwidth from

TABLE III: Summary of proposals with main focus on *capacity*

Proposal	Assumptions	Analytical tools	Topology	Other focus
Scheduler for multicarrier wireless system [19]	Full CSI Not fully backlogged Traffic: Generic	Greedy algorithms Dynamic programming Linear programming Lyapunov drift	Single cell (OFDMA) Multi-carrier Downlink	Fairness Stability
Scheduler for multichannel wireless systems with flow-level dynamics [20]	Full CSI Not fully backlogged Traffic: Generic	Markov chain Lagrangian multipliers Lyapunov drift	Single cell (OFDM) Multi-carrier Downlink	Throughput-optimal Stability
Optimal scheduler for HSDPA networks [21]	Full CSI Not fully backlogged Error free transmission Traffic: Bernoulli	Finite state Markov chain Dynamic programming Markov decision process	Single cell (HSDPA) Downlink	Throughput Delay Fairness
Throughput-optimal scheduler that accounts for flow-level dynamics [28]	Non-full CSI Not fully backlogged Traffic: Flow level with two class of flows	Markov chain Lagrangian multipliers Lyapunov drift	Single cell Downlink	Throughput-optimal Stability
Joint channel estimation and scheduler for wireless networks [31]	Non-full CSI	Markov chain Restless multi-armed bandit process Partially observable Markov decision process	Single cell Downlink	
Throughput-optimal scheduling with limited channel information [34]	Non-full CSI Not fully backlogged Traffic: Generic	Markov decision process Lyapunov drift Optimal stopping theory	Single cell Downlink	Throughput-optimal Stability
Optimal feedback allocation in multichannel wireless networks [38]	Non-full CSI Not fully backlogged Traffic: Poisson	Markov chain	Single cell (FDD) Downlink Multi channel	Near throughput-optimal Stability
Flow-level scheduler for wireless networks [37]	Non-full CSI Not fully backlogged Traffic: Bernoulli	Markov decision process Gilbert-Elliot model Lagrangian multipliers Dynamic programming	Single cell Downlink	Stability
Opportunistic scheduler for cognitive radio networks [25]	Non-full CSI Not fully backlogged Traffic: Bernoulli	Markov chain Lyapunov drift Lyapunov optimization Maximum weight match	Multi-cell	Throughput-optimal
Optimal scheduler for cooperative cognitive radio networks [26]	Full CSI Fully backlogged	Lyapunov drift Lyapunov optimization	Single cell	Stability

BE services when required. Therefore, each QoS flow borrows bandwidth from BE flows to maintain the promised average throughput. Similarly, QoS flows can borrow their extra bandwidth to BE flows. They show that integration of QoS and BE flows reduces the system capacity and leads to the so called *loss in opportunism* phenomenon (33% capacity reduction in the example provided in [41]). This loss is due to QoS requirements of flows, which forces the opportunistic scheduler to transmit packets of QoS flows, although that was not the *opportunistic* choice at that moment. QoS flows also affect the delay experienced by BE flows. This effect magnifies with lower SNR, higher guaranteed bandwidth, and larger number of QoS flows. If QoS flows remain in the system for a long time, BE flows are under-served until they have a chance to recover, i.e., QoS sessions leave the system. This phenomenon is called *local instability* which is caused by the coexistence of QoS and BE flows. Kim and de Veciana propose a call admission control (CAC) for BE flows which solves the local instability issue. Using numerical evaluation, they show that using CAC reduces the local instability and improves the delay for BE flows.

In [42], Sadiq *et al.* focus on the performance and robustness of packet schedulers in wireless networks. They prove that mean-delay-optimal policies show radial sum-rate-monotonicity (RSM) meaning that when users' queue grow linearly, the scheduler serves users in a way that de-emphasizes

queue balancing in order to maximize system throughput. This behavior is in contrast with the behavior of known throughput optimal schedulers (e.g., MaxWeight [13], Exp rule [15]). In that paper, *Log rule* policy is introduced as a new class of policies that are RSM and throughput optimal. The authors use a Markov chain to obtain the queue state transition probabilities which are used to minimize the average queue length of users. Sadiq *et al.* also evaluate the performance of Log rule, MaxWeight, and Exp rule schedulers via simulation. They show that Log rule and Exp rule exhibit comparable delay performance and perform better than MaxWeight under low traffic. In a high traffic scenario, users experience 20-80% less delay under Log rule, see Fig. 4. It is also shown that Log rule degrades more gracefully than the Exp rule and MaxWeight under high traffic load. In other words, when the system is overloaded more users meet their QoS requirements under Log rule in comparison to Exp rule and MaxWeight.

Some schedulers are designed to perform under worst case scenarios while others do not consider any specific network conditions. Sadiq *et al.* conclude that the schedulers which are optimized for overall system performance are more likely to be robust to changes than those optimized for the worst case.

In [43], Neely proposes an opportunistic scheduling algorithm with delay guarantees. He develops a novel virtual queue technique (i.e., the ϵ -persistent service queue) which guarantees a worst case delay for each users. He further

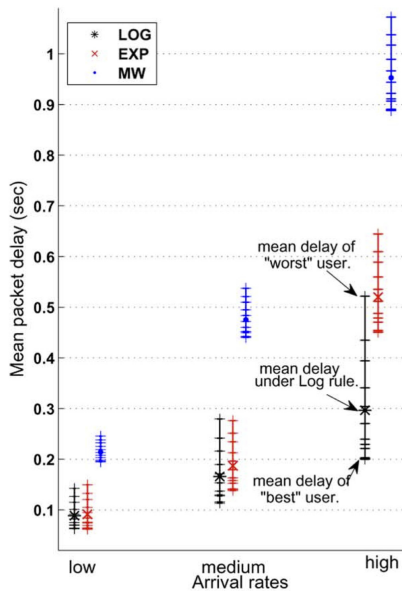


Fig. 4: Performance comparison among Log rule, EXP rule and MaxWeight(MW) scheduling algorithm. Mean packet delays are shown under three different arrival rate, namely low, medium and high [42].

uses Lyapunov drift and optimization techniques to obtain a throughput-optimal scheduling algorithm that guarantees bounded worst-case delay. The proposed scheduler is compatible with both ergodic and possibly non-ergodic channel and arrival settings. Moreover, it can be used for both single-hop and multi-hop scenarios. Finally, the author proves that the performance of the proposed algorithm is comparable to schedulers that have advance knowledge of channel variations (i.e., full CSI).

In [37], that was briefly discussed in the previous section, Jacko proposes an opportunistic scheduling scheme using a Whittle indexing policy that minimizes the transmission delay. Jacko's proposal is also throughput-optimal under not fully backlogged conditions.

B. Multiple QoS objectives

In this subsection, we list the proposals that intend to meet more than one QoS metric.

In opportunistic scheduling, it is common to observe that users with low channel quality frequently experience transmission rate fluctuations. These fluctuations result in larger queues and longer delays. Choi *et al.* proposed the AADTR metric to be able to measure and control these fluctuations and their resulting delays [44]. The algorithm proposed by Choi *et al.* in [44] targets orthogonal frequency division multiple access (OFDMA) wireless networks in which users can transmit over different subcarriers at the same time. Their proposal maximizes system throughput while meeting the *required average transmission rate* and the *average absolute deviation of transmission rate* (AADTR). The latter is a metric to control the transmission rate fluctuations. QoS flows have both average transmission rate and AADTR objectives. Average transmission rate is the only objective for BE flows. The proposal addresses both real time (i.e., video conferencing) and

BE traffic. The authors formulate the problem of scheduling in the OFDMA wireless communications which can be solved using the dual optimization technique [52]. The proposed algorithm calculates the optimal solution for every frame which guarantees average throughput with bounded fluctuations over time. The proposal performance is illustrated using computer simulations in both stationary and non-stationary channel conditions. In the simulations, it is assumed that the queues are fully backlogged and users move with the speed of 50 km/h invariably. Results show that the throughput of the proposed algorithm is on average 30% higher than that of Modified Largest Weighted Delay First (M-LWDF) [53]. M-LWDF is a heuristic that was originally designed for TDMA systems. It selects users based on a simple metric, taking into account both the current channel state and the head-of-line packet delay. Unlike M-LWDF, packet drop rate of the AADTR-based algorithm remains the same with increasing number of users.

In [45], throughput, delay and packet drop are the adopted QoS objectives of the scheduler. The authors map the QoS objectives into a unity cube that represents the QoS state of the flows, see Fig. 5. The delay and packet drop incurred by each flow is normalized over the flow's QoS requirements. The achieved throughput is, on the other hand, normalized over the *desired* throughput, and the inverse of the normalized throughput is used as third QoS dimension. Hence, a normalized value equal or less than one indicates that the flow is satisfied under that QoS objective. A flow that is mapped inside the unity cube is satisfied for all its QoS objectives.

