A Comparative Study of Resource Allocation Schemes in Heterogeneous Cellular Networks on the Downlink

by

Ararat Shaverdian

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Applied Science
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2016

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Network densification through heterogeneous networks (HetNets) is considered as a promising paradigm to address the ever increasing mobile users' data demands in 5G networks. A HetNet consists of macro cells (each with a macro base station) overlaid with a number of small cells (each with a low-power base station) and has been shown to significantly improve the network capacity when supported by carefully designed radio resource management (RRM) techniques. RRM is typically studied via a joint optimisation problem over three network processes, namely, resource allocation (RA), user association (UA) and user scheduling (US), and is the focus of this thesis. Our first objective is to characterise the optimal HetNet performance by jointly optimising these three processes through a unified framework under different channel deployment scenarios. Towards this, we focus on two RA schemes, namely, partially shared deployment (PSD) and co-channel deployment with almost blank subframes (ABS), proposed by 3GPP for future HetNets.

In the first part of the thesis, we revisit a unified optimisation framework under PSD that allows us to configure the network parameters (e.g., number of channels per-cell and power per-channel) and allocate optimal throughputs to users in a fair manner. The framework under consideration is based on a snapshot model where, in each snapshot, the number of users and channel gains are assumed to be fixed and known. Although the previous study on this framework provides many interesting engineering insights, it is primarily based on two wrong assumptions in terms of channel modelling and US which we correct in our work. We also revisit a similar framework but under ABS and conduct a thorough comparative study between ABS and PSD. We first show that the $\alpha-$fair scheduling problem under ABS is generally much more involved than that under PSD for $\alpha \neq 1$. To verify whether the US complexities involved from deploying ABS are justifiable, we compare the throughput performance of the two schemes under a static setting, where the number of users in each snapshot is assumed to be fixed. Our results indicate that PSD outperforms ABS for different choices of α and under different HetNet configurations.

In the second part of the thesis, we further study our frameworks under a dynamic

setting and continue our comparisons between the two RA schemes under different service-time models. The dynamic setting, as well as reaffirming the upper-hand of PSD, provides a number of new insights, most importantly the fact that the conventional physical-layer based UA schemes do not always work well. Motivated by this observation, we further explore the problem of UA under PSD with the objective of improving an existing online UA scheme. We show that when users are periodically triggered to re-associate (on an individual basis), the online UA scheme can significantly improve the system performance.

Acknowledgement

I would like to express my deepest gratitude to my supervisor Professor Catherine Rosenberg for her constant guidance, encouragement and patient without which the completion of this thesis would have been a distant dream. I will forever be grateful to her for giving me the opportunity to be a part of her group. I would also like to thank Professor Amir Keyvan Khandani and Professor Mohamed Oussama Damen for reviewing my thesis and their constructive comments.

I would also like to thank my fellow labmates Jagadish, Dariush, Santhana, Yigit and Shahida for all the invaluable discussions I had with them. I am beyond thankful to them for their comments and ideas, all of which helped me immensely with the writing of this thesis.

Last but not least, I gratefully acknowledge the financial support from Professor Rosenberg and The University of Waterloo in the form of scholarship and research assistantship throughout my studies here.

Dedication

To my teachers.

Contents

List of Figures

List of Tables

List of Acronyms

3GPP	3rd Generation Partnership Project
4G	Fourth Generation
5G	Fifth Generation
ABS	Almost Blank Subframes or CCD with Almost Blank Subframes
AG	Antenna Gain
BS	Base Station
C-RAN	Centralised-RAN
CCD	Co-channel Deployment
CoMP	Coordinated Multipoint Processing
CRE	Cell Range Expansion
FFR	Fractional Frequency Reuse
FUA	Fractional User Association
GM	Geometric Mean
HD	High Definition
HetNet	Heterogeneous Network
IP	Internet Protocol
KKT	Karush-Kuhn-Tucker
LTE	Long-Term Evolution
LTE-A	LTE-Advanced
MAC	Media Access Control
MBS	Macro Base Station
MCS	Modulation and Coding Scheme
MIMO	Multiple-input and Multiple-output
mmWave	Millimeter-wave
MUA	Multiple-BS User Association
NUM	Network Utility Maximisation
OD	Orthogonal Deployment
OFDM	Orthogonal Frequency-Division Multiplexing
PF	Proportional Fairness

PSD	Partially Shared Deployment
RA	Resource Allocation
RAN	Radio Access Network
RE	Range Extension
RRM	Radio Resource Management
SCF	Small-cell First
SC	Small Cell
SINR	Signal-to-Interference-plus-Noise-Ratio
SUA	Single-BS User Association
UA	User Association
US	User Scheduling

Chapter 1

Introduction

1.1 Overview

With the introduction of smart phones in the late 2000's and the broad range of IP-based applications that they offer, there has been an unprecedented surge in data demands from mobile customers. Being able to support these applications with, more or less, the same quality of service as fixed broadband connection coined the term *mobile broadband* and was a main driver for 4G systems [12]. With this in mind, operators started transitioning to an all-IP network to support rates from a few kbps per user, for voice calls, to 10's of Mbps per user for HD video streaming today and up to a target rate of 1 Gbps per user for future applications such as immersive multimedia [23]. In view of this massive surge in traffic demand, the industry is now aiming to increase today's network capacity by a factor of 100$\times$ within the next 20 years [32]. The fifth generation (5G) of wireless systems, currently under development, is expected to fulfill these objectives. 5G will offer a wide range of enabling technologies most of which revolve around three main paradigms [23, 9, 5, 3, 17, 27]:

1. Network densification through heterogeneous networks,
2. Exploiting higher spectrum frequencies (e.g., mm-Wave, tera-hertz bands),
3. Multi-antenna transmission and cooperative communications techniques (e.g., MIMO and CoMP).

Each of these three paradigms comes with potential gains and limitations. For example, network densification increases the spatial reuse and typically relies on topological optimisation rather than complex signal processing enhancements. However, it complicates resource, backhaul and mobility management [16, 24]. High frequency bands offer an abundance of unused spectrum but this comes at the cost of increased path-loss and often requires expensive equipments. Lastly, while most multi-antenna techniques enhance spectral efficiency through multiplexing, diversity, and antenna gain, they require tight synchronisation, complex signal processing capabilities and may fail due to inaccuracies in channel state information [23, 27].

Given these three paradigms with their potential gains and limitations, it is important to study the impact of each paradigm on network capacity over the years and identify the one(s) that has the most potential to help realise the 100-fold capacity target. The papers [38, 23] summarise the impact that each of these paradigms has had on network capacity from 1950 to 2000. According to their studies, there has been a million fold increase in wireless capacity over the 50 years. The breakdown of the capacity increase is as follows: 5× gain from designing better coding techniques, 5× gain from improved MAC and modulation schemes, 15× improvement from using wider spectrum and 2700× gain from network densification. According to this breakdown network densification seems to be the more promising technique and is the focus of this thesis. Although network densification will likely be incorporated along with one or a combination of the other techniques, we believe it will be the primary choice for capacity enhancement by researchers and operators.

Mobile operators have already started densifying their wide-area homogeneous networks into heterogeneous networks (HetNets) by deploying low-power base stations (BSs) within their existing macro cellular coverage. These low-power BSs are known as small cells (SCs) and come in different forms depending on their use. Home SCs (a.k.a, femto cells) are one form of SCs and are usually self-installed and maintained by the end user. Pico cells are another form of SCs and are installed and maintained by the operator, usually outdoors, to enhance the macro cell coverage. As of 2012, according to [18], the number of SCs was already greater than that of macro cells and this trend is expected to continue in years to come with 5G networks.

In the following, we will give a quick primer on HetNets and outline the issues limiting their performance.

1.2 Heterogeneous networks

A HetNet consists of a number of SCs with a small coverage area overlaying the existing homogeneous macro cellular network. Deploying these SCs within the existing networks brings a number of advantages. For example, it helps macro cells offload a portion of their users to SCs allowing more resources to be reused. In addition, it helps reduce the distance between users and access points which can mitigate the effect of path-loss and fading and improve the users' channel quality. This is particularly important for the cell-edge users who receive poor signals from the macro BSs (MBSs) and also suffer from the inter-cell interference from the neighbouring cells. Last but not least, because of the low installation and maintenance costs, SCs are an attractive solution for operators to increase their network capacity and fill the coverage holes of the MBSs.

HetNet deployment creates a number of challenges most of which impact the complexity of network operation. In particular, the disparity in power budget and load between the MBS and SCs makes the radio resource management[1] considerably more complicated than that in homogeneous networks where all cells typically have similar characteristics. Examples of other issues include increased number of handovers due to smaller cell coverage or bad quality of service for the SC users due to backhaul[2] limitations which call for new solutions in terms of mobility and backhaul management, respectively. Therefore, it is important to study the complexities involved in deploying HetNets in parallel with their potential gains. In this work, our focus will be only on the radio resource management aspect of HetNets and we will discuss its complexities in detail next. For more information on backhaul and mobility management, the reader may refer to [30, 25, 16].

1.2.1 Radio resource management: Network processes and their complex interplay

Operators design radio access networks with the objective of optimising a performance metric such as average delay or throughput per user, using a limited amount of radio resources. In other words, given a set of radio resources an operator wants to know *how to assign these resources to different BSs and users so as to optimise a performance metric?* Radio resource management (RRM) addresses this question, typically, via a joint optimisation problem over three network processes, namely, *resource allocation* (RA), *user association* (UA) and *user scheduling* (US), all of which are intricately coupled and, if designed carefully, can significantly improve the network performance [14]. Here, we only focus on the

[1]Radio resource management is the process of assigning radio resources, i.e., time, frequency and power, to different users and BSs in a network.

[2]Backhauls are the links between the MBS and SCs and between the MBS and the core network. Note that in this thesis, we assume that all the SCs overlaying a macro cell are connected to the MBS (in the macro cell) via these backhaul links.

problem of RRM for the downlink transmissions of OFDM-based systems. With this in mind, we define the three network processes as follows.

1. **Resource allocation** (RA) is the typically global (to the system) and slow-varying process of assigning channels[3] to the MBSs and SCs. RA is always a network-centric process which is often carried out centrally by the operator. In this thesis, we only deal with three RA schemes proposed by 3GPP [10] in which each MBS is allocated, say, M channels and these channels are further shared between the MBS and SCs via the following channel deployment schemes.

   (a) **Orthogonal Deployment** (OD): The MBS is assigned a subset of the dedicated channels among the M channels and the remaining channels are exclusively allocated to the SCs. This deployment relies on a simple frequency-domain interference avoidance mechanism between the MBS and its SCs.

   (b) **Partially Shared Deployment** (PSD): The MBS is assigned a set of dedicated channels similar to OD. The remaining channels are used by both the MBS and SCs. The reason for using the shared channels at the MBS is to help offload some of the users on the dedicated channels when necessary. The MBS would typically allocate a large portion of its power budget on its dedicated channels to provide a coverage umbrella for the macro cell and allocate the rest of it on its shared channels. Note that OD is a special case of PSD since if all the macro power budget is allocated to the dedicated channels, then PSD will reduce to OD. Therefore, PSD is always expected to provide an upper-bound to the performance of OD.

   (c) **Co-channel Deployment with Almost Blank Subframes** (CCD with ABS): All the BSs within a macro cell share all the available channels. However, since the MBS transmits at a much higher power level than the SCs it is only allowed to transmit data signals during some of the subframes. During the rest of the subframes, the MBS is only allowed to transmit reference signals at a very low power, hence the name *Almost Blank Subframes* (ABS). This way, the MBS will only cause negligible interference to its neighbouring co-channel cells during the ABS. This deployment relies on a simple time-domain interference avoidance mechanism between BSs of different tiers. We refer to the fraction of the time that the MBS is transmitting on the ABS as the *ABS duty cycle*. Also, we use the terms *CCD with ABS* and *ABS*, interchangeably, throughout this thesis.

[3] We use the terms channel and sub-channel interchangeably throughout this thesis.

2. **User association** (UA) refers to the process that selects a BS (or more) to be used by a user. It is also referred to as *cell selection* in the literature. UA rules typically require a user to measure the reference signals sent by each candidate BS and to associate with a BS based on those measurements. A commonly used UA rule in homogeneous networks is to associate a user upon arrival with the BS that offers the highest signal-to-interference-plus-noise-ratio (SINR). Such a UA rule, however, does not work quite well in the context of HetNets. This is because the MBS transmits at a much higher power level than the SCs and users, even at a close proximity of a SC, may still associate with the MBS resulting in an underutilisation of SCs. Therefore, HetNet deployment calls for new UA rules to incorporate fair and efficient load distribution among BSs. Considering the large variations in the number of users in a cell where users constantly enter and leave, clearly UA needs to be done rather more often than RA.

3. **User scheduling** (US) is the typically local (to a BS) quasi real-time process of allocating resource blocks[4] (RBs) at each BS to the users associated with it according to some fairness criteria. US is always a network-centric process, carried out at each BS (possibly with the help of other BSs) for all the users associated with it. A well-known example of a scheduling policy is Round Robin according to which a BS dedicates the entire network bandwidth for an equal amount of time to transmit to each associated user. Often, measures of fairness are introduced in scheduling policies by an operator to ensure a minimum quality of service for users with poor channel quality. The most widely-used measure of fairness in the context of cellular networks is $\alpha-$fairness first introduced in [28] where by changing the parameter $\alpha \in [0, \infty)$, one can achieve different levels of trade-off between maximum aggregate throughput and fairness. We will explain the notion of $\alpha-$fairness in more detail in Chapters 2 and 3. Note that an operator should ideally provide fairness among all of its users and not only among the set of users associated with a particular cell. Therefore, fairness should be a global criteria encompassing all users across an entire HetNet. Furthermore, it is noteworthy that although RA and US are always network-centric processes, UA can be both network-centric or user-centric, i.e., the association decision can be made for the user by the network and reported to her or it can be made by the user herself and reported to the network.

The three aforementioned processes impact one another and optimising each process independent of the other processes results in a sub-optimal HetNet performance. For example, consider a HetNet under the CCD with ABS scheme where an operator wants to

[4] A RB corresponds to one channel for one time slot.

decide on the value of the ABS duty cycle. This will depend on UA and US rules. If the operator decides to use a simple Max-SINR association rule and Round Robin scheduler, then it will be best to choose a small ABS duty cycle since most users in the system will tend to associate with the MBS because of its significantly higher transmit power. Conversely, if the UA rule is chosen such that more users are made to associate with the SCs, then it will be best to choose a large ABS duty cycle so that the SC users will experience less interference from the MBS. Studying the interplay of such intricately coupled processes entails a unified framework that parametrises all three processes and allows us to characterise the performance of the HetNet by configuring the network parameters. The ability to model these processes using a unified framework would enable us to conduct a comparative study of different deployment choices and heuristics which is one of the main objectives of this thesis.

1.2.2 Modelling HetNets

Apart from their complex interdependence, the network processes also highly depend on the HetNet configuration such as the number of BSs and their locations. For example, consider a HetNet based on ABS scheme. The larger the number of deployed SCs, the larger the ABS duty cycle is expected to be because of the increased SC coverage area (and, hence, the number of the SC users). Even more so, the performance of the network processes depends on the dynamics of the HetNet under consideration such as the number of users in the system, arrival rates, variations in users' channel gains, load per BS, traffic scenario, etc.. Therefore, formulating a HetNet model that encompasses all or some of these aspects is key in order to get realistic insights from the unified framework.

We start off with a *snapshot model* where we assume that the location of BSs, number of users, their locations and channel gains are fixed and known for a given snapshot or realisation. The snapshot model is a simple, yet insightful, tool that enables us to formulate many *network utility maximisation* problems and study the mean performance of the system by averaging the network utilities over a large number of realisations [29, 36]. We study the snapshot model under two settings: static and dynamic. In the static setting, each realisation corresponds to a random distribution of a *fixed number of static users* with *fixed channel gains* in each cell of the HetNet[5]. This setting is a good candidate for offline-static study to compare the performance of different deployment choices and configure the network parameters (e.g., shared channels in the case of PSD). Such a setting, however, does not quite reflect the behaviour of a real cellular network where there could be high

[5]Note that the channel gains and the location of the users in different realisations may be different.

variations in the number of users in different realisations depending on users' arrival or departure times. The dynamic setting will allow us to capture (some of) these effects. In this thesis, we focus on the service dynamics, i.e., we assume that users typically arrive in the system according to a predefined process, e.g., Poisson point process, and leave after a service-time where the service-time of a user will depend on the traffic scenario used for the system. We will consider two traffic scenarios: 1) a *fixed-delay* scenario where users stay for a fixed amount of time and 2) a *file-download* scenario where users download a file of a fixed size and, then, leave.

In the following, we summarise our objectives and contributions in this thesis.

1.3 Contributions

We study a HetNet comprising a number of macro cells, each overlaid with a number of SCs, by focusing on the downlink only. We revisit two unified optimisation frameworks, one under PSD and one under ABS, based on the snapshot model in both static and dynamic settings. We provide analytical and numerical results to make insightful remarks in each chapter about each of the two RA schemes. Our contributions are summarised below.