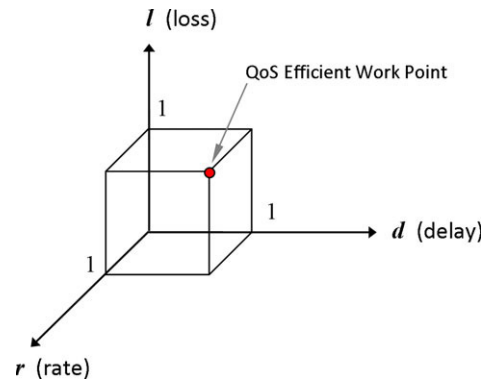


Fig. 5: QoS unity cube [45].

The QoS efficient work point is at (1,1,1) and the QoS-deviation metric is defined based on the distance from the current position of the QoS descriptor of the flow in the cube to the efficient work point. A negative value represents an unsatisfied flow and a positive value represents a satisfied flow. The authors propose three schedulers, namely, biggest QoS-deviation first (BQDF), adaptive QoS-deviation control (AQDC), and adaptive residual time control (ARTC). BQDF chooses the user with the highest deviation that corresponds to the least satisfied flow. BQDF causes throughput degradation because of ignoring channel condition of the flows. AQDC is proposed to fix the biased behavior of BQDF. AQDC builds a set which includes the $p\%$ of flows with highest QoS-deviation value, i.e., the flows that do not meet their QoS requirements.

Next, the flow with maximum transmission rate is chosen to be scheduled. If none of the flows in the set can transmit due to deep fading, then p is increased and the algorithm builds a new set of eligible users for transmission. The size of the set can be tuned to adjust the trade-off between QoS provisioning and throughput (i.e., the larger is the set, the higher is the diversity gain). ARTC uses the same approach as AQDC for scheduling flows from a chosen set. ARTC is different from AQDC in the sense that it builds the eligible user set based on the *residual time* obtained for the three adopted QoS objectives. *Residual time* is the time that a flow can await before transmission without violating its QoS requirements. The *residual time* was originally employed in Urgency Feasibility scheduler [54]. Urgency feasibility scheduler computes the residual time for all flows and serves the flow with lowest residual time. Furthermore, a call admission control policy is proposed to improve the QoS. A flow is rejected if the percentage of unsatisfied flows exceeds the admission percentage. In the end, the performance of BQDF, AQDC, ARTC, Urgency Feasibility, and MaxRate schedulers are evaluated via simulation using NS2 [55] under various scenarios. It is shown that AQDC and ARTC perform best in terms of maintaining QoS objectives. In terms of packet drops, AQDC and ARTC have more than 20% less packet drop than others. AQDC and ARTC achieve the same throughput as MaxRate which is more than 50% higher than Urgency Feasibility and BQDF. In conclusion, the schedulers that use the newly introduced QoS parameter (i.e., QoS deviation) outperform those which operate on a single QoS objective.

C. Summary

The authors of [37], [41]–[43] propose schedulers with only one QoS metric. Kim *et al.* propose a scheduler to avoid the performance anomaly caused by traffic integration [41]. The works in [37], [42], [43] are dedicated to delay improvement. Authors of [42] propose a log rule-based scheduler to achieve delay optimality. Neely uses Lyapunov drift and optimization techniques to propose a throughput-optimal scheduler with worst case delay guarantees and without any prior knowledge of channel conditions [43].

The schedulers in [44], [45] can support multiple QoS metrics. In [44], authors propose a scheduler for real-time and best effort services using the dual optimization technique. Authors in [45] propose schedulers which operate based on the *position* of the flow in the QoS unity cube or based on the *QoS residual time*. Table IV shows each proposal mainly focusing on QoS with details regarding the assumptions taken by the authors, analytical tools used for the proposal, the scenario in which the proposal is applicable, and other considerations taken into account besides QoS.

V. FAIRNESS

Due to the greedy behavior of opportunistic schedulers, their fairness performance is always a concern. Scheduling users opportunistically can result in under-serving some users due to their poor channel quality, while the rest are over-served because they are in a better channel conditions. As a result,

it is essential to monitor the way a scheduler allocates the resources to avoid unfairness among users in the long term.

There are different metrics defined for fairness (e.g., Jain's index, temporal fairness, utilitarian fairness). Jain's index is one of the popular fairness metrics for studying fairness performance of the schedulers. For a given set $X = \{x_1, x_2, \dots, x_n\}$ Jain's index is computed as follows [64]:

$$\text{Jain's index} = \frac{\left(\sum_i^n x_i\right)^2}{n \sum_i^n x_i^2}.$$

In [56], authors introduce optimal policies for opportunistic scheduling in OFDM systems with three different fairness criteria, namely *temporal fairness*, *utilitarian fairness*, and *minimum-performance guarantees*. Under temporal fairness criteria, all users are given at least a certain share of airtime, where under utilitarian fairness criterion users are given a certain share of throughput [57]. The policies with minimum-performance guarantees, as the name implies, aim to maximize the network performance while satisfying minimum user requirements. Temporal and utilitarian fairness methods oblige the scheduler to allocate a predefined share of resources (i.e., time, throughput) to every user. In contrast, with minimum-performance guarantees the scheduler is restrained to satisfy the minimum service requirement of the users. The authors [56] interpret the optimal policies as bipartite matching problem and solve them using Hungarian algorithms [58]. Simulation results shows that temporal, utilitarian, and minimum-performance guarantee policies provide 46%, 32%, and 31% gain over Round Robin, respectively.

One of the most diffused opportunistic approaches with fairness constraints is the proportional fair scheduler [6], [59]. This scheduler assigns priorities to users based on the ratio of two functions: the first function accounts for the rate potentially achievable in the current slot, while the second function accounts for historical average of the user's throughput.

In [60], Tsai introduces four scheduling algorithms that are based on proportional fair (PF) algorithms. All four proposed schedulers exploit user and channel diversity to improve the system performance. PF scheduler prioritizes users based on a balance between the current achievable rate and fairness [10], [61], [62]. The following illustrates the PF scheduler general scheduling policy and throughput monitoring [10]:

$$i^*(t) = \arg \max_{i \in N_b} \frac{R_i(t)}{T_i(t)}; \quad (1)$$

$$T_i(t+1) = \left(1 - \frac{1}{t_c}\right) T_i(t) + \frac{R_i(t)}{t_c} I_{i=i^*(t)}; \quad (2)$$

where $i^*(t)$ is the user scheduled at time t , N_b is the set of users, $R_i(t)$ and $T_i(t)$ are the peak feasible data rate and monitored throughput of user i at time slot t , $I_{\{\cdot\}}$ is the indicator function and t_c is an averaging time window. In the PF scheduler, it is assumed that the number of users is fixed (stationary scenario) and queues are fully backlogged. These assumptions are not realistic and variation in number of users and their queue length has negative impacts on system

TABLE IV: Summary of proposals with main focus on *QoS*

Proposal	Assumptions	Analytical tools	Topology	Other focus
Flow-level scheduler for wireless networks [37]	Non-full CSI Not fully backlogged Traffic: Bernoulli	Markov decision process Gilbert-Elliot model Lagrangian multipliers Dynamic programming	Single cell Downlink	Stability
Scheduler for wireless systems with integrated traffic [41]	Not fully backlogged Traffic: Poisson	Markov chain Lyapunov drift Foster theorem	Single cell (TDMA)	Stability
Delay-optimal Log rule based scheduler for wireless networks [42]	Not fully backlogged Traffic: Poisson	Markov decision process Dynamic programming Lyapunov drift Foster theorem	Single cell (HDR) Downlink	Throughput-optimal Delay-optimal
Opportunistic scheduler for wireless networks with worst-case guarantee [43]	No CSI Not fully backlogged	Lyapunov drift Lyapunov optimization	Single/multi-hop	
Scheduler for OFDMA systems with multimedia support [44]	Fully backlogged	Duality theory Lagrangian multipliers Convex optimization	Single cell (OFDMA) Downlink	
Adaptive QoS scheduler for wireless networks [45]	Not fully backlogged Traffic: Poisson & Pareto	Unity cube mapping	Single cell Downlink	

performance. As a remedy, Tsai [60] proposes to stop updating $T_i(t)$ for idle users which avoids giving idle users higher priority upon their next transmission. If $T_i(t)$ is updated for idle users, upon their activation, the scheduler assumes that they had packets to transmit within last few slots and they should be given higher priority to maintain fairness. The four proposed schedulers account for the dynamic changes in number of active users and queue length. These schedulers employ different methods with different level of sophistication to update users' $T_i(t)$. The first two algorithms only update $T_i(t)$ for active users and their difference is only the initial value of $T_i(t)$. The third algorithm uses a more sophisticated approach by updating $T_i(t)$ according to the status of the user in the previous slot (i.e. active or idle) and the last time it is scheduled. Finally, the fourth algorithm changes the averaging window size t_c with reference to the backlog of the users. Results show that proposed algorithms reduce the impact of dynamic changes in the network on packet delay and data rate.