1. In Chapter 3, we revisit a joint optimisation framework under PSD based on the snapshot model. The framework was first discussed in [14] where the authors make a wrong assumption on the shared channels model at the MBS. We improve upon this work by deploying the right model for the shared channels at the MBS. The corrected model raises two questions regarding the 1) joint scheduling of the macro users and 2) power apportioning on the shared and dedicated channels at the MBS. First, we show that the gain obtained from the scheduling of macro users on both sets of the macro channels is negligible. This property, as we will show, can be exploited to decompose the local MBS joint scheduling problem into two independent scheduling problems; one for the users on the dedicated channels and one for the users on the shared channels. We, then, show that by using a simple formula-based scheduler on each set of the MBS channels, we can achieve near-optimal throughput performance. Finally, we address the importance of power apportioning between the shared and dedicated channels at the MBS and show that the right choice of power apportioning can significantly impact the system's $\alpha-$mean throughput. Lastly, we compare the performance of PSD and OD and show that, by performing the right power apportioning at the MBS, PSD significantly outperforms OD.

2. In Chapter 4, we revisit a joint optimisation framework under ABS based on the snapshot model. The framework was first discussed in [6], where the authors show that the optimal PF scheduling problem under ABS is a simple linear-in-time algorithm. Interested in characterising the optimal scheduler for $\alpha \neq 1$, we prove that the optimal $\alpha-$fair scheduling under ABS can be NP-hard and, hence, much more involved than that under PSD (which is formula-based for all values of α). To verify whether the scheduling complexities involved from deploying ABS are justifiable, we further conduct a thorough comparative study between ABS and PSD in a static setting and show that the throughput gains achieved by deploying PSD are higher than ABS for different values of α and under different HetNet configurations and heuristics. Therefore, based on its simpler optimal $\alpha-$fair scheduler and higher throughput gains, we assert that PSD outperforms ABS in the static setting.

3. In Chapter 5, we further extend the optimisation framework to a dynamic setting to compare the performance of PSD and ABS under different traffic models. The motivation behind this is that the snapshot model under a static setting does not capture all the dynamics of a real cellular network. We consider two traffic scenarios as the bases of our comparisons, namely, fixed-delay and file-download, with a respective performance metric of average $\alpha-$mean throughput and per-user delay. We obtain tight upper-bounds and lower-bounds for the system performance in terms of the two performance metrics, respectively, and show the dominance of PSD over ABS for different values of α. Furthermore, we show that while re-association in the fixed-delay scenario (under both PSD and ABS) does not result in much gain in performance metric, it significantly improves the metric in the file-download scenario. Lastly, we show the inefficiency of a very good physical-layer based UA scheme in the file-download scenario, suggesting that a better-designed scheme is required for online systems.

4. Lastly, we consider the problem of user association under PSD in a dynamic setting in Chapter 6. Building up on the previous work in [15], we propose a simple device-centric association scheme where users are periodically and individually given a chance to re-associate to another BS in a greedy manner. Using simulation results, we show that our proposed scheme results in a very good performance compared to the optimal. Our proposed scheme is device-centric, eliminating the need for a network-centric global re-association of users, and only requires a small amount of information exchange from the network to each user.

1.4 Outline

The rest of the thesis is organised as follows. In Chapter 2, we present a summary of the related work. In Chapter 3, we present the optimisation framework under PSD for a given snapshot. In Chapter 4, we use the framework under PSD along with a similar framework under ABS to conduct a thorough comparative study between the two RA schemes in a static setting. In Chapter 5, we continue the comparisons between the two schemes under a dynamic setting. In Chapter 6, we focus on the UA problem in the dynamic setting under PSD. We present the UA scheme in [15] and propose a new re-association technique to further improve the scheme. In Chapter 7, we present a summary of the work done in this thesis and future research directions.

Chapter 2

Literature Review

In this chapter, we provide an overview of the literature work related to RA, UA and US under OD, PSD and ABS schemes in the context of OFDM-based HetNets. We will outline the novelties and limitations of the existing work and comment on how it is related to and improved upon in our work.

2.1 User scheduling

US in HetNets is a well-studied problem and various scheduling policies, often based on some throughput-based objective, have been proposed. The most widely-used objective is the sum of the $\alpha-$fair utility of user throughputs. These US problems are also known as $\alpha-$fair scheduling problems in the literature and fall in a more general category of problems known as *network utility maximisation* (NUM). The notion of $\alpha-$fairness was proposed in [28] and is widely used as a measure of fairness in throughput assignment in the context of cellular networks today. By changing the parameter $\alpha \in [0, \infty)$, one can achieve different levels of trade-off between aggregate throughput of the system and fairness in terms of throughput allocation to users. The following values of α and their corresponding objectives are often used in the literature for throughput assignment to users. For $\alpha \to \infty$, the $\alpha-$fair scheduling problem yields max-min throughput allocation, i.e., the scheduler tries to maximise the worst users' throughput. [37] shows that max-min scheduling sacrifices the efficiency (in terms of aggregate throughput) of the system for maximum fairness in throughput allocation. For $\alpha = 1$, the $\alpha-$fair scheduling problem, also known as *proportional fairness* (PF) scheduling, corresponds to sum of the logarithm of user throughputs and provides a good trade-off between fairness and efficiency [20]. For

$\alpha = 0$, the scheduling problem corresponds to the sum throughput maximisation. However, it makes no attempt to ensure fairness in throughput assignments.

Recall that fairness should be a global criteria encompassing all users across an entire HetNet. For a given RA and UA, we refer to this joint scheduling problem within the system as the *global scheduling* problem. Solving this global problem will require a high level of coordination and signalling among BSs which may be infeasible in today's HetNets. Hence, to decouple the global problem into local (per-BS) scheduling sub-problems, a lot of the work in the literature make the following assumptions [15, 14, 16]:

1. A BS transmits on all the channels allotted to it at a given time. With this assumption, the interference and, hence, the SINR and link rates, of all user locations in a HetNet will become fixed, making the local scheduling at each BS independent of scheduling at other BSs[1]. However, note that this assumption simplifies a time and frequency domain scheduling to a pure time-domain scheduling problem (where a BS allocates all of its sub-channels to one user at a given time) which implies that the channel-dependent scheduling aspect of the system can no longer be exploited.

2. Each SC is connected to the MBS via a high capacity wired backhaul. This is particularly an important assumption in terms of decoupling the global scheduling problem since if the MBS is of limited backhaul capacity, allowing each BS to independently schedule its users can lead to the violation of the MBS backhaul constraint.

3. Users are only allowed to associate with one BS; otherwise, joint scheduling of a user by two or multiple serving BSs will naturally require some level of coordination between conflicting radio resources (e.g., co-channel interference between the serving BSs). This assumption allows the network to schedule users locally (on a per-BS basis), if the two above-mentioned assumptions hold.

Note that if any of the aforementioned assumptions does not hold, then local scheduling may be sub-optimal. [16] shows that with the above-mentioned assumptions the global $\alpha-$fair scheduling problem can decouple into local scheduling sub-problems and derives closed-form solutions for each of the sub-problems. Furthermore, the authors show that for the scenario where the MBS backhaul is sufficiently provisioned but where the SC backhaul links have limited capacities, the global scheduling problem can still be decomposed into independent local sub-problems. However, the local α-fair schedules are different from

[1]Note that, here, we assume that users can report their channel gains to their associated BSs through *perfect* feedback channels. Therefore, each BS can compute the interference, SINR and link rates seen by its users using the channel gain reports and, then, schedule them accordingly.

those of the scenario of very large backhaul capacities. For the more general scenario where the MBS backhaul is also of limited capacity, the user schedule at a BS is affected by the channel gains of users in other BSs and, hence, the global problem in general does not decompose into local problems. The authors, however, propose two local US heuristics and show that, under some mild assumptions, they work well.

2.1.1 User scheduling with link adaptation

Due to the high variations in instantaneous channels, scheduling is usually performed along with an underlying mechanism that dynamically adjusts the transmission power or rate according to the channel state. Power adaptation mechanism adjusts the transmit power based on the channel state to maintain a near-constant data rate at the receiver which is, in particular, a desirable affect for voice services. Rate adaptation mechanism, on the other hand, maintains the transmit power at a constant level and adjusts the rate (by varying modulation and channel coding schemes) to compensate for channel variations. Scheduling with rate adaptation is suitable for packet-data traffic where a constant rate is not required as long as the (long-term) average rate is above a certain threshold [12]. Rate adaptation is also known as *Adaptive Modulation and Coding*. In this work, we assume that the HetNets under consideration use Adaptive Modulation and Coding with discrete rates as proposed in [26].

2.2 Resource allocation

RA is the typically global (to the system) and slow-varying process of assigning channels to the MBS and SCs. In OFDM-based homogeneous networks, RA is often simple as all BSs have similar power budget and load. A common way of allocating resources in homogeneous networks is to assign equal number of channels to each BS with some *spatial reuse factor*, r, greater than one to mitigate co-channel inter-cell interference. Another similar approach is outlined in [7], where a fraction of the available channels is used by all BSs and the rest of the channels is divided between BSs with a reuse pattern. The shared channels are used for *cell-interior* users who do not experience severe co-channel inter-cell interference, whereas the rest of the channels are used for *cell-edge* users who are more prone to such interference. This RA scheme is known as *Fractional Frequency Reuse* (FFR). In HetNets, the problem of RA is more complex due to the disparity in power budget, load and coverage of the MBSs and SCs. Currently, three RA schemes have been proposed by 3GPP [10] for future HetNets: *Orthogonal Deployment* (OD), *Partially Shared*

Deployment (PSD) and *Co-Channel Deployment with Almost Blank Subframes* (ABS). We will focus on these three RA schemes with an emphasis on the latter two.

Under a given RA scheme, the HetNet as a whole uses M' OFDM sub-channels and each macro cell (i.e., the MBS and its SCs) is allocated $M = \frac{M'}{r}$ sub-channels, where $r > 1$ is the reuse factor. Hence, a total of M OFDM sub-channels are available for each macro cell. In OD, the macro cell is assigned a fraction of the M available channels while the SCs jointly use the rest of the channels. Such a deployment protects SC users from the high-power MBSs interference and offers a simple inter-tier interference avoidance mechanism in the frequency domain. A closely similar RA scheme is PSD where a set of dedicated sub-channels are reserved for use only by the MBS and the remaining sub-channels are jointly used by the SCs and also MBS which uses a much lower power budget than the one used on the dedicated sub-channels. In ABS, the SCs and MBS transmit on the same channels. However, the MBS only transmits for a fraction of the subframes and is mute during the rest of the subframes. This RA scheme allows those SC users, who are rate-stifled by the high macro-interference when the MBS is on, to get prioritised service during the ABS duty cycle. See [33] for more details on the ABS scheme.

Each of the above-mentioned RA schemes involves tuning parameters, e.g., the number of MBS channels, the number of shared channels for PSD and OD and the ABS duty cycle for ABS. Therefore, to operate optimally, each scheme requires fine-tuning. In the case of PSD, for example, consider a snapshot of a cell in a HetNet with a fixed UA rule, US policy, channel gains and number of users. Finding the optimal RA scheme will involve searching through the set of all available channels and determining which channel(s) to be shared among the SCs and which one(s) to be exclusively assigned to MBS so that the objective function selected by the operator is maximised. This will be an arduous problem since channels would normally have different gains at different times for the same user-BS pair and determining the optimal channel allocation even for a small-sized set of available channels will be computationally difficult. Hence for the sake of simplicity, a lot of the work in the literature make the following assumptions [15, 14, 16]:

1. The channels assigned to each BS are flat, i.e., the channel gains across different sub-channels between a BS-user pair are equal at a given time.

2. The transmission power at each BS is equally distributed among its allotted channels.

3. The same RA parameters are used globally across the HetNet. This assumption will eliminate the need for finding the optimal RA parameter for each macro cell and, hence, significantly simplifying the process of channel allocation in the HetNet.

With these assumptions, the optimal RA problem will reduce to searching for the optimal *number* of shared channels for OD, the number of shared channels and the power budget for the shared channels for PSD and the ABS duty cycle for ABS so as to maximise the objective function.

2.3 User association

For a given user, UA occurs at least once at the arrival time and could occur at other instants, e.g, other users' arrivals or departures or triggered events. To distinguish between *UA upon arrival* and *UA at the instants other than the arrival time* of a given user[2], we refer to the latter as simply *user association* (or UA) and to the former as *user re-association*. Furthermore, from now on, we assume that a user retains her association with the same BS (that she associated with upon her arrival) until her departure from the system unless otherwise specified.

In homogeneous networks, UA is typically based on Max-SINR rule where a user upon arrival associates with the BS who offers the highest downlink pilot SINR. In HetNets, however, the problem of UA is more complicated due to the disparity of transmit power between the MBS and SCs. As a result, Max-SINR performs poorly in HetNet deployment since due to the much higher transmit power of the MBS most users will tend to associate with the MBS. This will, in turn, cause an underutilisation of the SCs. Hence, new UA rules have been envisaged to split the users between the MBS and SCs more efficiently and fairly. These rules generally fall into two categories: *single-BS user association* (SUA) and *multiple-BS user association* (MUA) rules. Under SUA rules, a user can only associate with one BS while, under MUA, a user can associate with multiple BSs.

SUA problems are generally formulated as integer optimisation problems with the objective of maximising the network's criteria (e.g., $\alpha-$fairness criteria). A number of iterative algorithms are proposed in the literature to solve such SUA problems under the global proportional fairness objective, for a fixed number of users and BSs in the system, e.g., [8, 22] to cite a few. However, such iterative algorithms usually involve high computational complexity and may not be suitable for real-time dynamic systems. For this reason, simpler UA rules have been proposed in the literature. Although these rules might result in sub-optimal performance, they have gained a lot of interest and are widely-used under different HetNet optimisation frameworks mainly due to their simplicity and ease of implementation. An example of such UA rules is Cell Range Expansion (CRE) introduced

[2]For simplicity, throughout this thesis we assume that all users in the system are static.

in [1]. Under CRE, a user at the time of association adds a positive biasing parameter to the SINR from her neighbouring SC and, so long as this sum value is greater than the SINR from the MBS, she associates with the SC. With CRE, the SCs coverage areas are virtually expanded and hence more users are off-loaded from the MBS to the SCs. The SC users who receive a stronger MBS SINR are called *CRE users* and the region where CRE users are located is called *CRE region* in the literature. CRE is often used along with CCD with ABS, since CRE users who experience severe downlink interference from the nearby MBS can get a chance to be served during the ABS duty cycle. Note that the biasing parameter for such UA rules may need to be fine-tuned and this can increase the complexity of the UA problem. Another similar UA rule introduced in [14] is *Small-cell First* (SCF). Under SCF, a user associates with a SC as long as the SINR received from the SC is greater than a pre-determined threshold. The authors in the paper show that, under PSD in a static setting, SCF performs near-optimally if the SCF threshold value and number of dedicated channels are fine-tuned. Another simple UA rule, proposed in [21], is *Range Extension* (RE) where a user associates with the BS with the lowest path-loss. The authors in [13] show that in a static setting, RE can perform better than Max-SINR under OD and PSD. All of the aforementioned UA rules (i.e., CRE, SCF and RE) rely on physical-layer measurements by the user and can be easily implemented without any computational complexity. However, these rules might result in sub-optimal performance in a dynamic system since they do not consider fairness in throughput assignment or network-level parameters such as BSs loads. Hence, new UA rules incorporating fairness and BSs load are needed to perform well in a dynamic system. In this thesis, we only focus on SUA rules. However, since MUA has recently gained a lot of interest, we will provide a brief introduction to it in the following.

The second category of UA rules in HetNets is MUA and, as another capacity-enabling technique, has recently gained a lot of interest. MUA allows data transmission from multiple BSs to a user and can potentially enhance the user's throughput. Although not implemented yet, it is anticipated that MUA can become a reality in near future via C-RAN making it a potential contender among capacity-enabling technologies. An example of MUA is Dual Connectivity (DC), currently being developed in LTE framework, where a user is allowed to associate with two BSs at the same time [4]. A typical use case of DC is splitting the voice and data flow between MBS and SCs, respectively. Moreover, DC can enhance mobility management since users are now being served by two access points and if the link to one of them is disconnected, the second one will continue to serve the user. A number of iterative (but complex) MUA rules have also been proposed to find optimal or near-optimal UA solutions for the PF problem in multi-cell HetNets. The authors in [40] have proposed a *fractional user association* (FUA) rule to allow users to associate with multiple BSs. They further extend FUA rule to a single-BS association by deploying a gradient projection method where users and BSs cooperatively find near-optimal association via iteratively exchanging a number of parameters.

An equally important problem to UA is deciding when to re-associate users. [35] studies the problem of designing an optimal UA under the PF objective metric. They propose an online algorithm where users report their information to a central node and based on an average throughput metric, the central node decides who should re-associate. They show via simulation that the proposed algorithm performs significantly better than Max-SINR. However, a major drawback of this scheme is that it is based on a network-centric paradigm and may not be a feasible option in terms of computational speed for systems with fast-varying channel gains and number of users.

2.4 Joint RA, UA and US under PSD

The problem of joint optimisation of RA, UA and US under PSD with PF objective metric, in a static setup, is well studied in [14, 13]. They show that, for a fixed UA rule and RA parameters (i.e., channels and power budget per BS), the optimal US rule under PSD is equivalent to local (per-BS) equal-time sharing under some mild assumptions. They compare a number of SUA rules and show that SCF outperforms Max-SINR and RE. Moreover, The authors derive tight upper-bounds for the joint optimisation problem with SUA and show that, if both the number of shared channel and the SCF (biasing) parameter are fine-tuned, SCF with equal-time sharing scheduler achieves near-optimal performance. It is noteworthy that they also show that CCD without ABS does very poorly compared to PSD. Although [14] provides valuable insights into the interplay of the network processes, it uses a wrong propagation model for the shared channels at the MBS since the authors assume that all channels at the MBS are flat and, yet, use a different SC channel model for the shared channels at MBS. [15] generalises the joint RA, US and UA optimisation framework under PSD with PF objective to the general-case α-fair objective and derives closed-form solution for the optimal schedules. The authors also study the problem of UA under PSD and OD in a dynamic setup where they propose a simple and user-centric $\alpha-$*fair UA* scheme under a global $\alpha-$fair throughput allocation framework and show that, if the backhaul is not a bottleneck, the proposed scheme yields optimal results. The proposed algorithm requires only a small amount of information exchange from the network to the users. This study is, however, limited to the case where re-association is not allowed.

2.5 Joint RA, UA and US under ABS

The authors in [6] show that with a fixed UA and ABS duty cycle, the optimal PF (i.e., $\alpha = 1$) scheduling can be done locally (at each BS) based on a simple threshold-based algorithm under some mild assumptions. The analysis is extended to the case where all users are allowed to have multiple associations. In [39], the authors allow users to associate with multiple BSs and show that the majority of users associate with a single BS during both ABS and non-ABS duty cycles, although only a small number of users retain their association during each cycle. [19] show that, for PF, the optimal ABS duty cycle is the ratio of the number of ABS users and total number of users, assuming users can be served either in ABS or non-ABS but not both. Based on this relationship, they reduce their joint UA and US problem to a pure UA problem and propose a near-optimal and iterative online algorithm. The authors in [34] propose a local RA scheme with PF objective, wherein the spectral efficiencies of SC users are assumed to be different during ABS and non-ABS duty cycles. Each MBS and SC maximises its local PF metric for all users associated with it, using CRE with a fixed biasing parameter. The results indicate that if the number of ABS is large, then the users outside of CRE region and closer to SC should be scheduled during ABS.

Although the above-mentioned works provide good insights into the interplay of different network proesses under ABS, they do not consider the case of $\alpha \neq 1$. In [11, 31], the authors consider 'approximate' versions of the joint RA, UA and US problem for $\alpha \neq 1$ and propose some solution techniques. However, they do not characterise the NP-hardness of the exact $\alpha \neq 1$-fair US problem.

Chapter 3

A Revisit of Joint Resource Allocation, User Association and User Scheduling Optimisation Framework under PSD: A Snapshot Model

Summary: In this chapter, we

- revisit the joint RA, UA and US optimisation framework under PSD based on the snapshot model in a static setting,
- using numerical results, show that there is not much gain obtained from scheduling the MBS users on both shared and dedicated channels,
- address the importance of power apportioning between shared and dedicated sub-channels at the MBS and
- compare the performance of PSD with that of OD.

In Chapter 1, we discussed a number of network processes, namely, RA, US and UA which jointly impact the performance of a HetNet. We also discussed how these processes have a complex interplay which entails a unified framework that encompasses all three processes and allows us to characterise the network performance. In this chapter, we revisit a unified optimisation framework, proposed by [14], which studies the joint optimisation of RA, UA and US under PSD and a global proportional fairness (PF) objective. The framework is based on the snapshot model where the number of users and their channel gains are fixed and known. The authors in [14] show that, under some mild assumptions, the global PF scheduling problem decomposes into a set of independent local (per-BS) PF scheduling sub-problems each of which is equivalent to equal-time sharing. Later, the author in [15] generalises these results for all $\alpha \in [0, \infty)-$fairness criteria and derives closed-form solution for the optimal schedules.

Although [14] provides valuable insights into the interplay of the network processes, it is primarily based on a number of wrong assumptions in terms of channel modelling and user scheduling at the MBS. Firstly, the authors deploy an asymmetric channel modelling approach at the MBS where they use a SC model for the shared channels and a MBS model for the dedicated channels. Such a disparity in channel modelling implies that a user experiences different channel gains from the same MBS (e.g., a low channel gain from the shared and a high channel gain from the dedicated channels), while in reality a user should see fairly similar gains across different channels of a BS, particularly, if the channels are assumed to be flat (which is the case in the paper). Apart from creating possible inaccuracies in the numerical results, this modelling approach undermines the importance of power apportioning on the two types of macro channels. For example, in the same work the authors consider a SC power budget on the shared channels and show that such a power allocation scheme works well with their modelling approach. However, this scheme will possibly no longer work if the same channel model was used on all channels. This is because MBSs typically have a much lower path loss and a much higher antenna gain compared to SCs and allocating the same amount of power budget as SCs' on the MBS shared channels may result in excessive interference on the neighbouring SCs. Therefore, we believe that the previous optimisation framework needs to be revisited with the right channel model along with its possible implications and that the performance gains of PSD over other RA schemes (as discussed in [14]) should be re-examined. Another restrictive assumption in [14] is that the authors confine the macro users to being scheduled on either shared or dedicated channels but not both. Such a restriction on user scheduling eliminates the potential gains, if any, obtained from joint scheduling of macro users on both types of macro channels.

In this chapter, our fist objective is to improve upon the work in [14] by deploying the

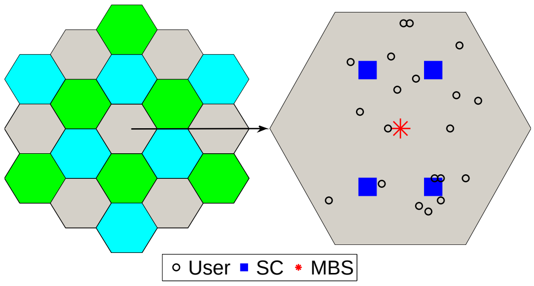


Figure 3.1: A multi-cell HetNet with a reuse factor of 3

right channel model at the MBS, i.e., by using the MBS channel model for all channels (both shared and dedicated) at the MBS. As we will see shortly, deploying the new channel model entails a detailed study of optimal apportioning of the MBS power budget on shared and dedicated channels which is missing from the previous work in [14]. Our second objective is to address the impact of joint scheduling of macro users on both types of the MBS channels and quantify the possible gains obtained from it.

In the following, we provide the details of the unified optimisation framework under PSD and address the required changes that need to be made.

3.1 System overview

We consider a cellular network comprising a set of macro cells as shown in Fig. 3.1. Each macro cell, in addition to a centrally placed MBS, has B low-power BSs making B SCs[1]. These SCs are connected to the MBS and the MBS is connected to the network core, all via wired backhaul links of infinite capacity.

3.2 Scope

We consider each macro cellular area, with its MBS, B SCs, and U users as a standalone HetNet system, and we optimise a number of network processes (RA, UA, US) within

[1]Note that the SCs are not always contained within a macro cell. For example, they can be placed at the coverage intersection of two macro cells. However, in this thesis, we only deal with the SCs contained within a macro cell.

the scope of such a single macro cellular area only. However, as we will show shortly our physical-layer SINR formulation allows us to take into account the inter and intra-cell interference coming from neighbouring macro and small cells. This way of decoupling a multi-cell HetNet at a macro cell level will eliminate the need for inter-macro cell coordination and, hence, greatly simplify the task of RRM. However, it may come with some penalty regarding system performance.

3.3 Main features of the model

Below, we outline the features that are incorporated in our optimisation framework.

1. **Different RA schemes:** The RA schemes involve channel assignment to different BSs and are considered as a natural mechanism for interference co-ordination between neighbouring BSs. Our framework is able to incorporate two RA schemes, i.e., OD and PSD, and characterise the network performance based on each of these schemes.

2. **Different UA schemes:** Various UA schemes have been proposed in the literature each with the objective of improving the network performance in terms of different criteria (e.g., best SINR, load-balancing, fairness criteria). Our optimisation framework is able to incorporate many UA schemes, e.g., Max-SINR, SCF and CRE, and can characterise the network performance based on each of these schemes.

3. **User scheduling:** US is typically a local and quasi real-time process of allocating RBs to the users associated with a BS according to some fairness criteria. Our optimisation framework is not only able to incorporate various US heuristics, but also provide globally (to the cell under consideration) optimal schedules for a given RA and UA scheme.

4. **Power apportioning:** Since power apportioning on the MBS is one of the main objectives of this study, our model is able to flexibly incorporate power apportioning between different channels and capture its affect on the network performance.

In the following, we describe our system model which encompasses all the above-mentioned features as well as their complex interplay into a unified framework and allows us to characterise the performance of the system.

3.4 System model

Our optimisation framework is based on a snapshot model. Under a given snapshot (or realisation), the cardinality and locations of BSs and users are assumed to be fixed and known. Each realisation corresponds to the random distribution of a fixed number of users in each cell in the HetNet according to a specific distribution. Note that because of the random distribution of the users, the channel gains change from one realisation to another. We assume all users in the network are active and greedy in the sense that there is an infinite backlog of packets for each user and that the users want to maximise their individual throughputs. We consider one cell in an OFDM-based multi-cell HetNet. The cell consists of a MBS, indexed by 0, overlaid by a set of B SCs, $\mathcal{B} = \{1, 2, ..., B\}$, indexed by j. Each SC is connected to the MBS via a wired backhaul link of infinite capacity. There are M sub-channels of bandwidth b available in the cell to serve a set of U users, indexed by i, on the downlink. We denote the set of users in realisation ω by $\mathcal{U}(\omega) = \{1, 2, ..., U\}$. The per-channel-use transmit power of the MBS and each SC is denoted by P_M and P_S, respectively, both of which are fixed and known. Throughout this thesis, we make the following assumptions to simplify our problem formulations.

1. Each BS transmits on all the channels allotted to it at a given time.
2. Users are only allowed to associate with one BS (SUA).
3. The BSs as well as the users are each equipped with one omni-directional antenna.

The HetNet as a whole uses M' OFDM sub-channels and each macro cell is allocated $M = \frac{M'}{r}$ sub-channels, where $r > 1$ is the reuse factor. Hence, a total of M OFDM sub-channels are available for the macro cell under consideration (i.e., to be used by the MBS in the middle of Fig. 3.1 and its B SCs).

3.4.1 Physical-layer characterisation

Under PSD, we assume that $k \in \{1, 2, ..., M\}$ sub-channels are jointly used between the MBS and SCs and $M - k$ sub-channels are dedicated to the MBS. We denote the MBS by $0_M, 0_S$ respectively when transmitting on the dedicated and shared channels. The respective per sub-channel transmit powers used by 0_M and 0_S are $\frac{P_M-\hat{P}_S}{M-k}$ and $\frac{\hat{P}_S}{k}$ where $\hat{P}_S$ is the power budget of the shared channels at the MBS. The SCs transmit with a power budget of $\frac{P_S}{k}$ per sub-channel. We assume that every BS in the HetNet transmits on all the channels allotted to it at all times.

Let g_{ji}, γ_{ji} represent the downlink channel gain and SINR from BS j to user i. The channel gain accounts for the antenna gain, path loss and slow fading. The slow fading at each location is modelled as a shadowing affect, with a log-normal pdf using a standard deviation recommended in [2]. We assume that the channel gains are time-invariant. Furthermore, we assume that the channels are flat and known to each BS through independent and perfect uplink feedback channels, allowing computation of SINRs at the BS. Every BS uses a prescribed function, $f(.)$, which maps SINR to efficiency in bits/symbol (see Table 3.2). For a given k and realisation ω,

$$\gamma_{ji} = \frac{P_j g_{ji}}{N_0 + \mathbf{1}_{\{j \in \mathcal{B} \cup \{0_S\}\}} \sum_{l \in \mathcal{B} \cup \{0_S\}, l \neq j} P_l g_{li} + I_{ji}}, \ \forall (i \in \mathcal{U}(\omega), j \in \mathcal{B}\{0_M, 0_S\}), \tag{3.1}$$

where $g_{0_S i} = g_{0_M i}$, $\forall i \in \mathcal{U}(\omega)$, I_{ji} is the interference from the neighbouring BSs transmitting on the same sub-channels as BS j and N_0 is the additive white Gaussian noise (AWGN) power. For simplicity we assume that the RA and power allocation patterns are repeated with some reuse factor in every macro cell in the HetNet and all cells are fully synchronised both in time and frequency domains. With this, the interference terms in (3.1) can be written as

$$\begin{aligned} I_{0_M i} &= \sum_{q \in \mathcal{M}} \left(\frac{P_M - \hat{P}_S}{M-k} \right) g_{qi}, \quad \forall i \in \mathcal{U}(\omega), \\ I_{ji} &= \sum_{r \in \mathcal{P}} P_r g_{ri}, \quad \forall (i \in \mathcal{U}(\omega), j \in \mathcal{B} \cup \{0_S\}), \end{aligned}$$

where $\mathcal{M}$ is the set of the MBSs in the HetNet that use the same set of channels as 0_M and $\mathcal{P}$ is the set of the MBSs and SCs in the HetNet that use the same set of channels as 0_S. Let r_{ji} be the user i's (maximum achievable) link rate from BS $j \in \mathcal{B} \cup \{0_S, 0_M\}$ (in bits per second). Then,

$$\begin{aligned} r_{0_M i} &= (M-k)b \times f(\gamma_{0_M i}), \quad \forall i \in \mathcal{U}(\omega), \\ r_{ji} &= kb \times f(\gamma_{ji}), \quad \forall (i \in \mathcal{U}(\omega), j \in \mathcal{B} \cup \{0_S\}). \end{aligned}$$

Note that r_{ji} is available only if user i is the only user associated with BS $j \in \mathcal{B} \cup \{0_M, 0_S\}$.

3.4.2 Utility functions and fairness criterion

Our goal is to perform joint RA, UA and US under PSD such that a global fairness in throughput assignment to users is guaranteed across a multi-cell HetNet. However, with our assumption that all BSs constantly transmit and assign equal power to each of their sub-channels, the interference terms and, hence, SINR and link rates become fixed. Therefore, the global problem can be decoupled into a set of independent per macro cell problems. We will henceforth refer to the joint RA, UA and US problem associated with one macro cell (i.e., the macro cell under consideration) in a multi-cell HetNet as the global (to the macro cell) problem.

A common global fairness objective which has been extensively used in the literature for throughput assignment to users is the sum of the $\alpha-$fair utility functions. If T_i is the throughput offered to user i, the utility corresponding to this allocation is given by

$$U_\alpha(T_i) = \begin{cases} \frac{{T_i}^{1-\alpha}}{1-\alpha} & \text{if } \alpha \geq 0 \text{ and } \alpha \neq 1 \\ \log(T_i) & \text{if } \alpha = 1 \end{cases}$$

The global fairness objective function is, then, $\sum_{i \in \mathcal{U}(\omega)} U_\alpha(T_i)$. Problems that maximise functions of utilities are typically known as *network utility maximisation problems.* Note that as $\alpha \to 0$, the objective function is efficient but not fair; as $\alpha \to \infty$, poorer users are favoured and at $\alpha = 1$, it is PF. Hence, higher values of α imply better user throughput-fairness at the cost of users' aggregate throughput.

Furthermore, for different values of α, we need to have different comparison metrics. Towards this, we can show that maximising the sum of the α-fair utility is equivalent to maximising the following throughput-based metric (see [15, Chapter 6]).

$$\bar{T}_\alpha(\{T_i\}_{i \in \mathcal{U}(\omega)}) = \begin{cases} \left(\frac{1}{|\mathcal{U}(\omega)|} \sum_{i \in \mathcal{U}(\omega)} T_i^{1-\alpha}\right)^{\frac{1}{1-\alpha}}, \ \alpha \geq 0, \alpha \neq 1, \\ \left(\prod_{i \in \mathcal{U}(\omega)} T_i\right)^{\frac{1}{|\mathcal{U}(\omega)|}}, \ \alpha = 1. \end{cases} \tag{3.3}$$

We will refer to $\bar{T}_\alpha(.)$ simply as the *α-mean throughput.* Note that for PF, this metric $\bar{T}_1(.)$ represents the geometric mean (GM) of user throughputs.

Now, we are ready to define the unified optimisation framework under PSD with the objective of maximising the global $\alpha-$mean throughput of the users within the cell under consideration.

3.5 Joint RA, UA and US problem under PSD: A unified optimisation framework

Let β_{ji} be the fraction of the time that user i is scheduled on BS j and x_{ji} be the binary UA variable, i.e., x_{ji} is 1 if user i is associated with BS j and is 0 otherwise. Given a network realisation ω, link rates r_{ji} and RA parameters $(k, \hat{P}_S)$, the $\alpha-$mean throughput maximisation problem under PSD can be written as follows.

$$\begin{aligned}
\textbf{[PSD-0}(k, \hat{P}_S)\textbf{]:} \max_{\{\beta_{ji}, x_{ji}\}} & \sum_{i \in \mathcal{U}(\omega)} U_\alpha \left(\sum_{j \in \mathcal{B} \cup \{0_M, 0_S\}} r_{ji} \beta_{ji} \right) \\
\text{s.t.} & \sum_{i \in \mathcal{U}(\omega)} \beta_{ji} \leq 1, \ \forall j \in \mathcal{B} \cup \{0_M, 0_S\}, && \text{(3.4a)} \\
& \sum_{j \in \mathcal{B} \cup \{0_M\}} x_{ji} = 1, \ \forall i \in \mathcal{U}(\omega), && \text{(3.4b)} \\
& 0 \leq \beta_{ji} \leq x_{ji}, \ \forall (i \in \mathcal{U}(\omega), j \in \mathcal{B} \cup \{0_M, 0_S\}), && \text{(3.4c)} \\
& x_{0_M i} = x_{0_S i}, \ \forall i \in \mathcal{U}(\omega), && \text{(3.4d)} \\
& x_{ji} \in \{0, 1\}, \ \forall (i \in \mathcal{U}(\omega), j \in \mathcal{B} \cup \{0_M, 0_S\}). && \text{(3.4e)}
\end{aligned}$$

Note that the variables $\hat{P}_S$ and k are implicitly included in the link rates r_{ji} in the formulation. (3.4a) represents the scheduling constraints at each BS. (3.4b) represents the SUA constraint. (3.4c) restricts the scheduling to the associated BS-user pairs. (3.4d) implies that a macro user can be scheduled on both shared and dedicated channels. (3.4e) represents the integrality constraint of UA variables.