The authors of [63] adapt the analytical model proposed for PF by Liu *et al.* [6] to an OFDMA networks with more realistic assumptions. Their model accounts for multiple subcarriers, but also for less realistic Poisson traffic arrivals. The adapted PF scheduler computes a matrix containing user rankings over all subcarriers. For every subcarrier, the user with the highest rank and *non-empty buffer* is scheduled. Non-empty buffer condition accounts for the fact that in real world a users can be eligible to be scheduled when it has no packets to transmit. In such cases, the scheduler selects the next best user with non-empty buffer to avoid wasting airtime. The proposed analytical model is validated in terms of average throughput and Jain's fairness index by simulation.

In [65], an adaptive resource allocation for OFDM system is proposed that accounts for each user's required data rate as a fairness measure. The authors formulate the optimization problem for subchannel and power allocation with a proportional fairness constraint. Since their proposed optimization problem requires linearization of nonlinear constraints, the authors propose a suboptimal solution with lower complexity. The suboptimal solution carries out the subchannel allocation

and power allocation separately. Via simulations, it is shown that the suboptimal solution can achieve 95% of the optimal capacity with much lower complexity.

In [66], Kwon *et al.* tackle the fairness issue in TDMA networks. The proposed opportunistic fair scheduler prior to [66] are either based on average rate-based utility functions or instantaneous rate-based utility functions. Average rate-based utility functions are suitable for elastic services (e.g., HTTP, Email, and FTP) for which instantaneous data rate does not affect the QoS. On the contrary, satisfaction of services such as video streaming depends on the instantaneous data rate. For such services, the utility function should be based on instantaneous rate. Kwon *et al.* propose a framework that can accommodate both elastic and non-elastic services. Instead of deterministic scheduling, they use a probabilistic scheduling policy that randomly schedules a user per time-slot with a certain probability. Kwon *et al.* model the channel using a finite state Markov chain (FSMC), formulate the scheduling problem based on convex optimization [67] and solve it using a Lagrangian function, the duality theorem [68]. An iterative algorithm is also proposed that can find the optimum solution in every time slot. It is shown via numerical simulation that their proposed scheduler meets the required fairness objective for users with elastic and non-elastic services.

A. Summary

The surveyed literature in this section considered fairness in opportunistic schedulers. As pointed out in [56], different metrics can be used to evaluate the fairness in a transmission system, e.g., *temporal fairness*, *utilitarian fairness*, and *minimum-performance guarantees*. Proportional fair schedulers attempt to balance between users with the best channel state, and users which received less throughput in the past [6]. Tsai, in [60], and Almatarneh *et al.*, in [63], point out the importance of throughput monitoring with proportional fair schedulers. Tsai proposes four variants of the original proportional fair algorithm that accounts for parameters such as users activity log and backlog. Kwon *et al.* [66] propose a fair opportunistic scheduler based on probabilistic scheduling approach using

convex optimization and dual theory. Table V shows each proposal mainly focusing on fairness with details regarding the assumptions taken by the authors, analytical tools used for the proposal, the scenario in which the proposal is applicable, and other considerations taken into account besides fairness.

VI. DISTRIBUTED SCHEDULING ALGORITHMS

In a centralized scheduling approach, the scheduler is aware of all user's channel condition or it will acquire an estimate of that information to make the scheduling decision. On the other hand, in a distributed scheduling approach, users make scheduling decisions independent of the central entity and possibly without an overall knowledge of the network.

Distributed opportunistic scheduling (DOS) did not get much attention until recent years. The authors of [70] and [71] took the first steps to study such systems under various scenarios. In [70], Zheng *et al.* study DOS in ad-hoc networks with random channel access under two scenarios: 1) all users cooperate to maximize the aggregate network throughput (i.e., cooperative); and 2) each user tries to maximize its own throughput (i.e., non-cooperative). In the first scenario (cooperative), all users contend for the channel but successful contention is not necessarily followed by transmission. Upon successful contention, the user transmits only if it is in good channel condition (i.e., high data rate), otherwise it allows the rest of the users to re-contend for the channel. This process goes on until a user in good channel condition wins the contention. The cooperation benefits all users because it leads to network capacity increment in the long run. Like any other opportunistic approach, the cooperation gain is depending on the channel diversity of the users. It can be seen that there is a trade-off between the gain obtained from probing more users and the time it takes to find a user with high rate. In Fig. 6, we can see that total available time T_{total} is divided between n contention period T_n and data transmission period T_{tran} .

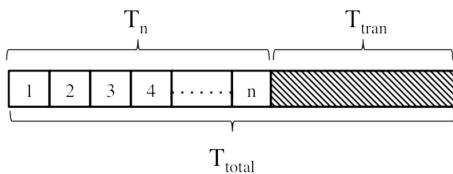


Fig. 6: DOS transmission example.

Clearly, the channel transmission time is reduced when more users probe the channel. Hence, there is a threshold beyond which, re-contending for channel becomes sub-optimal. The question is: *what is the optimal transmission rate so that a user does not need to drop its transmission chance and allow other users to re-contend for channel?* The authors propose an online iterative algorithm that can compute the optimal rate threshold based on the local information of each user. Next, they show that it is optimal to transmit when the current transmission rate is higher than an optimal threshold.

The second scenario analyzed in [70] is a non-cooperative game with selfish users trying to increase their throughput. Zheng *et al.* use game theory to model the behavior of users in this network. The authors investigate the existence

and uniqueness of the Nash equilibrium [72] using the best response strategy. For both scenarios (i.e., cooperative and non-cooperative), the optimal rate can be computed via an iterative algorithm based on the local information. Using local information alleviates the need for message exchange between users for reporting their CSI. Numerical simulations indicate that applying optimal stopping theory results in higher throughput gains. It is also shown that the non-cooperative scheme is less efficient than the cooperative scheme due to selfish behavior of users. However, the inefficiency of the non-cooperative scheme can be mitigated by forcing a price-based mechanism.

The DOS proposed in [71] is an extension to the work in [70] with the addition of average delay constraint. Hence, the proposed algorithm should maximize the throughput for both cooperative and non-cooperative scenarios mentioned in [70], while maintaining the designated delay constraints. The authors consider both network-wide and per-user average delay constraints. As for the network-wide average delay, they seek for the optimal transmission rate value in which all users cooperate to increase the total network capacity while maintaining the average delay less than α . In the case of individual user average delay constraint, that is defined for the non-cooperative scenario, each user tries to maximize its throughput while keeping its average delay less than α_m . In order to find the optimum threshold, Tan *et al.* again employ optimal stopping theory to formulate the problem and solve it using a stochastic Lagrangian approach. The results show that the optimal transmission rate threshold is upper bounded by a function of α if the average delay is less than a critical time ($\alpha < \alpha^*$). In other words, the average delay constraint does not affect the optimal threshold policy if ($\alpha > \alpha^*$); otherwise the optimal threshold policy will be a function of α . This result also applies to the non-cooperative scenario, where there exists an α_i^* for every user i . The authors evaluate the performance of the proposed DOS using numerical simulations which indicate that higher values of SNR reduce the effect of α on the optimal policy.

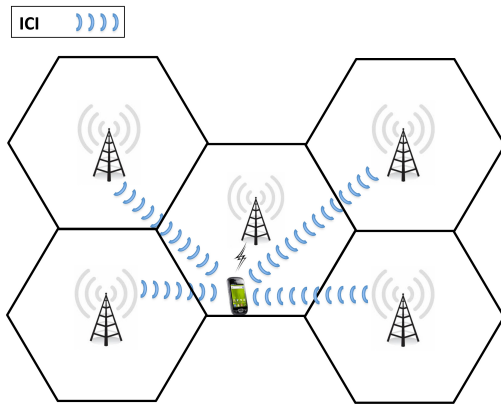
The authors of [73] propose a two-level probing for the DOS proposed in [70]. In practice, the channel estimation obtained from the first time probing is not very accurate due to noise. Hence, the authors [73] propose a two-level probing policy which appears to be a threshold-based policy. If the first channel estimation falls in between thresholds, then it is optimal to probe the channel for the second time. It is shown via simulation that the two-level probing policy can increase the gain over one-level probing policy up to 110%.