Remark 1. *The problem in (3.4) is parametrised with $(k, \hat{P}_S)$. In order to solve the problem, the parameters have to be chosen and fixed. Therefore, a joint optimal RA, UA, and US can be obtained by solving a set of parametrised problems*[2] *to find the optimal model parameters:*

$$\arg \max_{\{\hat{P}_S, k\}} \textit{\textbf{PSD-0}}(k, \hat{P}_S).$$

[2]Note that, we use the symbol $\mathbf{A}$ to represent the optimal value (i.e., the value of the objective function when the variables are chosen optimally) of problem [$\mathbf{A}$].

3.5.1 Problem approximation and analysis

The problem [$\textbf{PSD-0}(k, \hat{P}_S)$] allows the MBS to schedule a user on both shared and dedicated channels which can complicate the scheduling at the MBS. To simplify the scheduling, we only allow the MBS to schedule a given user on either shared or dedicated channels but not both. As we will verify in section 3.6, such a restriction will only result in a negligible loss in the objective function. Hence, we approximate [$\textbf{PSD-0}(k, \hat{P}_S)$] by

$$\begin{aligned}[\textbf{PSD-1}(k, \hat{P}_S)]\textbf{:} \max_{\{\beta_{ji}, x_{ji}\}} & \sum_{i \in \mathcal{U}(\omega)} U_\alpha \left(\sum_{j \in \mathcal{B} \cup \{0_M, 0_S\}} r_{ji}\beta_{ji} \right) \\ \text{s.t. } & (3.4a), (3.4c), (3.4e), \\ & \sum_{j \in \mathcal{B} \cup \{0_M, 0_S\}} x_{ji} = 1, \ \forall i \in \mathcal{U}(\omega). \end{aligned} \tag{3.5}$$

The problem above is an integer (and non-linear if $\alpha \neq 0$) program and solves for optimal UA and US when channel and power allocation parameters, $(k, \hat{P}_S)$, are fixed.

Remark 2. *[15, Theorem 2] If all x_{ji}'s (the UA rule) are given, a) Decomposition: The global problem (3.5) can be decoupled into a set of $|\mathcal{B} \cup \{0_M, 0_S\}|$ independent local α-fair problems, one for each BS, where the local problem for BS j is*

$$\begin{aligned}[P^j_{Local}] : \max_{\{\beta_{ji} \geq 0\}_{i \in \mathcal{U}_j(\omega)}} & \sum_{i \in \mathcal{U}_j(\omega)} U_\alpha(r_{ji}\beta_{ji}) \\ s.t. & \sum_{i \in \mathcal{U}_j(\omega)} \beta_{ji} \leq 1. \end{aligned}$$

b) Closed-form solution: The following schedule is optimal for the local problem $[P^j_{Local}]$.

$$\beta_{ji} = \frac{r_{ji}^{\frac{1-\alpha}{\alpha}}}{\sum_{i' \in \mathcal{U}_j(\omega)} r_{ji'}^{\frac{1-\alpha}{\alpha}}},$$

where $\mathcal{U}_j(\omega)$ is the set of users associated with BS j at realisation ω.

Remark 2 suggests that by fine-tuning the parameters $(k, \hat{P}_S)$ and choosing a good UA scheme, we can achieve a good $\alpha-$mean throughput performance using an optimal formula-based scheduler under PSD. It is noteworthy that the scheduler is extremely simple for $\alpha = 1$, as the optimal scheduling reduces to equal-time sharing (i.e., Round Robin).

3.5.2 User association

Solving [**PSD-1**$(k, \hat{P}_S)$] for optimal x_{ji}'s yields optimal user association for a fixed $(k, \hat{P}_S)$ which can be used as a benchmark to evaluate the performance of simple UA schemes. We study three different simple but sub-optimal UA schemes which are based on simple rules that a user can use to perform its association decision. We will consider Max-SINR, Small-cell First (SCF) [14] and Cell Range Expansion (CRE) [1] as representatives of such schemes. Formally, under PSD we define these UA schemes as follows.

1. **Max-SINR:** A user i associates with BS j^* if $j^* = \arg\max_{j\in\mathcal{B}\cup\{0_M,0_S\}} \gamma_{ji}$.
2. **SCF:** A user i associates with BS j^* if $j^* = \arg\max_{j\in\mathcal{B}\cup\{0_S\}} \gamma_{ji}$ and $\gamma_{j^*i} > \delta$; otherwise, $j^* = 0_M$.
3. **CRE:** A user i associates with SC j^* if $j^* = \arg\max_{j\in\mathcal{B}\cup\{0_S\}} \gamma_{ji}$ and $\gamma_{j^*i} + \epsilon > \gamma_{0_M i}$; otherwise, $j^* = 0_M$.

δ (in dB) and ϵ (in dB) are two configurable parameters, also called the *biasing* parameters, associated with SCF and CRE, respectively, and allow us to configure an association bias in favour of SCs. We assume that these values are fixed and known to all users and BSs a priori when using SCF or CRE. Furthermore, we assume that the biasing parameter δ (ϵ) is equal for all SCs when using SCF (CRE).

All of the above-mentioned rules are simple in the sense that they do not involve any real-time load-balancing and are easy to implement (since each user can do it individually). They also provide feasible SUA solutions and, hence, provide lower-bounds for the optimal SUA solution (which can be obtained by solving the problem in (3.4) for optimal x_{ji}'s). Studying these UA schemes helps us understand how simple association schemes perform compared to the optimal UA. In the absence of power apportioning at the MBS, [14] already shows that SCF works well. Our study allows us to see whether this observation extends to the case of PSD with power apportioning.

3.6 Numerical results

3.6.1 Parameter settings

We consider the middle cell in a 19-cell HetNet with a reuse factor of $r = 3$ as shown in Fig. 3.2. All lengths unless specified are in meters (m). Each cell, with a radius

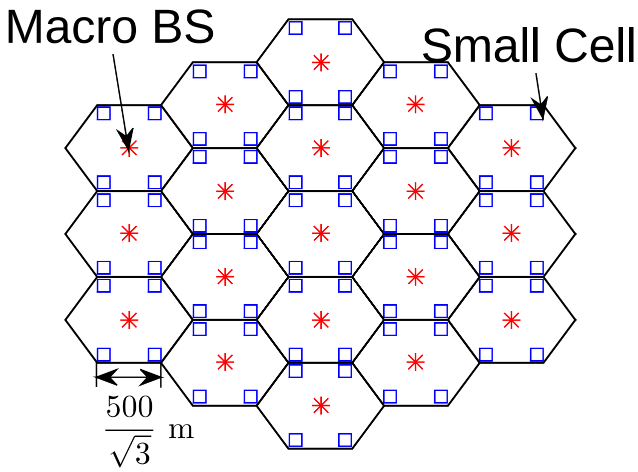


Figure 3.2: The HetNet configuration with 4 SCs per macro cell

Table 3.1: Physical-layer parameters

Noise power	-174 $\frac{\text{dBm}}{\text{Hz}}$	T_{subframe}	1 ms
P_S	30 dBm	P_M	46 dBm
MBS ant. gain	15 dBi	SC ant. gain	5 dBi
User ant. gain	0 dBi	Sub-channel Bandwidth	180 KHz
Shadowing s.d.	8 dB	Penetration loss	20 dB
SC_{OFDM}	12	SY_{OFDM}	14
SC-user path loss	$140.7 + 36.7\log_{10}(d/1000),\ d \geq 10$ m $140.7 + 36.7\log_{10}(10/1000),\ d < 10$ m		
MBS-user path loss	$128 + 37.6\log_{10}(d/1000),\ d \geq 35$ m $128 + 37.6\log_{10}(35/1000),\ d < 35$ m		

of $500/\sqrt{3}$ m, consists of one MBS, located in the centre, and four SCs located $\Delta = 230$ m away from the MBS. Each cell has $M = 33$ sub-channels available for downlink transmission. The set of user locations are restricted to a grid composed of 1387 uniformly distanced points within each cell of the HetNet. We assume that each user receives a non-zero link rate from at least one BS, i.e., there are no uncovered locations in the system.

The physical-layer parameters are based on the 3GPP evaluation methodology document [2] used for HetNets in LTE. These parameters are reproduced here for ease of

Table 3.2: Modulation and coding schemes - LTE

SINR thresholds (in dB)	-6.5	-4	-2.6	-1	1	3	6.6	10	11.4	11.8	13	13.8	15.6	16.8	17.6
Efficiency (in bits/symbol)	0.15	0.23	0.38	0.60	0.88	1.18	1.48	1.91	2.41	2.73	3.32	3.90	4.52	5.12	5.55

reading. The SINR model is given in equation (3.1). The channel gain g_{ji} accounts for antenna gain, path loss and slow fading. Following [14], users experience log-normal shadowing with standard deviation of 8 dB. The path loss for the SC and macro users at a distance of d from the corresponding BS are given in the last two lines of Table 3.1.

We assume that the system uses adaptive modulation and coding with discrete rates. Table 3.2 taken from [26] gives us the mapping between the SINR and efficiency (in bits/symbol) per sub-carrier for the modulation and coding schemes under the LTE framework. The bit rate obtained by a user that has an SINR between level ℓ and level $\ell + 1$ is $r = \frac{\mathrm{SC_{ofdm}} \times \mathrm{SY_{ofdm}}}{T_{\mathrm{subframe}}} e_\ell$ where e_ℓ is the efficiency (bits/symbol) of the corresponding level ℓ, $\mathrm{SC_{ofdm}}$ is the number of data sub-carriers per sub-channel bandwidth, $\mathrm{SY_{ofdm}}$ is the number of OFDM symbols per subframe, and T_{subframe} is the subframe duration in time units. The value of these parameters are shown in Table 3.1.

3.6.2 Relaxation and optimality gap

The optimisation problems in (3.4) and (3.5) are integer (and non-linear if $\alpha \neq 0$) programs which are typically hard to solve quickly with the existing integer program solvers (e.g., Bonmin) even for a relatively small set of variables. For this reason, we convert the two integer programs into continuous convex optimisation problems by allowing fractional user association (FUA), i.e., relaxing the integrality constraints (i.e., $x_{ji} \in [0, 1]$). These relaxed problems, termed as *PSD-FUAWithJointSched* and *PSD-FUA*, clearly provide upper-bounds to their corresponding integer problems specified in (3.4) and (3.5). We obtain these upper-bounds by solving the relaxed convex optimisation problems for each realisation using the commercial solver, *Minos 5.51*. To see that the relaxations are tight, we obtain a feasible set of solution, $\{x_{ji} \in \{0, 1\}\}_{j,i}$, from the *PSD-FUA* problem as follows. For user i, choose the BS $j_i^* = \arg\max_{j \in \mathcal{B} \cup \{0_S, 0_M\}} r_{ji}\beta_{ji}$ and break the ties in favour of the SCs. We refer to the *PSD-FUA* problem with this feasible SUA solution as *PSD-SUA*, since it provides a lower-bound for the optimal solution to the SUA problem in (3.5). Similarly, we define *OD-FUA* and *OD-SUA* under OD. Recall that OD is a special case of PSD where $\hat{P}_S = 0$.

3.6.3 The static setting

As stated in Section 3.4, we consider a snapshot model for our optimisation framework. We assume a set of $U = 20$ users are distributed i.i.d. uniformly on the grid locations in the cell. We generate a set of 100 realisations Ω (i.e., $|\Omega| = 100$) and, for each realisation $\omega \in \Omega$, compute a solution to the *PSD-FUAWithJointSched*, *PSD-FUA* and *PSD-SUA* (as discussed in Section 3.6.2) for each value of $k \in \{1, 2, ..., 32\}$ and $\hat{P}_S \in \{20, 15, 10, 5, 3, 0, -1, -10\}$ dBm. For each problem, we compute the (average) $\alpha-$mean throughput corresponding to that problem over the 100 realisations and plot it as a function of k for $\alpha \in \{1, 2\}$.

3.6.4 Validation of the upper-bounds

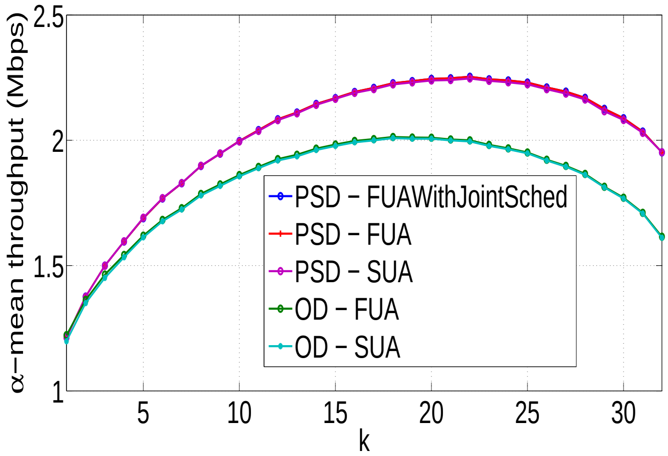
Fig. 3.3 shows the average $\alpha-$mean throughput of the users versus k corresponding to the *PSD-FUAWithJointSched*, *PSD-FUA* and *PSD-SUA* problems. The $\hat{P}_S$ is set to 3 dBm for all the curves in the figure, since our numerical experiments suggested that this value performs near-optimally for all values of k for all three problems (and, hence, we only include the results corresponding to $\hat{P}_S = 3$ dBm in the figure). A close examination of the curves reveals that the *PSD-FUAWithJointSched*, *PSD-FUA* and *PSD-SUA* perform very closely for all values of k. Therefore, we conclude that the problem in (3.4) can be with high accuracy approximated by (3.5). Indeed, the exact solution of (3.4) is somewhere in between *PSD-FUAWithJointSched* and *PSD-SUA*. However, since the performance gap between the two is reasonably negligible and US using the model in (3.5) is simpler,

> *...from now on we will only consider the joint optimisation model in* (3.5) *whenever dealing with PSD.*

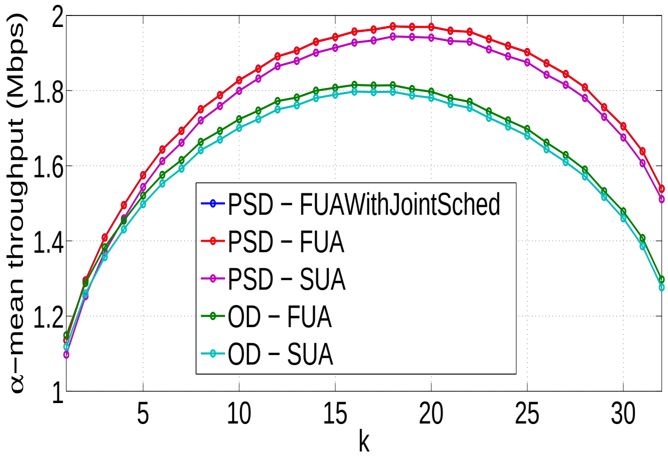
Furthermore, the figure shows that, by deploying the right power apportioning at the MBS, *PSD-FUA* outperforms *OD-FUA* by 11.94% and 8.83% for $\alpha = 1$ and 2, respectively.

3.6.5 The impact of power apportioning at the MBS

Fig. 3.4 shows the average $\alpha-$mean throughput corresponding to *PSD-FUA* for different values of $\hat{P}_S$ as a function of k. The figure suggests that the right choice of $\hat{P}_S$ does significantly impact the $\alpha-$mean throughput of the system. As can be seen, a much lower amount of power compared to the SCs power budget on the MBS shared channel

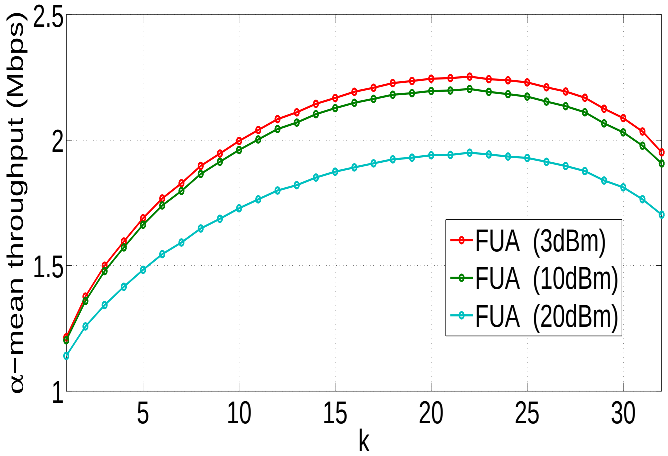


(a) $\alpha = 1$

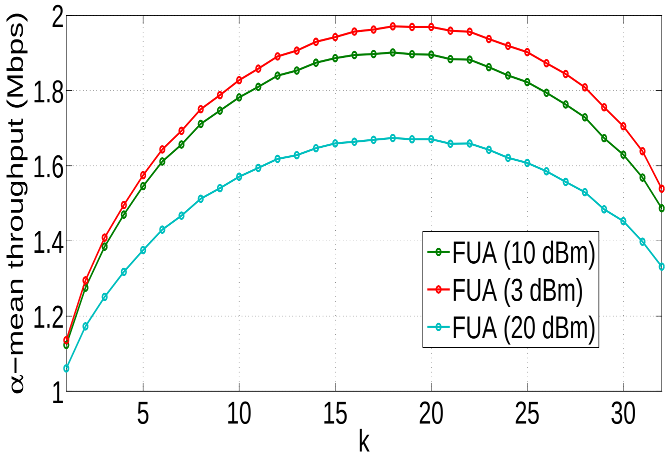


(b) $\alpha = 2$

Figure 3.3: $\alpha-$mean throughput as a function of k. Settings: $U = 20$ and $M = 33$.



(a) $\alpha = 1$



(b) $\alpha = 2$

Figure 3.4: $\alpha-$mean throughput as a function of k. Settings: $U = 20$ and $M = 33$.

is preferable. This can be justified by the fact that the much higher antenna gain and lower path loss of the MBS compensate for the low value of $\hat{P}_S$ without causing excessive interference on the neighbouring SCs. It is noteworthy that our numerical experiments suggested that the optimal value of $\hat{P}_S$ (i.e., 3 dBm) is almost invariant for different values of α, U, M, Δ and P considered in this thesis. Therefore,

> *...from now on, we set $\hat{P}_S$ to 3 dBm in all figures related to PSD unless otherwise specified.*

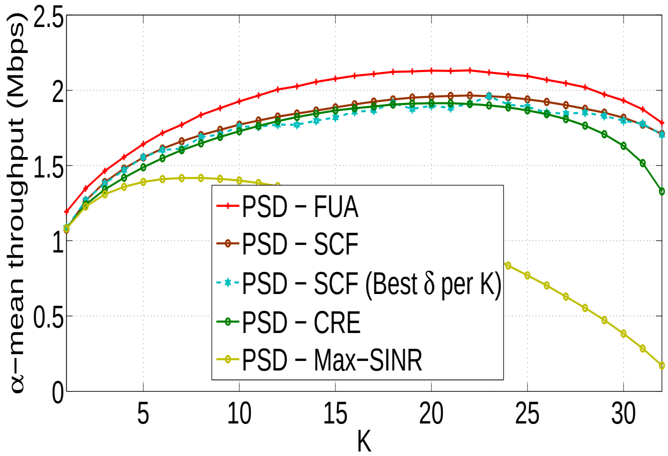
3.6.6 Comparison between different simple UA schemes

Fig. 3.5 shows the average $\alpha-$mean throughput corresponding to simple UA rules (i.e., Max-SINR, SCF, CRE) as a function of k. SCF and CRE have a biasing parameter δ and ϵ, respectively. We assume that δ can take any one of the SINR threshold values shown in Table 3.2 and that ϵ can take any of the SINR threshold values from the set $\{2i|i \in \mathcal{Z}, 0 \leq i \leq 10\}$ dB. For *PSD-SCF*, the value of δ in each realisation is selected so that the $\alpha-$mean throughput of users is maximised in the given realisation. The curves suggest that by fine-tuning the parameters $(k, \hat{P}_S, \delta)$ for each realisation, *PSD-SCF* can perform very closely to the upper-bound, i.e., *PSD-FUA*.

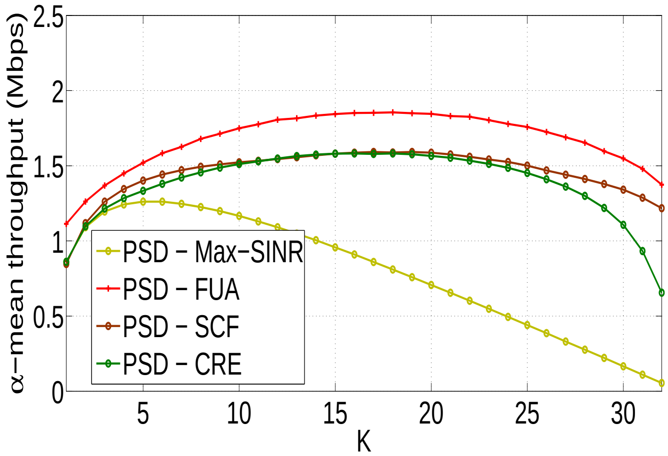
Since it may not be a feasible option to vary δ too frequently in a dynamic system, we also plot the average $\alpha-$mean throughput per k for the best value of δ. To do so, for each k, we compute the average $\alpha-$mean throughput over the 100 realisations for each value of δ and select the δ that yields the best average $\alpha-$mean throughput. We refer to this as *PSD-SCF (Best δ per k)* in the figures. Similarly, we define *PSD-CRE* and *PSD-CRE (Best ϵ per k)*. The results show that fine-tuning the UA parameters δ per k for SCF results in roughly the same $\alpha-$mean throughput as when selecting the optimal δ per realisation (which otherwise can be a bottleneck in a dynamic network). Lastly, the results suggest that *PSD-SCF* generally performs better than *PSD-CRE* for both values of α and significantly better than Max-SINR.

3.7 Conclusions

In conclusion, we revisited the problem of joint RA, UA and US under PSD in a static setting, first discussed in [14] where the authors make a wrong assumption on the shared



(a) PSD - $\alpha = 1$



(b) PSD - $\alpha = 2$

Figure 3.5: $\alpha-$mean throughput as a function of k and θ. Settings: $U = 20$, $M = 33$.

channels model at the MBS. We improved upon this work by deploying the right model for the shared channels at the MBS, i.e., by using the MBS channel model for the shared channels. The correct modelling of the channels raised two questions regarding the 1) joint scheduling of the macro users and 2) power apportioning on the shared and dedicated channels at the MBS. First, we showed that there was not much gain obtained from scheduling macro users on both types of the channels which consequently allowed us to decompose the local MBS joint scheduling problem into two independent problems; one for the users on the dedicated channels and one for the users on the shared channels. We showed that deploying a simple formula-based scheduler for each of the two sets of macro users (independent of the other set of macro users) results in a near-optimal throughput performance, while significantly simplifying the scheduling. Finally, we addressed the importance of power apportioning between the shared and dedicated channels at the MBS and showed that the right choice of $\hat{P}_S$ could significantly impact the system's $\alpha-$mean throughput performance. Luckily, our numerical results indicated that the optimal $\hat{P}_S$ was robust to different HetNet configurations and choices of α for all values of k. Lastly, our results showed that PSD could significantly outperform OD by fine-tuning $\hat{P}_S$.

Chapter 4

A Thorough Comparison Between PSD and ABS: A Static Setting

Summary: In this chapter, we

- revisit the joint RA, UA and US optimisation framework under ABS based on the snapshot model in a static setting,
- show that the scheduling problem for ABS can be NP-hard for $\alpha \neq 1$, and
- using numerical results, compare the best PSD scheme to the best ABS scheme and show that PSD outperforms ABS, under various network configurations and heuristics.

4.1 Introduction

In Chapter 3, we studied a unified optimisation framework, encompassing the three network processes, based on a snapshot model under PSD. We obtained tight upper-bounds for $\alpha-$mean throughput of a cell in a multi-cell HetNet for different values of α and saw that SCF association scheme along with an optimal formula-based $\alpha-$fair scheduler can perform very well if the RA and biasing parameters are fine-tuned. Furthermore, we saw that PSD can outperform OD if the MBS power budget is carefully apportioned between the shared and dedicated channels. Based on our offline studies so far, PSD seems to be a strong contender among the 3GPP's proposed RA schemes both in terms of ease of implementation and throughput gains. In this chapter, we compare PSD with another 3GPP proposed RA scheme, i.e., (CCD with) ABS. Our objective is to conduct a comparative study between the two RA schemes both in terms of ease of implementation and performance gains. Towards this, we revisit a unified framework similar to that in Chapter 3 but under ABS. The framework under consideration, first discussed in [6], studies the joint optimisation of RA, UA and US under ABS and a global PF criteria based on a static model. The authors show that, under some mild assumptions, the global PF scheduling problem decomposes into a set of independent local (per-BS) scheduling problems each of which can be solved to optimality based on a simple threshold-based algorithm. Building upon their work, our first objective in this chapter is to consider the unified framework under a global α-fairness criteria and comment on the tractability of the global scheduler for $\alpha \neq 1$. Our second objective is to conduct a thorough comparison between ABS and PSD in terms of throughput performance under different HetNet configurations and heuristics.

4.2 System model

We consider a snapshot model similar to the one described in Section. 3.4 and use the same notations and assumptions unless otherwise specified.

4.2.1 Physical-layer characterisation

Under ABS, we assume that all of the M sub-channels are shared between the MBS and SCs. We denote the MBS by 0. The respective per sub-channel transmit power used by the MBS is $\frac{P_M}{M}$ and 0 during the non-ABS and ABS duty cycles[1]. The per sub-channel transmit power used by the SCs is $\frac{P_S}{M}$ during both cycles. We assume that the SCs transmit at all times while the MBSs transmit for only $0 \leq \theta \leq 1$ fraction of the time. We refer to θ as the RA parameter under ABS. Given the parameter θ and a network realisation ω, the per sub-channel SINR from BS j to user i can be written as

$$\gamma_{ji} = \begin{cases} \frac{P_j g_{ji}}{N_0 \, + \, \sum_{l \in \mathcal{B} \cup \{0\}, l \neq j} P_l g_{li} + I_{Mi} + I_{Si}} & \text{non-ABS cycle,} \\ \frac{P_j g_{ji}}{N_0 + \sum_{l \in \mathcal{B}, l \neq j} P_l g_{li} + I_{Si}} & \text{ABS cycle,} \end{cases} \tag{4.1}$$

[1]Note that we assume the MBS is completely mute during the ABS duty cycle.

where $I_{Mi} = \sum_{m\in\mathcal{M}} \left(\frac{P_M}{M}\right) g_{mi}$ is the co-channel interference from the neighbouring MBSs and $I_{Si} = \sum_{q\in\mathcal{P}} \left(\frac{P_S}{M}\right) g_{qi}$ is the co-channel interference from the neighbouring SCs. Note that γ_{ji} is a function of θ and takes two values depending on whether the BS j is transmitting during ABS or non-ABS to user i. Clearly, the value of γ_{ji} for the SCs during ABS is at least as good as that during non-ABS because of the reduced interference as a result of muting the MBSs in the system. We denote the (maximum achievable) link rates seen by user i from BS j with $r_{ji}, \tilde{r}_{ji}$ (in bits per second), during non-ABS and ABS respectively, which can be computed by $M \times b \times f\left(\gamma_{ji}\right)$.

4.3 Joint RA, UA and US problem under ABS: A unified framework

As before, we use the sum of the $\alpha-$fair utility functions as our global fairness criteria. Let β_{ji} and $\tilde{\beta}_{ji}$ be the fraction of time that user i is scheduled on BS j during non-ABS and ABS duty cycles, respectively, and x_{ji} be the binary UA variable. Given a network realisation ω, rates $r_{ji}(\omega)$, and the RA parameter θ, the global $\alpha-$mean throughput maximisation problem under ABS can be written as follows [6].

$$\begin{aligned}
\textbf{[ABS}(\theta)\textbf{]:} \max_{\{\tilde{\beta}_{ji},\beta_{ji},x_{ji}\}} & \sum_{i\in\mathcal{U}(\omega)} U_\alpha\left(\sum_{j\in\mathcal{B}\cup\{0\}} (\beta_{ji}r_{ji} + \tilde{\beta}_{ji}\tilde{r}_{ji})\right) \\
\text{s.t.} & \sum_{i\in\mathcal{U}(\omega)} \beta_{ji} \leq \theta, \quad \forall j \in \mathcal{B}\cup\{0\}, \\
& \sum_{i\in\mathcal{U}(\omega)} \tilde{\beta}_{ji} \leq 1-\theta, \quad \forall j \in \mathcal{B}, \\
& \sum_{j\in\mathcal{B}\cup\{0\}} x_{ji} = 1, \quad \forall i \in \mathcal{U}(\omega), \\
& 0 \leq \beta_{ji}, \tilde{\beta}_{ji} \leq x_{ji}, \ \forall (i\in\mathcal{U}(\omega), j \in \mathcal{B}\cup\{0\}), \\
& x_{ji} \in \{0,1\}, \ \forall (i\in\mathcal{U}(\omega), j\in\mathcal{B}\cup\{0\}).
\end{aligned} \tag{4.2}$$

Note that $\tilde{\beta}_{0i} = 0$ (since the MBS is off during the ABS duty cycle). The first two constraints apportion the time that each BS transmits to its associated users when the MBS is on and off, respectively. The last three constraints are identical to their PSD counterparts.

The problem above is an integer (and non-linear if $\alpha \neq 0$) program and solves for optimal UA and US when the RA parameter, θ, is given.

Remark 3. *The problem in (4.2) is parametrised with θ. In order to solve the problem, the parameter has to be chosen and fixed. A joint optimal RA, UA, and US can, then, be obtained by solving a set of parametrised problems to find the optimal parameter:*

$$\arg\max_{\theta} \boldsymbol{ABS(\theta)}.$$

Remark 4. *[6] For $\alpha = 1$, if all x_{ji}'s (the UA rule) are given: a) Decomposition: The global problem (4.2) can be decoupled into a set of $|\mathcal{B} \cup \{0\}|$ independent local (per-BS) PF problems, and b) Near-optimal algorithm: The following algorithm yields a near-optimal schedule for any of the local problems. Each BS will arrange its users in a descending order based on the ratios $r_i/\tilde{r}_i$ and split the users into two groups. The group containing the users with higher $r_i/\tilde{r}_i$ ratios will be scheduled during non-ABS and the rest of the users during ABS. The users in each group will receive equal service time.*

Remark 4 suggests that by fine-tuning the parameter θ and choosing a good UA rule, we can achieve a good $\alpha-$mean throughput performance using a simple scheduler that scales linearly with the number of users for $\alpha = 1$. In the following, we will extend this result and comment on the tractability of optimal scheduling for the case of $\alpha \neq 1$.

Remark 5. *For $\alpha \neq 1$, if all x_{ji}'s (the UA rule) are given, obtaining the optimal schedules can be a difficult problem. Consider the following. User i is defined to have **higher priority** over user u if user u is scheduled in MBS off implies user i is scheduled in MBS off. Then,*

1. *If $\{([\tilde{r}_k/r_k] > [\tilde{r}_{k+1}/r_{k+1}]) \cap (\tilde{r}_k > \tilde{r}_{k+1})\}$, then it is easy to show that user k gets higher priority than user $k+1$ independent of the other user associations and ordering produces the optimal schedule.*

2. *If $\{([\tilde{r}_k/r_k] > [\tilde{r}_{k+1}/r_{k+1}]) \cap (r_k < r_{k+1}) \cap (\tilde{r}_k < \tilde{r}_{k+1})\}$, the priority decision between users $k, k+1$ depends on the other users' associations and rates.*

Proof. *The proof follows by applying the KKT conditions to the local (per-BS) $\alpha-$fair scheduling problems.*

Consider the worst case, with all users having rates obeying the second condition above. Finding the optimal ordering for such users is identical to poset ordering which is known to be NP-hard, indicating that

... the $\alpha \neq 1-$fair scheduling problem under ABS can be NP-hard.

This is while an optimal formula-based solution exists for the general-case $\alpha-$fair scheduling problem under PSD as we showed n Chapter 3.

4.3.1 User association

As in Section 3.5.2, we use a number of simple UA schemes that can be used along with the optimal scheduler under ABS (by solving (4.2) for β_{ji} and $\tilde{\beta}_{ji}$) to allocate throughputs to users. Studying these UA schemes helps us understand how simple association rules perform compared to the optimal UA (which can be obtained by solving the problem in (4.2) for the optimal x_{ji}'s). As before, we use Max-SINR, SCF and CRE as representative of such schemes to also provide a fair basis of comparison to PSD. The definition of the UA schemes in Section 3.5.2 follows similarly for ABS by replacing (the set of BSs under PSD) $\mathcal{B} \cup \{0_M, 0_S\}$ by $\mathcal{B} \cup \{0\}$. Note that under ABS since a given user i sees two SINRs from a SC j, all the schemes above can be performed based on either of the two SINRs. We denote the respective UA schemes under ABS by *Max-SINR, SCF, CRE* when γ_{ji} during ABS is used, and by *Max-SINR′, SCF′, CRE′* when γ_{ji} during non-ABS is used in the corresponding UA scheme.

4.4 Numerical results

4.4.1 Parameter settings

We consider the middle cell in a 19-cell HetNet with a reuse factor of $r = 3$ (see Fig. 3.2) and use the same notations and parameter settings as in Section 3.6 unless otherwise specified.