Garcia-Saaverdra *et al.*, in [74], address the shortcomings of previous DOS proposals such as [70], [71], [73] and propose an adaptive distributed opportunistic scheduling (ADOS). Previous DOS proposals were designed for a scenario with fully saturated users (i.e., fully backlogged user's transmission queues). In a real networks users have different queue sizes, in fact there are situations when user contends for the channel but it has not enough data to transmit during the whole contention period. Hence, there are periods of time in which no one is transmitting. These non-transmission periods are called *empty mini-slots* in [74]. This is the drawback of DOS

TABLE V: Summary of proposals with main focus on *fairness*

Proposal	Assumptions	Analytical tools	Topology	Other focus
Proportional fair scheduler for wireless forward link data services [60]	Full CSI Not fully backlogged Traffic: Poisson		Single cell (TDM) Downlink	
Proportional fair scheduler for wireless forward link data services [66]	Full CSI Not fully backlogged Traffic: Poisson	Convex optimization Duality theory Lagrangian multipliers	Single cell (TDMA)	Jain's fairness Index
Analytical model for proportional fair scheduler in OFDMA systems [63]	Not fully backlogged		Single cell (OFDMA)	Jain's fairness Index
Opportunistic schedulers optimized under different fairness consideration [56]	Full CSI Fully backlogged	Maximal bipartite matching Hungarian algorithms	Single cell (OFDM) Downlink	Temporal fairness Utilitarian fairness Minimum performance guarantee
Adaptive opportunistic scheduler for Multiuser OFDM Networks [65]	Full CSI	Linear programming Non-linear programming Newton-Raphson method [69]	Single cell (OFDM)	

which makes it sub-optimal for unsaturated scenarios. ADOS overcome the aforementioned drawback by controlling both the optimal transmission rate threshold (i.e., the rate at which users transmit after a successful contention) and the access probability (i.e., the probability that a user attempts to contend for the channel). The authors exploit access probability to discriminate among users with large backlog and those with small backlog. Hence, the probability of empty mini-slot occurrence is reduced by assigning higher access probabilities to users with large backlog. The optimal rate and access probabilities are computed locally via a simple algorithm based on control theory that adapts itself to the current state of the network and converges to the optimal operation point. The drawback of ADOS can be the convergence time of the adaptive algorithm. Using simulations, the authors show that ADOS outperforms previous DOS proposals in terms of proportional fairness, and achieves 30% higher throughput.

**Fig. 7:** Inter-Cell interference in cellular networks.

All previously reviewed distributed approaches suit carrier sense multiple access (CSMA). However, distributed scheduling algorithms can also be used in multi-cell OFDMA scheduling. In a real network, apart from environmental noise, mobile users are also affected by the interference from neighboring cells, see Fig. 7. Indeed, it has been shown that Inter-Cell interference (ICI) can be reduced by exchanging information between base stations using multi-cell scheduling [75]–[82].

In the multi-cell scheduling approach, every base station first computes the ICI from the signal measurements sent by

its users, and then exchanges this information with neighboring base stations. A drawback of this approach is that a high speed connection should be allocated to inter-base station communications. Another drawback is the high computation overhead on the central entity that processes this information for further scheduling decisions (EPC in LTE [2]).

Tang *et al.* [75] propose two joint multi-cell scheduling and beam coordination schemes, namely *SINR feedback* and *ABC*. In the *SINR feedback* scheme, each base station selects m beams randomly and sends beam-pilots within the cell. The users will send their feedback with respect to their SINR which are also affected by the beam pilots received from the neighboring cells. Each base station makes scheduling decisions based on the received SINR feedbacks. Intra-cell scheduling decisions are made based on the PF algorithm proposed in [83].

In the *ABC* scheme, the cellular network is partitioned dividing base stations into A/B/C subsets. Although the network is partitioned, it operates as a reuse factor 1. In the first step, base stations tagged as *A* select m beams, send beam pilots and make their scheduling decision. Note that users of subset B and C also listen to beam pilots to be able to estimate their SINR. After the scheduling decision has been made, the identity of the scheduled users and the beams assigned to them is sent to the neighboring base stations tagged as subset B. This helps the base stations in subset B to avoid choosing beams that interfere with subset A. Base station in subset B will perform the same operation and inform to those in subset C. Due to priority (in term of frequency selection) given to the base stations in subset A, they will provide better service to their users in comparison to the base stations in subset B and C. To avoid such unfairness, the authors propose to assign the A/B/C tags in a Round Robin fashion.

In their work, Tang *et al.* compare the proposed schemes with two other schemes. In the first scheme, network operates using a frequency reuse partition with a factor 3. In the second scheme, each base station operates fully independently and without considering ICI. Using simulations, it is shown that *SINR feedback* and *ABC* outperform the two other schemes. *ABC* scheme provides more than 100% throughput gain in comparison to schemes without ICI control mechanism. In addition, users on the edge of the cell receive higher throughput

with *ABC* scheme.

In [76], Bendlin *et al.* propose a distributed multi-cell scheduling, namely cooperative eigen beamforming (CEB), that is tolerant to delay and capacity limitations of backhaul links. In many proposals, authors assume that backhaul links have zero delay and unlimited capacity, which is not a realistic assumption. In CEB, authors assume that each base station schedules one user per slot and it has full CSI knowledge of its users. Scheduling is performed in two steps in CEB. In the first step, each base station chooses the proper beamforming which minimizes ICI and in the second step, a user will be scheduled. The main advantage of CEB is the low amount of data exchanged among base stations. Bendlin *et al.* show that CEB can perform very close to schemes that disseminate the full CSI information in the network. CEB also exhibits robustness towards delay in backhaul links.

A. Summary

In this section we have reviewed proposals focusing on distributed opportunistic scheduling strategies. The body of work [70], [71], [73], [74] investigates distributed scheduling for wireless networks with random medium access (e.g., CSMA). In [70], [71] a distributed scheduler is proposed, which uses optimal stopping theory to make transmission decisions according to the channel state. The scheduler proposed in [71] includes delay constraints in addition to throughput. In [74] a distributed algorithm based on control theory is proposed that improves the proposals in [70], [71], [73] by considering dynamic traffic conditions.

Proposals in [75], [76] focus on distributed scheduling in cellular networks. Tang *et al.* [75] propose two approaches, namely *SINR feedback* and *ABC*, to reduce the ICI in a multi-cell scenario. Bendlin *et al.* [76] use *cooperative eigen beamforming* for ICI reduction. Table VI shows each proposal focusing on distributed scheduling with details regarding the assumptions taken by the authors, analytical tools used for the proposal, the scenario in which the proposal is applicable, and other considerations taken into account.

VII. SURVEY'S SUMMARY

In this section, we conclude the work with summarizing the common techniques, evaluation methods, and possible future directions of research in opportunistic scheduling algorithms. Moreover, the proposals under different categories are summarized in Table VIII. We also provide a performance comparison among different schemes based on the results provided in the reviewed papers.

A. Common techniques

Markov chain is commonly used in many papers, especially for modeling the channel variations in wireless networks [20], [21], [25], [28], [31], [34], [37], [41], [42], [66], [95]. The Lyapunov drift and optimization techniques are widely employed to prove the stability of the schedulers or to optimize them [19], [20], [25], [26], [28], [34], [41]–[43]. Many proposals exploit other mathematical techniques such

TABLE VII: Common techniques used in opportunistic scheduling

Category	Analytical tools
Capacity	- Markov chain [20], [21], [25], [28], [31], [37], [38] - Lyapunov [19], [20], [25], [26], [28], [34] - Dynamic programming [19], [21], [37] - Optimal stopping theory [34] - Greedy algorithms [19] - Restless multi-armed bandit [31], [37] - Lagrangian multiplier [19], [20], [28], [37]
QoS	- Markov chain [41], [42] - Lyapunov [41]–[43] - Dynamic programming [42] - Duality theory [44] - Lagrangian multiplier [44]
Fairness	- Markov chain [66] - Duality theory [66] - Hungarian algorithms [56]
Distributed	- Optimal stopping theory [70], [71], [73], [74] - Nash equilibrium [70], [74] - Control theory [74] - Duality theory [71] - Game theory [70], [71]

as duality theory [44], [66], optimal stopping theory [34], [70], [71], [74] and restless multi-armed bandit [31], [37] for optimization purposes. It was of surprise that the common optimization techniques (e.g., linear programming, non-linear programming) are rarely used in recent works. Table VII illustrates the popularity of each technique in every category of opportunistic scheduling, according to the taxonomy we defined in Section II.