4.4.2 Relaxation and optimality gap

The optimisation problem in (4.2) is an integer (and non-linear if $\alpha \neq 0$) program and hard to solve quickly with the existing integer program solvers. For the ease of computation, we convert the integer program into a continuous convex optimisation problem by allowing fractional user association (FUA), i.e., relaxing the integrality constraints (i.e.,

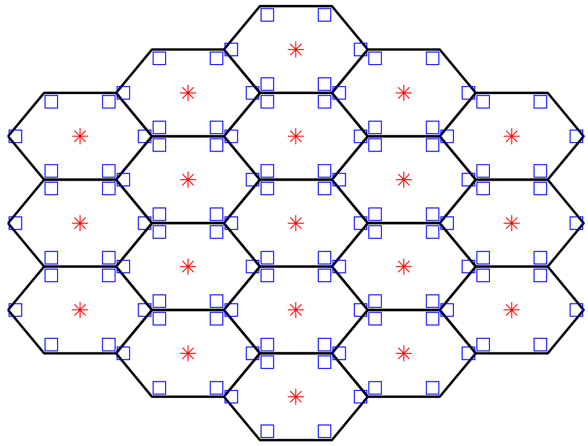
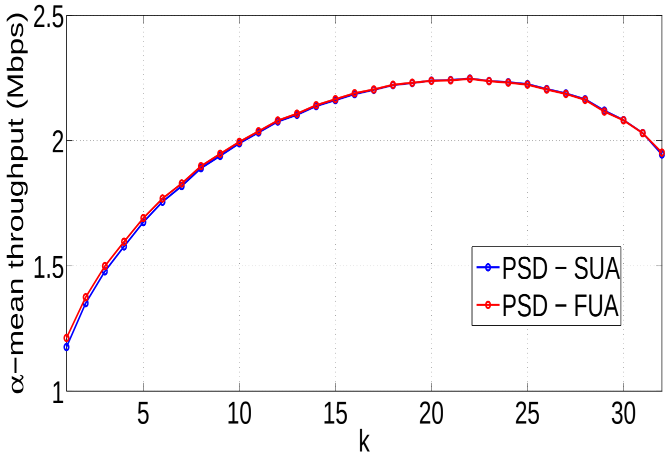


Figure 4.1: The HetNet configuration with 6 SCs per macro cell

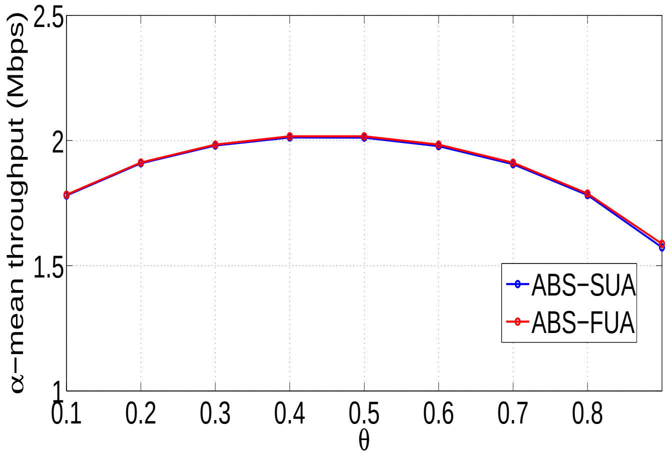
$x_{ji} \in [0,1]$). This relaxed problem, termed as *ABS-FUA*, provides an upper-bound to its corresponding integer program specified in (4.2). We obtain this upper-bound by solving the relaxed convex optimisation problem for each realisation using the commercial solver *Minos 5.51*. To see that the relaxation is tight, we obtain a feasible set of solutions, $\{x_{ji} \in \{0,1\}\}_{j,i}$, from the *ABS-FUA* problem as follows. For user i, choose the BS $j_i^* = \arg\max_{j \in \mathcal{B} \cup \{0\}} r_{ji}\beta_{ji} + \tilde{r}_{ji}\tilde{\beta}_{ji}$ and break the ties in favour of the SCs. We refer to the *ABS-FUA* problem with this SUA feasible solution as *ABS-SUA*, since it provides a lower-bound for the optimal solution to the SUA problem in (4.2).

4.4.3 The static setting

We consider a snapshot model for our optimisation framework similar to the one in Section 3.6. All the parameters and notations used in our static setting are as described before unless otherwise specified. We generate a set of 100 realisations Ω (i.e., $|\Omega| = 100$) and, for each realisation $\omega \in \Omega$, compute a solution to the *ABS-FUA* and *ABS-SUA* problems for each value of $\theta \in \{0.1, 0.2, ..., 0.9\}$. For each problem, we compute the (average) $\alpha-$mean throughput corresponding to that problem over the 100 realisations and plot it as a function of θ for $\alpha \in \{1, 2\}$.



(a) PSD - $\alpha = 1$



(b) ABS - $\alpha = 1$

Figure 4.2: $\alpha-$mean throughput as a function of k and θ. Settings: $U = 20$, $M = 33$, $\Delta = 230\ m$ (MBS-to-SC distance).

4.4.4 Comparison between PSD and ABS: The upper-bounds

Fig. 4.2 shows the $\alpha-$mean throughput as a function of k and θ for FUA and SUA problems corresponding to PSD and ABS. The curves corresponding to ABS in Fig. 4.2(b) show that *ABS-FUA* and *ABS-SUA* perform very closely for all values of θ, indicating that the problem in (4.2) can be with high accuracy approximated by *ABS-FUA*.

The optimum $\alpha-$mean throughputs (with $\alpha = 1$) for PSD and ABS are 2.247 $Mbps$ and 2.017 $Mbps$, obtained at $k = 20$ and $\theta = 0.5$ respectively, indicating a 11.40% higher performance for *PSD-FUA* compared to *ABS-FUA*.

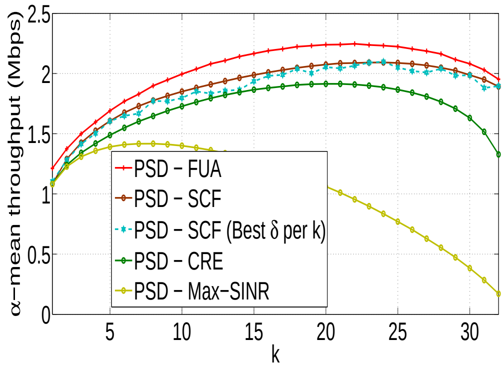
4.4.5 Comparison between different UA schemes under ABS

Fig. 4.3 shows the $\alpha-$mean throughput as a function of k and θ under different simple UA schemes (i.e., Max-SINR, SCF, CRE). For *ABS-CRE*, the value of ϵ in each realisation is selected so that the $\alpha-$mean throughput of users is maximised in the given realisation. The curves suggest that by fine-tuning the parameters (θ, ϵ), *ABS-CRE* can perform very closely to the upper-bound, i.e., *ABS-FUA*.

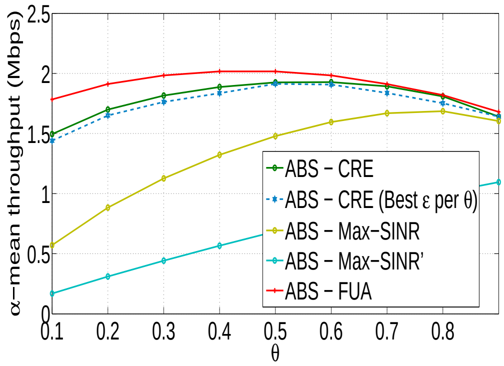
Since it may not be a feasible option to vary ϵ too frequently in a dynamic system, we also plot the average $\alpha-$mean throughput per θ for the best value of ϵ. To do so, for each θ, we compute the average $\alpha-$mean throughput over the 100 realisations for each value of ϵ and select the ϵ that yields the best average $\alpha-$mean throughput. We refer to this as *ABS-CRE (Best* ϵ *per* θ*)* in the figures. Similarly, we define *ABS-SCF* and *ABS-SCF (Best* δ *per* θ*)*. The results show that fine-tuning the UA parameter ϵ per θ for CRE results in roughly the same $\alpha-$mean throughput as when selecting the optimal ϵ per realisation. This indicates that fine-tuning the parameter ϵ for each realisation, which can be a bottleneck in a dynamic network, does not incur much gain in performance. Lastly, the results suggest that *ABS-CRE* generally performs better than *CRE-SCF* for both values of α and significantly better than Max-SINR.

4.4.6 Comparison between PSD and ABS under different UA schemes

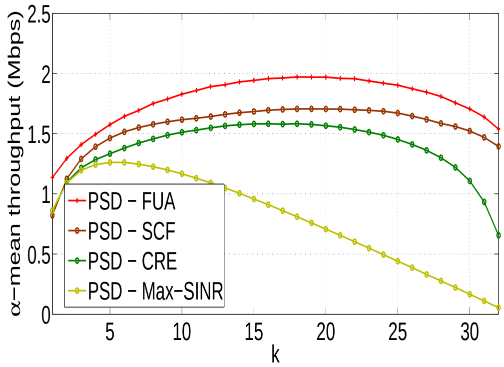
Fig. 4.3 shows that *PSD-SCF (Best* δ *per* k*)* and *ABS-CRE (Best* ϵ *per* θ*)* result in the highest $\alpha-$mean throughput under PSD and ABS respectively, among the UA schemes under study, and both perform close to their corresponding (FUA) upper-bounds. By



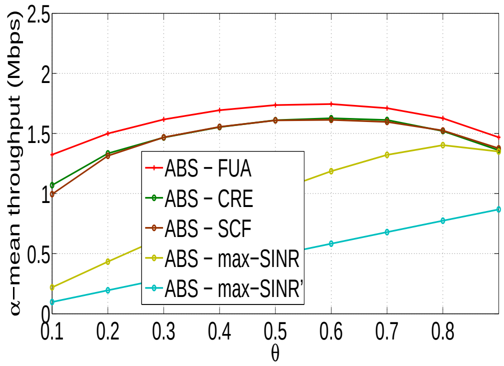
(a) PSD - $\alpha = 1$



(b) ABS - $\alpha = 1$

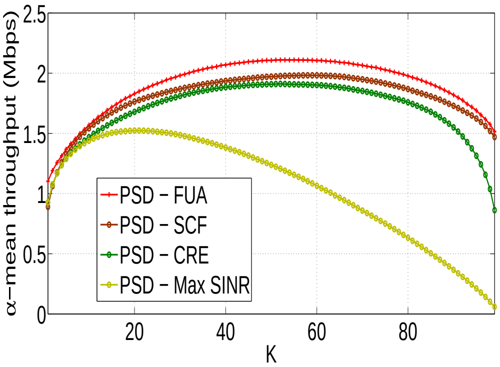


(c) PSD - $\alpha = 2$

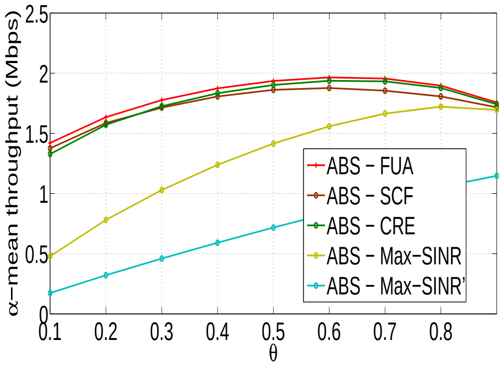


(d) ABS - $\alpha = 2$

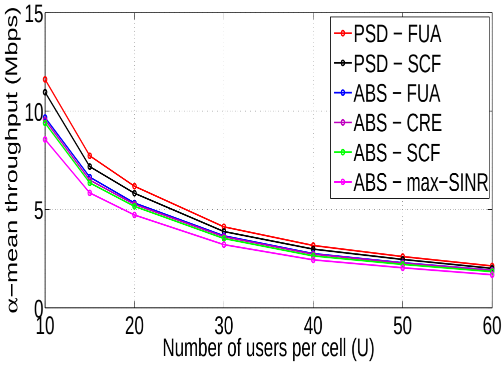
Figure 4.3: $\alpha-$mean throughput as a function of k and θ. Settings: $U = 20$, $M = 33$, $\Delta = 230\ m$.



(a) PSD - $U = 60$, $M = 100$, $\alpha = 1$



(b) ABS - $U = 60$, $M = 100$, $\alpha = 1$



(c) PSD - $\alpha = 1$

Figure 4.4: Comparison between PSD and ABS under different settings

comparing the resulting $\alpha-$mean throughputs of the two schemes for $\alpha = 1$, we see a 10.52% gain from deploying PSD over ABS. For $\alpha = 2$ the figure indicates that the *PSD-SCF* has a marginally better performance (of about 4.20%) than the *ABS-CRE*. However, recall that scheduling under ABS for $\alpha = 2$ can be much more involved (see Remark 5) than under PSD.

Fig. 4.3 and 4.4 reconfirm the upper hand of PSD over ABS for $\alpha = 2$ and other values of U and M in the static setup.

We also conducted a number of experiments to compare the performance of ABS and PSD under the three UA schemes (i.e., Max-SINR, SCF, CRE) for different U, M, Δ and B. Our objective was to see which UA scheme performs better with which RA scheme. We summarise our observations below.

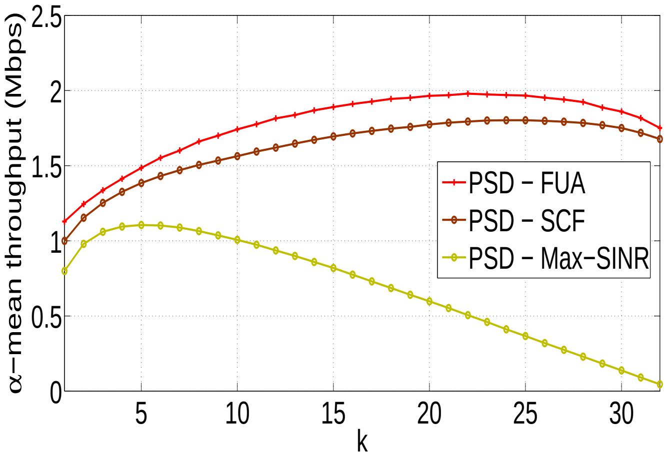
Our numerical experiments suggested that, for different HetNet configurations,

1. the UA schemes under ABS perform better if based on γ_{ji} during the ABS duty cycle and
2. SCF performs better under PSD while CRE performs better under ABS.

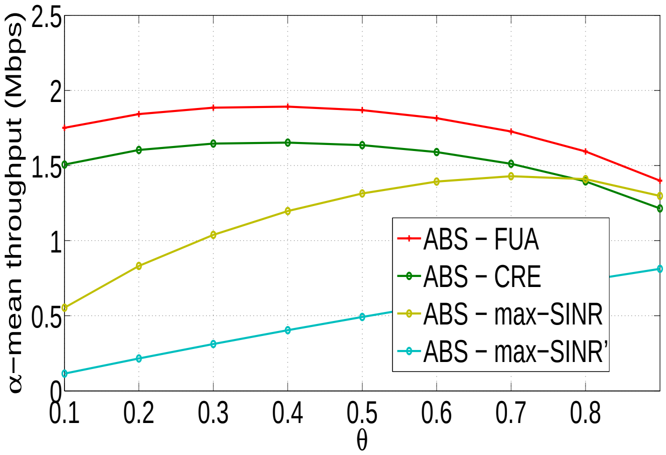
Hence, from now on, we will only include *PSD-SCF* and *ABS-CRE* in our figures unless necessary.

4.4.7 Extended results

1. *Impact of SC placement on $\alpha-$mean throughput:* Fig. 4.5 shows the $\alpha-$mean throughput as a function of k and θ for $\alpha = 1$ for a shorter MBS-to-SC distance (than in the previous figures), i.e., $\Delta = 145$ m. The results indicate a 4.76% higher performance for *PSD-FUA* compared to *ABS-FUA*. It is noteworthy that for smaller values of Δ, the optimal k for PSD (θ for ABS) increases (decreases) since more users will tend to associate with the SCs. In particular, the users in the middle of the cell will likely associate with the SCs and the cell-edge users will likely associate with the MBS. This is because, as the SCs are placed closer to the MBS, the distance between the SCs and the (uniformly distributed) users reduces and their channel quality improves. This also implies that it is best to locate SCs near the cell edges to achieve



(a) PSD - $\alpha = 1$



(b) ABS - $\alpha = 1$

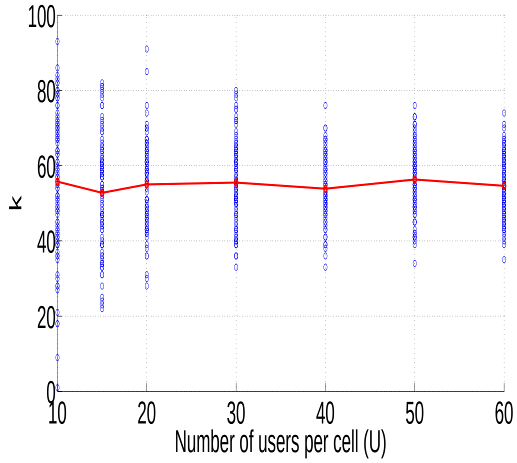
Figure 4.5: $\alpha-$mean throughput as a function of k and θ. Settings: $U = 20$, $M = 33$, $\Delta = 145\ m$.

a higher mean throughput, since otherwise cell-edge users who are already suffering from inter/intra-cell interference from the neighbouring co-channel cells, will be served by the MBS far apart with poor channel quality.

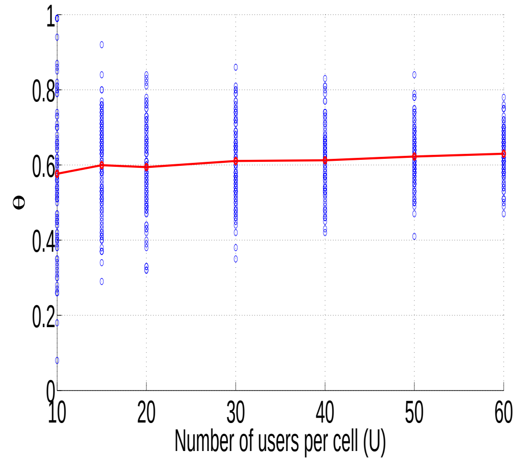
2. *Impact of SC placement on k and θ:* Figs. 4.6(a)-4.6(d) show the impact of SCs locations on k and θ. The optimal values of k and θ over the 100 realisations for each U are shown by the blue circles on the y-axis and the average of these optimal values are shown by the red circles on the y-axis. The results suggest that k is more robust to the SCs placement than θ (under the assumption that users are uniformly distributed).

3. *Impact of number of SCs on $\alpha-$mean throughput:* Fig. 4.7 shows the impact of deploying a higher number SCs on the $\alpha-$mean throughput of PSD and ABS. The results indicate that by deploying 6 SCs in each cell of the HetNet (see Fig 4.1), we observe a 10.47% better performance for *PSD-FUA* over *ABS-FUA*. The optimal value of k (and θ) increases (decreases) compared to 4-SC case since now more users will associate with the SCs and, hence, the increase (increase) in the number of shared channel resources (ABS duty cycle).

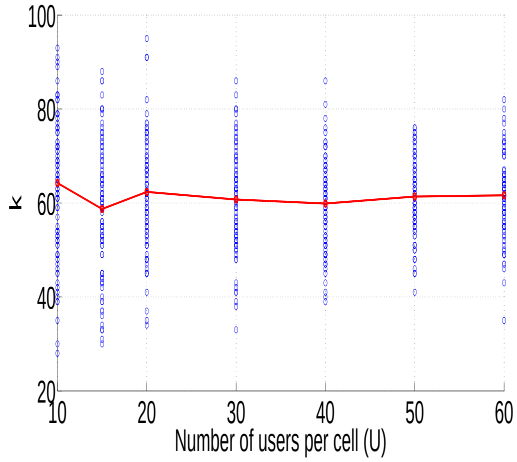
4. *Impact of fixing k and θ on $\alpha-$mean throughput:* So far, we have looked at the performance of the HetNet, under the assumption that the RA parameters k and θ can be optimise for each realisation. However, such an assumption might not be feasible in a fast-changing dynamic network and, hence, it is interesting to see if fixing the RA parameters will result in significant loss in the throughput performance or not. To do this, we compute the per-realisation relative gain for *PSD-SUA* by $\frac{\bar{T}_\alpha(k^*_\omega,\omega)-\bar{T}_\alpha(\bar{k},\omega)}{\bar{T}_\alpha(\bar{k},\omega)}$ where $\bar{T}_\alpha(k,\omega)$ is the $\alpha-$mean throughput of all users in the cell as a function of k and ω, k^*_ω is the optimal k at realisation ω and $\bar{k}$ is a fixed value for all ω's from the set $\{1, ..., M-1\}$. We compute the relative gain for *ABS-SUA* similarly. We set $\bar{k}$ and $\bar{\theta}$ to the average of k^*_ω's and θ^*_ω's over the 100 realisations, respectively and plot the relative gain in $\alpha-$mean throughput as a function of ω. Fig. 4.8 shows that fine-tuning $\bar{k}$ and $\bar{\theta}$ will result in less than 5% (PSD) and 8% (ABS) loss in $\alpha-$mean throughput performance, indicating the robustness of the two RA schemes to their respective RA parameters.



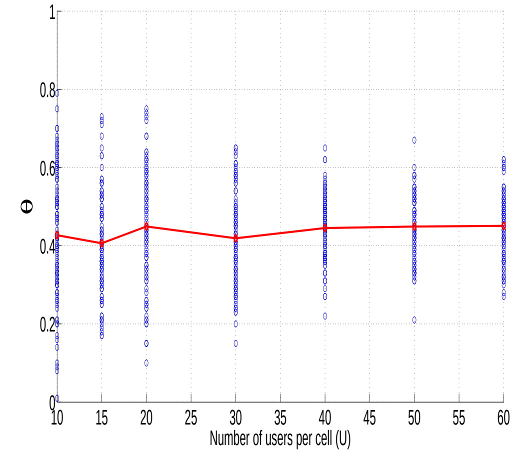
(a) PSD - $\Delta = 230\ m$



(b) ABS - $\Delta = 230\ m$

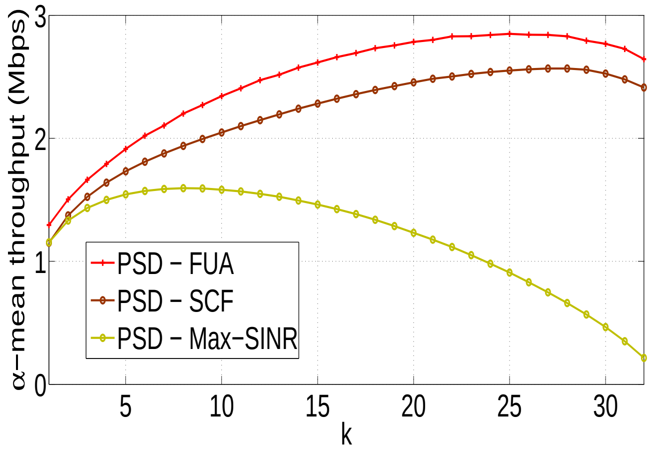


(c) PSD - $\Delta = 145\ m$

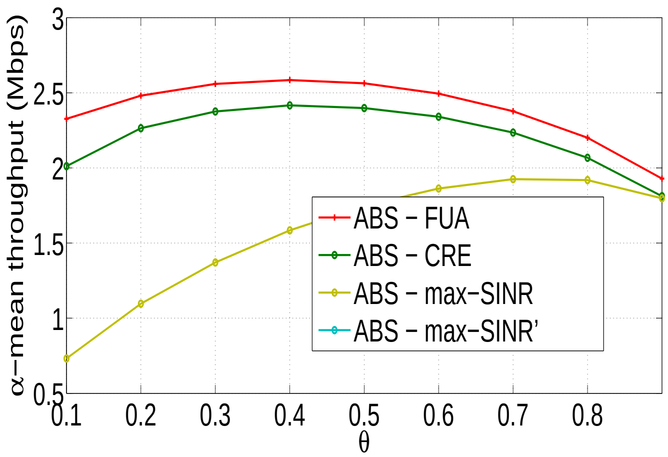


(d) ABS - $\Delta = 145\ m$

Figure 4.6: k as a function of U. Settings: $\alpha = 1$, $M = 100$.

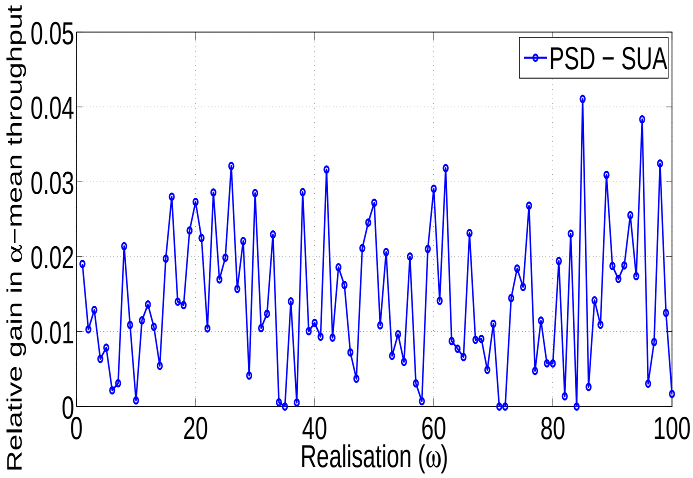


(a) PSD - $\alpha = 1$

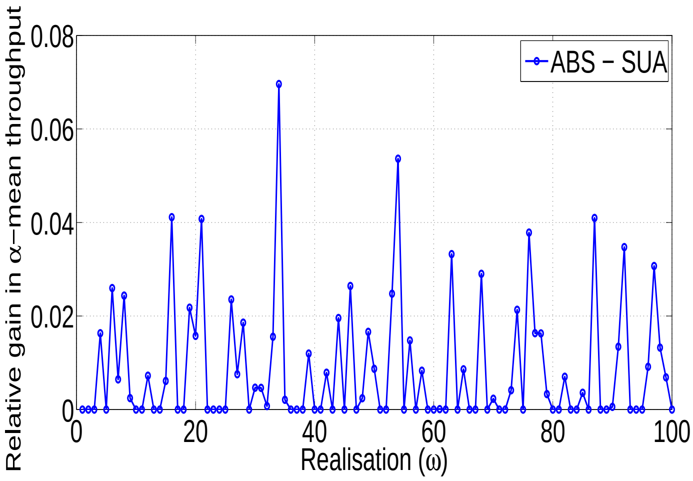


(b) ABS - $\alpha = 1$

Figure 4.7: $\alpha-$mean throughput as a function of k. Settings: $U = 20$, $M = 33$ and $|\mathcal{B}| = 6$.



(a) PSD - $\alpha = 1$



(b) ABS - $\alpha = 1$

Figure 4.8: Relative gain in $\alpha-$mean throughput from optimising k (and θ) for each realisation as a function of k (and θ). Settings: $U = 60$, $M = 100$.

4.5 Conclusions

In conclusion, we revisited the problem of joint RA, UA and US under ABS scheme, first discussed in [6], where the authors show that the optimal PF (i.e., $\alpha = 1$) scheduling problem under ABS is a simple linear-in-time algorithm. Interested in characterising the optimal scheduler for $\alpha \neq 1$, we proved that the optimal $\alpha-$fair scheduling under ABS can be NP-hard and, hence, much more involved than that under PSD (which, as we discussed in Chapter 3, is formula-based for all values of α). To verify whether the scheduling complexities involved from deploying ABS were justifiable, we further made a thorough numerical comparison between ABS and PSD in a static setting and showed that the throughput gains achieved by deploying PSD were higher than ABS for different values of α and under different HetNet configurations and heuristics. Therefore, based on its simpler optimal $\alpha-$fair scheduler and higher throughput gains, we assert that PSD outperforms ABS in a static setting.

Chapter 5

A Thorough Comparison Between PSD and ABS: A Dynamic Setting

Summary: In this chapter, we

- consider the optimisation frameworks under PSD and ABS in a dynamic setting,
- through simulations, provide upper-bounds to the performance of the system for the two RA schemes,
- show that the best PSD scheme outperforms the best ABS scheme, under various network configurations, and
- show that the physical-layer based UA schemes, i.e., SCF and CRE, could perform quite poorly in a dynamic network and that new UA schemes are required to compensate for the performance loss.

In Chapter 4, we studied the joint optimisation of RA, UA and US under ABS and made a thorough comparison with its PSD counterpart under a snapshot model in a static setting. Most notably, we observed the following remarks: 1) For a given UA rule (x_{ji}'s) and RA parameter (k or θ), while optimal $\alpha-$fair scheduling under PSD is formula-based for all values of α, it can be NP-hard under ABS for $\alpha \neq 1$ and 2) using numerical results, we showed that PSD performs better than ABS in terms of *average*[1] $\alpha-$mean throughput for different values of α and under different HetNet configurations and UA heuristics. Although these remarks clearly indicate the upper-hand of PSD both in terms of ease of implementation and average throughput performance, they are only valid under the static setting. As mentioned before, a static setting is only limited to an offline study and does not necessarily reflect the behaviour of an online system including all of its dynamics. Therefore, in order to confirm whether our conclusions from the static setting still hold in an online system, in this chapter we further extend our framework to a dynamic setting.

5.1 System model

Our optimisation framework is based on the dynamic setting in which users come, according to a predefined random process, and depart after being served. For simplicity, we do not consider mobility so users only depart when they have been fully served. We consider only the active users in the network and assume that the users are greedy in the sense that they want to maximise their individual throughputs. We study one cell in an OFDM-based multi-cell HetNet under PSD. Note that, we only show the steps to develop the dynamic framework under PSD. The framework under ABS can be developed following similar steps. The physical-layer characteristics including the SINR and link rate models are all identical to the ones described in Section 3.4.1.

Let $\mathcal{U}(t)$ represent the set of users in the cell under consideration at time t. We define the network realisation $\omega(t) = \{g_{ji}(t)\}_{j\in\mathcal{B}\cup\{0_M,0_S\},i\in\mathcal{U}(t)}$ to be the set of channel gains between all user-BS pairs in the cell. We assume time-invariant channels for all users in the system and, therefore, the network realisation $\omega(t)$ would change only when a user arrives or departs. For simplicity, we assume that no users arrive in the system at the same time. Let a_u and d_u respectively be the arrival and departure times of user u. Then,

$$\begin{aligned}\mathcal{U}(a_u) &= \mathcal{U}(a_u^-) \cup \{u\},\\ \mathcal{U}(d_u) &= \mathcal{U}(d_u^-) \backslash \{u\},\end{aligned} \tag{5.1}$$

where a_u^- and d_u^- represent the time just before the arrival and the departure of the $n-$th user, respectively. We assume that an association decision for user u will be taken at the time of her arrival and denote the binary UA variable at time t by $x_{ji}(t)$. The values of $x_{ji}(t)$ may vary depending on whether a user is triggered to re-associate or not, i.e., the

[1]Note that although we have shown that PSD performs better than ABS on average over a large number of snapshots, this does not imply that PSD performs better in all snapshots.

UA variable $x_{ji}(t)$ at instant t is determined by the decision carried out at the most recent association or re-association event. Furthermore, we denote the proportion of the time allotted to user i by BS j at time t *in the next available frame*[2] by $\beta_{ji}(t)$.

5.1.1 Modelling the network processes

In an online system, network processes (i.e., RA, UA, and US) are often computed at different time-scales, some more frequently and some less, either by the network or the users. For example, the RA parameter k under PSD is typically computed offline by the operator and changes at a very slow time-scale. US, on the other hand, is carried out by the BSs and is a quasi real-time process over equally partitioned transmission time intervals (e.g., sub-frames). UA decisions, however, can be carried out by either the user or the network (or possibly jointly) whenever an association or re-association event is triggered. Therefore, the frequency of UA will depend on the deployed UA scheme, although it is always faster than RA and slower than US. In the following, we describe the assumptions made to model the network processes for our dynamic system and the time-scales at which each process is carried out.

1. **Resource allocation:** For a given RA scheme, i.e., PSD or ABS, we assume that the RA parameters k or θ are fixed and known (to both users and BSs) a priori. Clearly, to obtain upper-bounds to the performance of our system under different deployment choices these parameters need to be carefully fine-tuned. We describe the process of fine-tuning the RA parameters in Section 5.3.

2. **User association:** For a given user i, we assume that the UA variables $\{x_{ji}\}_j$ are computed at least once at the user's arrival instant but they can also be recomputed at other times. Most of the existing UA schemes (e.g., SCF) trigger an association at the arrival time of a user without allowing further re-associations of that user. Depending on whether re-association is allowed or not, different upper-bounds can be obtained for the performance of the system. We consider two of such upper-bounds described as follows. 1) If the *network* jointly *recomputes* user associations of *all users* in the system every time there is a change in $\omega(t)$ so as to maximise the global sum of user utilities, we can obtain an upper-bound on the performance of all UA schemes that allow re-association. We refer to this as the joint global *PSD-SUAWithReassociation* problem and formulate it in Section 5.2. 2) If the *network* *computes* the user association of only *a newly arriving user* in the system while

[2] We assume scheduling is performed on a per-frame basis.

retaining the old users' association so as to maximise the global sum of user utilities, we can obtain an upper-bound on the performance of all UA schemes that allow association at the time of arrival only without further re-association of any of the users. We refer to this as the joint global *PSD-SUAWithoutReassociation* problem and formulate it in Section 5.2. Clearly, *PSD-SUAWithReassociation* provides an upper-bound to *PSD-SUAWithoutReassociation*.

3. **User scheduling:** For a given user i, we assume that the US variables $\{\beta_{ji}\}_j$ are computed on a per-frame basis. Under our dynamic framework, we only need to recompute the scheduling variables $\beta_{ji}(t)$ at BS j if there is a change in the realisation $\omega(t)$ that affects the set of users associated with j[3], i.e., if there is an arrival or departure occurring at BS j. Upon a change in the realisation $\omega(t)$ and given the latest UA decisions, the optimal US scheme would (re-)compute $\beta_{ji}(t)$'s so that the global sum of user utilities is maximised in the next frame.

In the following section, we formulate and analyse the joint global $\alpha-$fair SUA with re-association problems under PSD and ABS.