B. Evaluation Method

The proposed opportunistic algorithms can be evaluated via numerical calculation obtained from analysis and simulations using available network simulators (e.g., NS2 [55], NS3 [97], OPNET [98], etc.). The nature of opportunistic schedulers requires cross-layer information that is not always available at MAC layer. The time and effort required to adapt a new algorithm into the current protocols makes implementation less popular than numerical results. Hence, most of the proposals are supported by mathematical analysis and simulation via ad-hoc simulators. As such, the evaluation does not consider the impact of the proposed schedulers on other layers (e.g., application, network, and transport) and vice versa.

C. Open issues

So far we have reviewed opportunistic scheduling from various aspects. In this subsection, we highlight the open research issues in opportunistic scheduling.

1) *Optimality under realistic assumptions:* There are many proposals in opportunistic scheduling that are optimal under certain assumptions. Authors of [17], [99] focus on this fact in OFDMA systems by showing the real world constraint such as limited computational capacity of base stations and its affect on scalability of the network. The following is a list of most common assumptions:

- Full CSI availability of mobile users' channel at base station.
- Fully or infinitely backlogged queues.
- Mobile users with i.i.d. channel distributions.

TABLE VI: Summary of proposals with *distributed scheduling*

Proposal	Assumptions	Analytical tools	Topology	Other focus
Distributed opportunistic scheduler for ad-hoc networks [70]	Non-full CSI Fully backlogged	Optimal stopping theory Nash equilibrium Game theory	Single hop (ad-hoc) Random access	Throughput Fairness
Distributed opportunistic scheduler for ad-hoc networks under delay constraints [71]	Non-full CSI Fully backlogged Game theory	Optimal stopping theory Duality theory Nash equilibrium Stochastic Lagrangian approach	Single hop (ad-hoc) Random access	Throughput Fairness
Distributed opportunistic scheduler for ad-hoc networks with two-level probing [73]	Non-full CSI Fully backlogged	Optimal stopping theory Lebesgue's convergence theorem [84]	Single hop (ad-hoc) Random access	Throughput
Distributed opportunistic scheduler for ad-hoc networks using control theory [74]	Non-full CSI Not fully backlogged Traffic: Stations transmitting at half saturation rate	Control theory	Single hop (ad-hoc) Random access	Throughput Fairness Stability
Multi-cell opportunistic scheduler for cellular networks using random beamforming [75]	Non-full CSI		Multi-cell Downlink	Throughput Fairness
Multi-cell scheduling with consideration of delay and capacity limitations in backhaul links [76]	Full CSI Error-free backhaul		Multi-cell Downlink	Throughput Fairness

- Fixed number of mobile users (no one leaves or joins the channel).
- Scheduling decision can be made offline.
- Scheduling single user per frame.
- Scheduling in single channel network.

Some of these assumptions can highly affect the performance of a proposal. For instance, base station obtains the CSI via feedback received from users. Assuming that every user is sending feedbacks on every slot and for every available subchannel, a big chunk of bandwidth is wasted for feedbacks. Another example is the assumption of fully backlogged queues that made the MaxWeight [13] scheduler optimal, although, it was proven to be sub-optimal under queue dynamics (i.e., not fully backlogged queues).

There have been efforts to eliminate such unrealistic assumptions. However, by removing these assumptions, we would increase computation and modeling complexity. We believe that opportunistic scheduler proposals are not implemented in practice due to high complexity of the proposals and their unrealistic assumptions which may results in scarce interest for practical applications. It would be interesting to investigate the performance of optimal schedulers in real implementations or under realistic assumptions.

2) *Fairness*: Fairness is also an open issue in opportunistic schedulers. In comparison with other categories of schedulers, fairness in opportunistic scheduling is an under-explored field. In a real network, it is crucial for the operator to achieve the service rate committed to users. A first step toward encouraging operators to implement opportunistic schedulers into their networks is a proposal with tight fairness control features. As mentioned previously, there is always trade-off between fairness and opportunism and the challenge is to decide when a scheduler should stop being opportunistic and starts being fair. However, increasing fairness does not always result in throughput reduction. For example when multiple users are experiencing the same channel conditions, the scheduler can improve the fairness without affecting the throughput. In other words, in the event that multiple users can transmit at the same rate, the scheduler can tune the fairness by allowing the

least served user to transmit. As a result, the fairness can be improved without throughput impairment.

3) *Data offloading*: Nowadays, many mobile devices are equipped with WiFi. In [85] and [86], the authors propose to take advantage of WiFi to offload the elastic traffic on WiFi and real time traffic on cellular network. Since a large portion of cellular traffic is formed by elastic traffic, *opportunistic* data offloading can reduce the congestion in cellular networks and improve the QoS of users. Data offloading can also be performed by exploiting the user's behavior. In event that people with shared interest gather in one place (e.g., an sport match), it is very likely that they try to access the same content multiple times. In such case, users who already acquired the content, can offload the content to users who are requesting for it. This approach benefits cellular network by saving bandwidth for re-sending the same content. In addition, it can save the network from saturation in places where the density of networks varies much (e.g. stadium).

4) *Cooperative networks*: We believe that opportunistic scheduling can effectively take advantage of cooperative diversities created in cooperative networks [87]–[89]. For example, users with good channel quality can forward data for the users with poor channel quality. Hence, using, e.g., relay nodes among the mobile users, the scheduler would be not required to wait for channel quality improvements of users in bad channel states, or waste bandwidth by scheduling users with low transmission rates. As observed in [26], users can benefit from cooperation both individually and network wise. There are several approaches for cooperative communications, such as cooperative MIMO [90]–[92], and relaying [93]. In cooperative MIMO or virtual antenna array, users with single antenna join and create a virtual MIMO device [92]. There are various relaying techniques but they all operate based on the concept of forwarding data to an entity with better or more reliable channel [93]. Another interesting scheme is cooperative sensing in cognitive network in which users cooperate for detecting more accurate spectrum opportunities [94]. For instance, in a very recent work, the authors of [100] propose a cooperative packet delivery for hybrid wireless networks.

TABLE VIII: Performance comparison of different categories of opportunistic scheduling algorithms.

Category	Assumptions	Analytical tools	Topology	Other focus	Achieved improvement
Capacity	-Non-full CSI -Full CSI -Not fully backlogged -Fully backlogged -Traffic: Generic, Poisson, Bernoulli -Error free transmission	-Markov chain -Lyapunov drift -Lyapunov optimization -Lagrangian multipliers -Finite state Markov chain -Markov decision process -Partially observable Markov decision process -Dynamic programming -Linear programming -Restless multi-armed bandit process -Optimal stopping theory -Gilbert-Elliot model -Maximum weight match	-Single cell -Multi-cell -OFDMA -OFDM -HSDPA -FDD -Downlink and generic	-Throughput -Throughput-optimal -Near throughput-optimal -Delay -Fairness -Stability	-Up to 37% throughput gain over Round Robin. -Up to 70% delay reduction compared to Round Robin. -Up to 95% average queue size improvement over MaxWeight. -Up to 4% improvement over expected utility compared to optimal policy.
QoS	-Non-full CSI -Not fully backlogged -Fully backlogged -Traffic: Poisson, Pareto	-Markov chain -Lyapunov drift -Lyapunov optimization -Lagrangian multipliers -Foster theorem -Convex optimization -Dynamic programming -Duality theory -Unity cube mapping -Maximum weight match	-Single cell -Single hop -multi-hop -TDMA -OFDMA -HDR -Downlink and generic	-Throughput -Throughput-optimal -Delay-optimal -Delay -Stability	-Between 30~40% throughput gain over M-LWDF. -Achieving up to 92% of maximum achievable throughput. -Average delay gain of 50% over MaxRate. -Between 20~80% delay improvement over Exp rule which itself outperform MaxWeight.
Fairness	-Full CSI -Not fully backlogged -Fully backlogged -Traffic: Poisson	-Convex optimization -Lagrangian multipliers -Duality theory -Maximal bipartite matching -Hungarian algorithms -Linear programming -Non-linear programming -Newton-Raphson method	-Single cell -OFDMA -OFDM -TDMA -TDM -Downlink and generic	-Jain's fairness -Temporal fairness -Utilitarian fairness -Minimum performance guarantee	-Achieves a minimum Jain's index of 0.6 with varying data rate. -Up to 46% throughput improvement over Round Robin. -Delay improvement of 12% over PFA.
Distributed	-Full CSI -Not-full CSI -Fully backlogged -Not fully backlogged -Traffic: Stations transmitting at half saturation rate -Error free backhaul communications	-Optimal stopping theory -Nash equilibrium -Duality theory -Stochastic Lagrangian approach -Lebesgue's convergence theorem -Control theory -Game theory	-Single hop -Multi-cell -Random access -Ad-hoc -Downlink and generic	-Throughput -Fairness -Stability	-Throughput gain of 100% over an independent scenario. -Up to 40% throughput gain over the DOS proposed in [70] that obtains 13.9~42.8% gain over the proposal in [96] that improves the throughput of existing protocols up to 50%.

They use coalitional game theory to group mobile users that assist each other by relaying traffic. Today's cellular phones are equipped with a secondary interface (e.g. WiFi) as well as the cellular interface. The authors propose to take advantage of this capability to form coalitions among mobile users using the secondary interface. Once a coalition is formed, coalition members can send their traffic, on secondary interface, to the user in the best channel to relay their traffic to cellular network.

5) *Implementation:* To the best of our knowledge, no one explored the gain obtained via opportunism in an actual experiment. Hence, we believe such implementations would provide insights into unforeseen problems and consequences of opportunistic scheduling.

D. Future trends

After two decades of research, we can say that opportunistic scheduling became very mature. This maturity calls for new opportunistic schedulers whose designs are backed up with analysis. Indeed, many authors not only show the performance advantage of their proposal, but also prove the stability of the schedulers.

Currently, researchers are active towards two major directions. First, researchers evaluate the performance of existing proposals under more realistic scenarios such as flow-level dynamics, multi-user multi-carrier scheduling, and mobility. This helps us to have a better overview of the system performance in a real world scenario. Second, researchers seek for novel applications and new challenges for opportunistic scheduling. Use of opportunistic scheduling in cooperative communications is one of the newly explored areas which has attracted the interest of many researchers.

An interesting example to integrate opportunistic scheduling with cooperative communications would be forming clusters among mobile users. In this scenario, each cluster chooses the member experiencing the best channel to handle the cluster traffic.

VIII. CONCLUSIONS

In this paper, we categorize the opportunistic scheduling proposals based on the approach they took for formulating the problem of scheduling, i.e., capacity, QoS, fairness, distributed scheduling.

The opportunistic proposals that aim to improve the system capacity are able to achieve throughput optimality. However, these proposals may not be able to guarantee any fairness and QoS metrics. QoS related opportunistic schedulers can also achieve throughput-optimal or near throughput-optimal results while maintaining the QoS constraints. Throughput gain obtained by these proposals depends on the number of QoS metrics and their desired values. We also observed that the unfairness issue which is caused by the greedy nature of opportunistic schedulers can be resolved at the cost of throughput reduction. Although these schedulers achieve fairness at the cost of throughput reduction, they still outperform the non-opportunistic schedulers. In order to reduce the complexity of opportunistic schedulers, some proposed opportunistic distributed scheduling.

Opportunistic scheduling appears to be a promising solution to wireless channel bandwidth limitations. In this survey, the importance of opportunistic schedulers in wireless networks was emphasized with extensive reviews and comparisons among them. It was observed that the proposals with focus on capacity can provide up to 37% capacity gain and 70% delay reduction compared to Round Robin. QoS oriented proposals offered throughput improvement up to 30% throughput gain and delay improvement up to 80%. Proportional fair based schedulers achieved a minimum Jain's index of 0.6 while maintaining the QoS requirements. Finally distributed scheduling proposals appeared to have high throughput gain without the need for a central managing entity. We also introduced some of the open issues and discussed possible solutions based on the state-of-the-art proposals and researches done within past two decades.

REFERENCES

- [1] 3GPP, "Dual-cell high speed downlink packet access (HSDPA) operation (release 8)," *3GPP TR 25.825*.
- [2] H. Ekstrom, A. Furuskar, J. Karlsson, M. Meyer, S. Parkvall, J. Torsner, and M. Wahlqvist, "Technical solutions for the 3G long-term evolution," *IEEE Communications Magazine*, vol. 44, no. 3, pp. 38 – 45, March 2006.
- [3] IEEE 802.16-2004, "IEEE standard for local and metropolitan area networks part 16: Air interface for broadband wireless access systems amendment 3: Advanced air interface," 2004.
- [4] I. Mansfield. (2012, March) Mobile internet usage is doubling year on year. [Online]. Available: <http://www.cellular-news.com/story/53388.php>
- [5] ——. (2012, March) Wireless network data problems increase as more subscribers use web apps. [Online]. Available: <http://www.cellular-news.com/story/53329.php>
- [6] E. Liu and K. Leung, "Proportional fair scheduling: Analytical insight under rayleigh fading environment," in *Proceedings of IEEE WCNC*, 2008, pp. 1883–1888.
- [7] M. L. Pinedo, *Scheduling: Theory, Algorithms, and Systems*, 4th ed. Springer, 2012.
- [8] R. Knopp and P. Humblet, "Information capacity and power control in single-cell multiuser communications," in *Proceedings of IEEE ICC*, vol. 1, June 1995, pp. 331 –335 vol.1.
- [9] —, "Information capacity and power control in single-cell multiuser communications," in *Proceedings of IEEE ICC*, vol. 1. IEEE, 1995, pp. 331–335.
- [10] P. Viswanath, D. Tse, and R. Laroia, "Opportunistic beamforming using dumb antennas," in *IEEE Transactions on Information Theory*, June 2002, p. 449.
- [11] N. Sharma and L. Ozarow, "A study of opportunism for multiple-antenna systems," *IEEE Transactions on Information Theory*, vol. 51, no. 5, pp. 1804 – 1814, May 2005.
- [12] X. Qin and R. Berry, "Exploiting multiuser diversity for medium access control in wireless networks," in *Proceedings of IEEE INFOCOM*, vol. 2. IEEE, 2003, pp. 1084–1094.
- [13] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, R. Vijayakumar, and P. Whiting, "Scheduling in a queuing system with asynchronously varying service rates," *Probability in the Engineering and Informational Sciences*, vol. 18, pp. 191–217, April 2004.
- [14] P. van de Ven, S. Borst, and S. Shneer, "Instability of maxweight scheduling algorithms," in *Proceedings of IEEE INFOCOM*, April 2009, pp. 1701 –1709.
- [15] J.-W. Lee, R. Mazumdar, and N. Shroff, "Opportunistic power scheduling for multi-server wireless systems with minimum performance constraints," in *Proceedings of IEEE INFOCOM*, vol. 2, March 2004, pp. 1067 – 1077 vol.2.
- [16] C. Wong, R. Cheng, K. Lataief, and R. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 10, pp. 1747–1758, 1999.
- [17] L. Badia, A. Baiocchi, A. Todini, S. Merlin, S. Pupolin, A. Zanella, and M. Zorzi, "On the impact of physical layer awareness on scheduling and resource allocation in broadband multicellular IEEE 802.16 systems [radio resource management and protocol engineering for IEEE 802.16]," *IEEE Wireless Communications Magazine*, vol. 14, no. 1, pp. 36–43, 2007.
- [18] A. Asadi and S. W. Tan, "An enhanced cross layer downlink scheduling algorithm for IEEE 802.16 networks," in *Proceedings of IEEE ICIN*, January 2011, pp. 212 –217.
- [19] M. Andrews and L. Zhang, "Scheduling algorithms for multi-carrier wireless data systems," in *Proceedings of ACM MobiCom*. ACM, 2007, pp. 3 –14.
- [20] S. Liu, L. Ying, and R. Srikant, "Scheduling in multichannel wireless networks with flow-level dynamics," in *Proceedings of the ACM SIGMETRICS*, 2010, pp. 191–202.
- [21] H. Al-Zubaidy, I. Lambadaris, and J. Talim, "Optimal scheduling in high-speed downlink packet access networks," *ACM Transactions on Modeling and Computer Simulation*, vol. 21, pp. 3:1–3:27, December 2010.
- [22] L. Tassiulas, "Scheduling and performance limits of networks with constantly changing topology," *IEEE Transactions on Information Theory*, vol. 43, no. 3, pp. 1067 –1073, May 1997.
- [23] M. Neely, E. Modiano, and C. Rohrs, "Dynamic power allocation and routing for time-varying wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 1, pp. 89 – 103, January 2005.
- [24] A. Eryilmaz and R. Srikant, "Fair resource allocation in wireless networks using queue-length-based scheduling and congestion control," in *Proceedings of INFOCOM*, vol. 3, March 2005, pp. 1794 – 1803 vol. 3.
- [25] C. Luo, F. R. Yu, H. Ji, and V. C. Leung, "Optimal channel access for TCP performance improvement in cognitive radio networks," *Wireless Networks*, vol. 17, pp. 479–492, February 2011.
- [26] K. Khalil, M. Karaca, O. Ercetin, and E. Ekici, "Optimal scheduling in cooperate-to-join cognitive radio networks," in *Proceedings of IEEE INFOCOM*, April 2011, pp. 3002 –3010.
- [27] A. Lyapunov, "The general problem of the stability of motion," *International Journal of Control*, vol. 55, no. 3, pp. 531–534, 1992.
- [28] S. Liu, L. Ying, and R. Srikant, "Throughput-optimal opportunistic scheduling in the presence of flow-level dynamics," *IEEE/ACM Transactions on Networking*, vol. 19, no. 4, pp. 1057 –1070, August 2011.
- [29] L. Kleinrock, "Queueing systems. volume 1: Theory," 1975.
- [30] D. Bertsekas, D. Bertsekas, D. Bertsekas, and D. Bertsekas, "Dynamic programming and optimal control," 1995.
- [31] W. Ouyang, S. Murugesan, A. Eryilmaz, and N. Shroff, "Exploiting channel memory for joint estimation and scheduling in downlink networks," in *Proceedings of IEEE INFOCOM*, April 2011, pp. 3056 –3064.
- [32] K. Glazebrook, H. Mitchell, and P. Ansell, "Index policies for the maintenance of a collection of machines by a set of repairmen," *European Journal of Operational Research*, vol. 165, no. 1, pp. 267 – 284, 2005.
- [33] S. Guha, K. Munagala, and P. Shi, "Approximation algorithms for restless bandit problems," *Journal of the ACM (JACM)*, vol. 58, pp. 3:1–3:50, 2010.
- [34] P. Chaporkar, A. Proutiere, H. Asnani, and A. Karandikar, "Scheduling with limited information in wireless systems," in *Proceedings of ACM MobiHoc*, 2009, pp. 75–84.

- [35] M. Neely, "Stochastic network optimization with application to communication and queueing systems," *Synthesis Lectures on Communication Networks*, vol. 3, no. 1, pp. 1–211, 2010.
- [36] G. Peškir, A. Shiryaev, and A. Shiryaev, *Optimal stopping and free-boundary problems*. Birkhauser, 2006, vol. 10.
- [37] P. Jacko, "Value of information in optimal flow-level scheduling of users with Markovian time-varying channels," *Journal of Performance Evaluation*, vol. 68, no. 11, pp. 1022–1036, November 2011.
- [38] M. Ouyang and L. Ying, "On optimal feedback allocation in multi-channel wireless downlinks," in *Proceedings of ACM MobiHoc*. ACM, 2010, pp. 241–250.
- [39] X. Lin and N. Shroff, "The impact of imperfect scheduling on cross-layer rate control in wireless networks," in *Proceedings of IEEE INFOCOM*, vol. 3, March 2005, pp. 1804 – 1814 vol. 3.
- [40] P. Chaporkar, K. Kar, and S. Sarkar, "Throughput guarantees through maximal scheduling in wireless networks," in *Proceedings of 43rd Annual Allerton Conference on Communication, Control and Computing*, 2005, pp. 28–30.
- [41] H. Kim and G. de Veciana, "Losing opportunism: Evaluating service integration in an opportunistic wireless system," in *Proceedings of IEEE INFOCOM*, May 2007, pp. 982–990.
- [42] B. Sadiq, S. J. Baek, and G. de Veciana, "Delay-optimal opportunistic scheduling and approximations: The log rule," in *Proceedings of IEEE INFOCOM*, April 2009, pp. 1692–1700.
- [43] M. Neely, "Opportunistic scheduling with worst case delay guarantees in single and multi-hop networks," in *Proceedings of IEEE INFOCOM*, April 2011, pp. 1728–1736.
- [44] K. W. Choi, W. S. Jeon, and D. G. Jeong, "Resource allocation in OFDMA wireless communications systems supporting multimedia services," *IEEE/ACM Transactions on Networking*, vol. 17, no. 3, pp. 926–935, June 2009.
- [45] R. K. Beidokhti, M. H. Yaghmaee Moghaddam, and J. Chitizadeh, "Adaptive QoS scheduling in wireless cellular networks," *Wireless Networks*, vol. 17, pp. 701–716, April 2011.
- [46] S. Patil and G. de Veciana, "Managing resources and quality of service in heterogeneous wireless systems exploiting opportunism," *IEEE/ACM Transactions on Networking*, vol. 15, no. 5, pp. 1046–1058, October 2007.
- [47] —, "Reducing feedback for opportunistic scheduling in wireless systems," *IEEE Transactions on Wireless Communications*, vol. 6, no. 12, pp. 4227–4232, December 2007.
- [48] S. Shakkottai, "Effective capacity and QoS for wireless scheduling," *IEEE Transactions on Automatic Control*, vol. 53, no. 3, pp. 749–761, April 2008.
- [49] P. Key, L. Massoulie, A. Bain, and F. Kelly, "Fair internet traffic integration: network flow models and analysis," *Annals of Telecommunications*, vol. 59, pp. 1338–1352, 2004.
- [50] E. Altman, A. Orda, and N. Shimkin, "Bandwidth allocation for guaranteed versus best effort service categories," in *Proceedings of IEEE INFOCOM*, vol. 2, March 1998, pp. 617–624 vol.2.
- [51] T.-J. Lee and G. de Veciana, "Model and performance evaluation for multiservice network link supporting ABR and CBR services," *IEEE Communications Letters*, vol. 4, no. 11, pp. 375–377, November 2000.
- [52] S. Borst and P. Whiting, "Dynamic channel-sensitive scheduling algorithms for wireless data throughput optimization," *IEEE Transactions on Vehicular Technology*, vol. 52, no. 3, pp. 569–586, May 2003.
- [53] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, "Providing quality of service over a shared wireless link," *IEEE Communications Magazine*, vol. 39, no. 2, pp. 150–154, February 2001.
- [54] C. Hartmann, R. Vilzmann, A. Schmitt-Nilson, and J. Eberspacher, "Channel aware scheduling for user-individual QoS provisioning in wireless systems," in *Proceeding of IEEE VTC*, vol. 2, September 2004, pp. 1009–1013 Vol. 2.
- [55] NS2. [Online]. Available: <http://www.isi.edu/nsnam/ns/>
- [56] Z. Zhang, Y. He, and E. Chong, "Opportunistic downlink scheduling for multiuser OFDM systems," in *Proceedings of IEEE WCNC*, vol. 2. IEEE, 2005, pp. 1206–1212.
- [57] X. Liu, E. Chong, and N. Shroff, "A framework for opportunistic scheduling in wireless networks," *Computer Networks*, vol. 41, no. 4, pp. 451–474, 2003.
- [58] H. Kuhn, "The Hungarian method for the assignment problem," *Naval research logistics quarterly*, vol. 2, no. 1-2, pp. 83–97, 1955.
- [59] J. Yang, Z. Yifan, W. Ying, and Z. Ping, "Average rate updating mechanism in proportional fair scheduler for HDR," in *Proceedings of IEEE GLOBECOM*, 2004, pp. 3464–3466 Vol.6.
- [60] J.-T. Tsai, "State-dependent proportional fair scheduling algorithms for wireless forward link data services," in *Proceedings of IEEE INFOCOM*, April 2008, pp. 2414–2422.
- [61] P. Bender, P. Black, M. Grob, R. Padovani, N. Sindhushyana, and S. Viterbi, "CDMA/HDR: a bandwidth efficient high speed wireless data service for nomadic users," *IEEE Communications Magazine*, vol. 38, no. 7, pp. 70–77, July 2000.
- [62] A. Jalali, R. Padovani, and R. Pankaj, "Data throughput of CDMA-HDR a high efficiency-high data rate personal communication wireless system," in *Proceedings of IEEE VTC*, vol. 3, 2000, pp. 1854–1858 vol.3.
- [63] R. Almatarneh, M. Ahmed, and O. Dobre, "Performance analysis of proportional fair scheduling in OFDMA wireless systems," in *Proceedings of Vehicular Technology Conference (VTC)*. IEEE, 2010, pp. 1–5.
- [64] R. Jain, D. Chiu, and W. Hawe, "A quantitative measure of fairness and discrimination for resource allocation in shared computer systems," *DEC research report TR-301*, 1984.
- [65] Z. Shen, J. Andrews, and B. Evans, "Adaptive resource allocation in multiuser OFDM systems with proportional rate constraints," *IEEE Transactions on Wireless Communications*, vol. 4, no. 6, pp. 2726–2737, 2005.
- [66] J.-A. Kwon, B.-G. Kim, and J.-W. Lee, "A unified framework for opportunistic fair scheduling in wireless networks: a dual approach," *Wireless Networks*, vol. 16, pp. 1975–1986, October 2010.
- [67] D. Bertsekas, A. Nedi, A. Ozdaglar *et al.*, "Convex analysis and optimization," 2003.
- [68] D. Bertsekas, "Nonlinear programming," 1999.
- [69] C. Kelley, *Solving nonlinear equations with Newton's method*. Society for Industrial Mathematics, 2003, vol. 1.
- [70] D. Zheng, W. Ge, and J. Zhang, "Distributed opportunistic scheduling for ad hoc networks with random access: An optimal stopping approach," *IEEE Transactions on Information Theory*, vol. 55, no. 1, pp. 205–222, January 2009.
- [71] S.-S. Tan, D. Zheng, J. Zhang, and J. Zeidler, "Distributed opportunistic scheduling for ad-hoc communications under delay constraints," in *Proceedings of IEEE INFOCOM*, March 2010, pp. 1–9.
- [72] M. Osborne and A. Rubinstein, *A course in game theory*. The MIT press, 1994.
- [73] P. Thejaswi, J. Zhang, M. Pun, H. Poor, and D. Zheng, "Distributed opportunistic scheduling with two-level probing," *IEEE/ACM Transactions on Networking (TON)*, vol. 18, no. 5, pp. 1464–1477, 2010.
- [74] A. Garcia-Saavedra, A. Banchs, P. Serrano, and J. Widmer, "Distributed opportunistic scheduling: A control theoretic approach," in *Proceedings of IEEE INFOCOM*, vol. 3, March 2012.
- [75] X. Tang, S. Ramprasad, and H. Papadopoulos, "Multi-cell user-scheduling and random beamforming strategies for downlink wireless communications," in *Proceedings of IEEE VTC*, September 2009, pp. 1–5.
- [76] R. Bendlin, Y.-F. Huang, M. Ivrlac, and J. Nossek, "Fast distributed multi-cell scheduling with delayed limited-capacity backhaul links," in *Proceedings of IEEE ICC*, June 2009, pp. 1–5.
- [77] W. Choi and J. Andrews, "The capacity gain from base station cooperative scheduling in a MIMO DPC cellular system," in *Proceedings of IEEE International Symposium on Information Theory*, July 2006, pp. 1224–1228.
- [78] —, "Base station cooperatively scheduled transmission in a cellular MIMO TDMA system," in *Proceedings of IEEE Information Sciences and Systems*, March 2006, pp. 105–110.
- [79] —, "The capacity gain from intercell scheduling in multi-antenna systems," *IEEE Transactions on Wireless Communications*, vol. 7, no. 2, pp. 714–725, February 2008.
- [80] H. Skjævlving, D. Gesbert, and A. Hjørungnes, "Low-complexity distributed multibase transmission and scheduling," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, pp. 116:1–116:9, January 2008.
- [81] S. Kiani, D. Gesbert, J. Kirkebo, A. Gjendemsjo, and G. Oien, "A simple greedy scheme for multicell capacity maximization," in *Proceedings of International Telecommunications Symposium*, September 2006, pp. 435–440.
- [82] S. Kiani, G. Oien, and D. Gesbert, "Maximizing multicell capacity using distributed power allocation and scheduling," in *Proceedings of IEEE WCNC*, March 2007, pp. 1690–1694.
- [83] M. Kountouris and D. Gesbert, "Memory-based opportunistic multi-user beamforming," in *Proceedings of IEEE International Symposium on Information Theory*, September 2005, pp. 1426–1430.

- [84] R. Bartle, *The elements of integration and Lebesgue measure*. Wiley Online Library, 1995.
- [85] A. Balasubramanian, R. Mahajan, and A. Venkataramani, "Augmenting mobile 3G using WiFi," in *Proceedings of ACM MobiSys*. ACM, 2010, pp. 209–222.
- [86] B. Han, P. Hui, and A. Srinivasan, "Mobile data offloading in metropolitan area networks," *ACM SIGMOBILE Mobile Computing and Communications Review*, vol. 14, pp. 28–30, November 2010.
- [87] M. Uysal, *Cooperative communications for improved wireless network transmission: framework for virtual antenna array applications*. Information Science Publishing, 2010.
- [88] E. Hossain, D. Kim, and V. Bhargava, *Cooperative cellular wireless networks*. Cambridge Univ Pr, 2011.
- [89] J. Laneman, D. Tse, and G. Wornell, "Cooperative diversity in wireless networks: Efficient protocols and outage behavior," *IEEE Transactions on Information Theory*, vol. 50, no. 12, pp. 3062–3080, 2004.
- [90] S. Cui, A. Goldsmith, and A. Bahai, "Energy-efficiency of mimo and cooperative mimo techniques in sensor networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, pp. 1089–1098, 2004.
- [91] S. Jayaweera, "Virtual mimo-based cooperative communication for energy-constrained wireless sensor networks," *IEEE Transactions on Wireless Communications*, vol. 5, no. 5, pp. 984–989, 2006.
- [92] J. Choi, "Opportunistic beamforming with single beamforming matrix for virtual antenna arrays," in *Proceedings of IEEE International Conference on communications*, 2011, pp. 1–5.
- [93] Q. Deng and A. Klein, "Relay selection in cooperative networks with frequency selective fading," *EURASIP Journal on Wireless Communications and Networking*, vol. 2011, no. 1, p. 171, 2011.
- [94] K. Choi, E. Hossain, and D. Kim, "Cooperative spectrum sensing under a random geometric primary user network model," *IEEE Transactions on Wireless Communications*, vol. 10, no. 6, pp. 1932–1944, 2011.
- [95] S. Bodas, S. Shakkottai, L. Ying, and R. Srikant, "Scheduling in multi-channel wireless networks: rate function optimality in the small-buffer regime," in *Proceedings of ACM SIGMETRICS*, 2009, pp. 121–132.
- [96] B. Sadeghi, V. Kanodia, A. Sabharwal, and E. Knightly, "Opportunistic media access for multirate ad hoc networks," in *Proceedings of ACM MobiCom*. ACM, 2002, pp. 24–35.
- [97] NS3. [Online]. Available: <http://www.nsnam.org/>
- [98] OPNET. [Online]. Available: <http://www.opnet.com/>
- [99] S. Ali, K. Lee, and V. Leung, "Dynamic resource allocation in OFDMA wireless metropolitan area networks [radio resource management and protocol engineering for IEEE 802.16]," *IEEE Wireless Communications Magazine*, vol. 14, no. 1, pp. 6–13, 2007.
- [100] K. Akkarajitsakul, E. Hossain, and D. Niyato, "Cooperative packet delivery in hybrid wireless mobile networks: A coalitional game approach," *IEEE Transactions on Mobile Computing*, no. 99, pp. 1–1, 2012.



Vincenzo Mancuso obtained his master degree in Electronics from University of Palermo, Italy, in 2001, and a PhD in Electronics, Computer Science and Telecommunications from the same University in 2005. After the PhD, he has collaborated with University of Roma "Tor Vergata" and University of Palermo. He has been visiting scholar at the ECE Department of Rice University, Houston, Texas, and postdoc in the MAESTRO team at INRIA Sophia Antipolis, France. Since September 2010, Vincenzo is Staff Researcher at Institute IMDEA Networks, working on analytical and experimental projects on wireless networks (802.11 and 802.16/LTE) and energy efficient network protocols. He's also associate professor at University Carlos III of Madrid since 2011.



Arash Asadi received a master with highest distinction in Telematics Engineering from University Carlos III of Madrid, Spain, in October 2012, and a master degree in Telecommunication Engineering from Multimedia University, Malaysia in January 2011. He obtained his bachelor degree in Electronic Engineering from Islamic Azad University, Iran, in 2006. Since 2011, he has worked as research officer at Multimedia University, focusing on channel capacity improvements of wireless networks. Since July 2011, Arash is with Institute IMDEA Networks,

pursuing his PhD studies while being involved in research project addressing resource allocation in wireless network and next-generation cooperative networks.