5.2 The global $\alpha-$fair SUA with re-association problem

Let τ be the set of all events that cause the realisation $\omega(t)$ to change, i.e., either a new arrival or departure. Then, for a fixed k and $\hat{P}_S$, the global $\alpha-$fair *PSD-SUAWithReassociation* problem can be formulated as follows for $\forall t \in \tau$,

[PSD-SUAWithReassociation$(t, k, \hat{P}_S)$**]:**

$$\max_{\{\beta_{ji}(t), x_{ji}(t)\}} \sum_{i \in \mathcal{U}(t)} U_\alpha \left(\sum_{j \in \mathcal{B} \cup \{0_M, 0_S\}} r_{ji} \beta_{ji}(t) \right)$$

$$\text{s.t.} \sum_{i \in \mathcal{U}(t)} \beta_{ji}(t) \le 1, \ \forall j \in \mathcal{B} \cup \{0_M, 0_S\}, \tag{5.2a}$$

$$\sum_{j \in \mathcal{B} \cup \{0_M\}} x_{ji}(t) = 1, \ \forall i \in \mathcal{U}(t), \tag{5.2b}$$

[3] Since otherwise the schedules will take the same values as the previous frame and, so, no re-computation is required.

$$0 \leq \beta_{ji}(t) \leq x_{ji}(t), \ \forall (i \in \mathcal{U}(t), j \in \mathcal{B} \cup \{0_M, 0_S\}), \tag{5.2c}$$

$$x_{ji}(t) \in \{0, 1\}, \ \forall (i \in \mathcal{U}(t), j \in \mathcal{B} \cup \{0_M, 0_S\}). \tag{5.2d}$$

Note that even though $\beta_{ji}(t)$'s will be computed at each time instant $t \in \tau$, a user i associated with BS j will not be scheduled until the next available frame. Furthermore, note that the problem in (5.2) is a function of time and is called by the system every time the realisation $\omega(t)$ changes.

Similarly, for a fixed θ, we define the global $\alpha-$fair *ABS-SUAWithReassociation* problem as follows.

[ABS-SUAWithReassociation(t, θ)]:

$$\max_{\{\tilde{\beta}_{ji}(t), \beta_{ji}(t), x_{ji}(t)\}} \sum_{i \in \mathcal{U}(t)} U_\alpha \left(\sum_{j \in \mathcal{B} \cup \{0\}} r_{ji}\beta_{ji}(t) + \tilde{r}_{ji}\tilde{\beta}_{ji}(t) \right)$$

$$\text{s.t.} \sum_{i \in \mathcal{U}(t)} \beta_{ji}(t) \leq \theta, \quad \forall j \in \mathcal{B} \cup \{0\}, \tag{5.3a}$$

$$\sum_{i \in \mathcal{U}(t)} \tilde{\beta}_{ji}(t) \leq 1 - \theta, \quad \forall j \in \mathcal{B}, \tag{5.3b}$$

$$\sum_{j \in \mathcal{B} \cup \{0\}} x_{ji}(t) = 1, \quad \forall i \in \mathcal{U}(t), \tag{5.3c}$$

$$0 \leq \beta_{ji}(t), \tilde{\beta}_{ji}(t) \leq x_{ji}(t), \ \forall (i \in \mathcal{U}(t), j \in \mathcal{B} \cup \{0\}), \tag{5.3d}$$

$$x_{ji}(t) \in \{0, 1\}, \ \forall (i \in \mathcal{U}(t), j \in \mathcal{B} \cup \{0\}), \tag{5.3e}$$

where $\tilde{\beta}_{ji} = 0$.

The problems in (5.2) and (5.3) provide the absolute best upper-bounds for the $\alpha-$mean throughput of the system as they trigger the re-computation of all UA and US variables by the network upon every change in the realisation $\omega(t)$, i.e., this is the best that can be done among all classes of US and UA schemes. However, one should note that although these problems provide upper-bounds for performance, they are only useful as benchmarks for comparison with other (simple) schemes; otherwise, they would be prohibitively complex to solve to optimality as a network-centric[4] online scheme. This is because finding an optimal

[4]In a network-centric UA scheme, the UA decisions are made by the network (e.g., the serving BS, or a network controller in C-RAN) and are communicated to the users. The decisions are made with the help of link-level measurement feedback provided by the users (e.g., channel conditions of a user from all candidate BSs).

solution would involve an exhaustive search over the set of all possible binary variables x_{ji}'s and, consequently, take too long under strict end-to-end latency constraints. Therefore, these problems are only useful as a way to find upper-bounds for the performance of the system which is the main objective of this chapter. We will delay the achievability of these upper-bounds using heuristic schemes to Chapter 6.

Similar to *SUAWithReassociation* problems, we can define two respective problems for PSD and ABS to provide upper-bounds for all classes of US and UA schemes that only allow association of users upon their arrival without allowing further re-association. We refer to these problems as *PSD-SUAWithoutReassociation* and *ABS-SUAWithoutReassociation* each of which can be obtained by adding a set of constraints to the problems in (5.2) and (5.3), respectively, to enforce old users' association. Formally, let $\mathcal{U}'(t) \subset \mathcal{U}(t)$ be the set of old users in the system (who have not departed from the system yet) at time t and let $\{x_{ji}(a_i)\}_j$ denote the user i's association parameters upon her arrival. Then, for $\forall t \in \tau$, we have

$$\textbf{[PSD-SUAWithoutReassociation}(t, k, \hat{P}_S)\textbf{]:}$$

$$\begin{aligned}
&\max_{\{\beta_{ji}(t), x_{ji}(t)\}} \sum_{i \in \mathcal{U}(t)} U_\alpha \left(\sum_{j \in \mathcal{B} \cup \{0_M, 0_S\}} r_{ji} \beta_{ji}(t) \right) \\
&\quad \text{s.t. } (5.2a) - (5.2d), \\
&\qquad x_{ji}(t) = x_{ji}(a_i), \ \forall (i \in \mathcal{U}'(t), j \in \mathcal{B} \cup \{0_M, 0_S\}),
\end{aligned} \tag{5.4}$$

and

$$\textbf{[ABS-SUAWithoutReassociation}(t, \theta)\textbf{]:}$$

$$\begin{aligned}
&\max_{\{\tilde{\beta}_{ji}(t), \beta_{ji}(t), x_{ji}(t)\}} \sum_{i \in \mathcal{U}(t)} U_\alpha \left(\sum_{j \in \mathcal{B} \cup \{0\}} r_{ji} \beta_{ji}(t) + \tilde{r}_{ji} \tilde{\beta}_{ji}(t) \right) \\
&\quad \text{s.t. } (5.3a) - (5.3e), \\
&\qquad x_{ji}(t) = x_{ji}(a_i), \ \forall (i \in \mathcal{U}'(t), j \in \mathcal{B} \cup \{0\}),
\end{aligned} \tag{5.5}$$

where $x_{ji}(a_i)$'s are fixed and known to the network. Note that the problems in (5.4) and (5.5) are only useful as benchmarks for comparison with other schemes; otherwise, they would be too complex to implement as a network-centric online scheme (because of the integrality constraints). Next, we compare the performance of PSD and ABS in a dynamic setting using our developed frameworks.

5.3 Numerical results

5.3.1 Parameter settings and integrality relaxation

We consider the middle cell in a 19-cell HetNet with a reuse factor of $r = 3$ (see Fig. 3.2) and use the same notations and parameter settings as in Section 3.6 unless otherwise specified. To obtain a good and fast approximation to the optimisation problems in (5.2)-(5.5), we relax the integrality constraints in the problems and obtain a feasible solution from their corresponding relaxed versions as explained in Sections 3.6.2 and 4.4.2.

5.3.2 The dynamic setting

In the dynamic setting, the users arrive in the system according to a homogeneous Poisson point process of rate λ and choose their locations i.i.d. uniformly. Recall that the users channels are time-invariant, i.e., each user observes the same channel gain as upon arrival for her complete stay in the system. We consider each frame duration to be 10^{-2} *sec*. Every arriving user is scheduled in the next available frame.

5.3.3 Service-time models

We consider two service-time scenarios.

1. *File-download* scenario: In this scenario, each user leaves the system when she downloads a file of a fixed size. The service-time of a user depends on the old and new arrivals and departures.

2. *Fixed-delay* scenario: In this scenario, the life-time of all users in the system is fixed. This service-time model is suitable for modelling users watching media streams of fixed length where the quality of a stream (i.e., its coding rate) is adjusted to match the available end-to-end throughput. In this scenario, higher throughput to a particular user translates into a better quality of service, but the amount of time a user spends in the system is independent of the allocated throughput.

We choose the average per-user delay and the average α-mean throughput as the performance metrics for the file-download and fixed-delay scenarios, respectively.

For a given arrival rate λ, we run 5 simulations each with a different random seed. Each simulation runs for a period of at least 1000 seconds. For convergence, the period of simulation is increased at a step of 1000 seconds until the performance metric is within 5% of the performance metric before the increase.

For each simulation with file-download scenario, we compute the average per-user delay as follows. For each user i, we record the total time that the user spends in the system, i.e., until she completes downloading a file of size F. We compute the average per-user delay per-simulation by taking the arithmetic mean of the delays of all users who have departed from the system over the simulation period. We, then, take the arithmetic mean of these quantities over the set of 5 simulations to obtain average per-user delay.

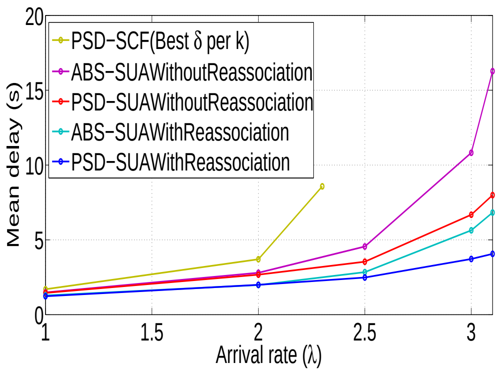
For each simulation with fixed-delay scenario, we compute the average α-mean throughput as follows. For each user i, we compute her throughput at her departure instant as $T_i = \frac{\bar{m}_i}{\hat{t}}$ where $\bar{m}_i$ is the total bits transmitted to user i during her life-time of $\hat{t}$ seconds. We calculate the $\alpha-$mean throughput per-simulation of all recorded user throughputs T_i in the simulation period and, then, take the arithmetic mean of these quantities over the set of 5 simulations to obtain the average α-mean throughput.

5.3.4 Fine-tuning the RA parameters in the dynamic setting

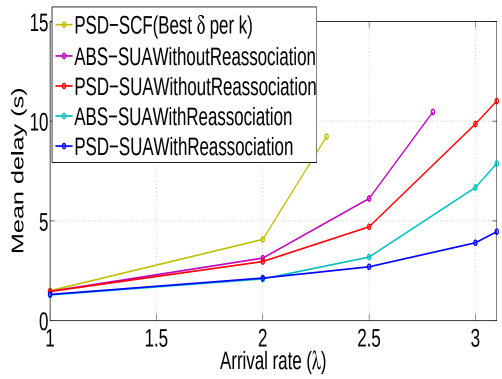
In order not to bias the results in favour of one scheme, we obtain the performance metric for each scheme when the RA parameter k is optimised for that scheme for each value of λ. More precisely, for each value of λ, we run a simulation per scheme for all possible values of k and keep the best results over all k for that λ. Furthermore, for SCF, we jointly optimise the pair of parameters δ and k for each λ.

5.3.5 Comparison between PSD and ABS: The upper-bounds

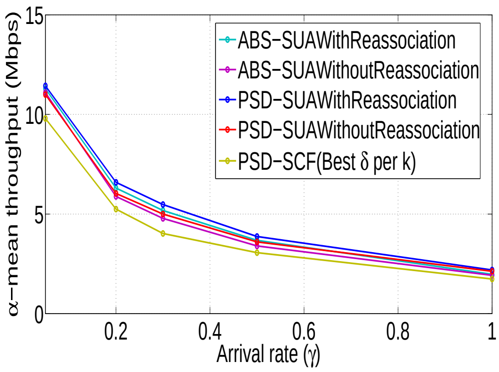
As before, we show the results for $\alpha \in \{1, 2\}-$fairness for ABS and PSD by approximating their corresponding problems as described in Section 5.3.1. Figs. 5.1 show the variation of the performance metrics with the arrival rate λ. We have only included arrival rates that result in an average number of users between 1 and 30. For both $\alpha = \{1, 2\}-$fairness, it is evident that the optimal ABS with (without) re-association has higher average delay and lower α-mean throughput than PSD with (without) re-association. We summarise the performance gap between the two service-time scenarios as follows.



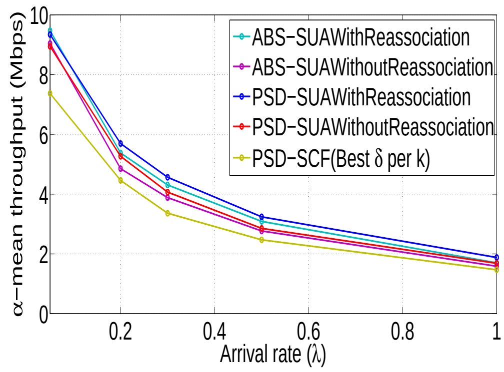
(a) file-download model: $\alpha = 1$



(b) file-download model: $\alpha = 2$



(c) Fixed-delay model: $\alpha = 1$



(d) Fixed-delay model: $\alpha = 2$

Figure 5.1: Performance metric as a function of arrival rate (λ). Settings: $\hat{t} = 20$ sec and $F = 10$ $Mbit$ for all users and $M = 33$.

- *File-download*: We can achieve up to 1) 11.91% and 8.30% better performance by deploying *PSD-SUAWithoutReassociation* over *ABS-SUAWithoutReassociation* for $\alpha = 1$ and 2, respectively, and 2) 14.81% and 15.67% better performance by deploying *PSD-SUAWithReassociation* over *ABS-SUAWithReassociation* for $\alpha = 1$ and 2, respectively.

- *Fixed-delay*: We can achieve up to 1) 5.83% and 8.66% better performance by deploying *PSD-SUAWithoutReassociation* over *ABS-SUAWithoutReassociation* for $\alpha = 1$ and 2, respectively, and 2) 5.60% and 5.59% better performance by deploying *PSD-SUAWithReassociation* over *ABS-SUAWithReassociation* for $\alpha = 1$ and 2, respectively.

5.3.6 Efficiency of SCF under different service-time models

Figs. 5.1 indicate that while PSD with SCF performs quite well in the fixed-delay scenario, it performs quite poorly in the other scenario indicating the inefficiency of the physical-layer based UA rule for the latter scenario. Furthermore, note that while re-association in the fixed-delay scenario (under both PSD and ABS) does not result in much gain in performance metric, it significantly improves the metric in the file-download scenario. This suggests that incorporating re-association along with a network-aware UA scheme (e.g., *PSD-SUAWithoutReassociation*) can help improve the performance.

5.4 Conclusions

In this chapter, we further extended the static optimisation framework to a dynamic framework to compare the performance of PSD and ABS under different traffic models. We considered two traffic models as the bases of our comparisons, i.e., fixed-delay and file-download, with a respective performance metric of average $\alpha-$mean throughput and per-user delay. We obtained tight upper-bounds and lower-bounds for the system performance in terms of the two performance metrics, respectively, and showed the dominance of PSD over ABS for different values of α. Furthermore, we showed that while re-association in the fixed-delay scenario (under both PSD and ABS) does not result in much gain in performance metric, it significantly improves the metric in the file-download scenario. Lastly, we showed the inefficiency of SCF in the file-download scenario.

Chapter 6

Device-centric α-fair User Association under PSD

Summary: In this chapter, we

- revisit the problem of user association under PSD in a dynamic setup,
- propose a simple device-centric re-association rule where users are periodically and individually given a chance to re-associate to another BS in a greedy manner,
- show that the proposed scheme outperforms a very good physical-layer based UA scheme, i.e., SCF, and can reduce the loss in performance with respect to the global network-centric approach.

6.1 Introduction

In the previous chapters, we focused on the joint optimisation of the three network processes under PSD and ABS. Using extensive numerical and simulation results, we obtained the respective upper and lower-bounds for the average $\alpha-$mean throughput and delay under static and dynamic settings for both of the RA schemes. We showed the upper hand of PSD over ABS both in terms of 1) ease of optimal α-fair scheduling and 2) achievable bounds for the system performance. To verify whether the PSD bounds are achievable using simple UA schemes (along with the optimal US), we also studied the performance of a very good physical-layer based UA scheme, i.e., SCF. In Chapter 5, we saw that although SCF performs quite well for the fixed-delay traffic model, it performs quite poorly for the file-download model. The reason for this is as follows. As we showed in Chapter 3, if k is fine-tuned, the average $\alpha-$fair throughputs of a fixed number of users (per snapshot) over a large number of snapshots performs close to optimality with SCF. Likewise, we should expect a similar behaviour in a dynamic setting with a fixed number of users (per scheduling interval) over a large number of scheduling intervals as is the case in the fixed-delay scenario since the number of users in any scheduling interval is approximately equal to $\lambda \times \hat{t}$ in the long run. However, for the file-download model, the situation is quite different since the number of users in different scheduling intervals may significantly vary[1] as the users' departure time will depend on the other users' service-time. Therefore, the expected behaviour of the system under the file-download model is quite different from that of the static setting or fixed-delay model under the dynamic setting. Hence, new UA schemes are required to enhance the performance of the system with such a traffic model. A good UA under this traffic model should ideally consider 1) the per-BS load (as well as the physical-layer measurements), and 2) the network's criteria (e.g., the fairness criteria) [15, 13]. In Chapter 5, we studied a network-centric (joint RA, US and) UA scheme, namely, *PSD-SUAWithoutReassociation* that encompassed both of these elements and provided a lower-bound in terms of the per-user average delay among the class of all physical-layer based UA schemes (e.g., SCF). Moreover, we saw that re-association along with this UA scheme can significantly help improve the performance metric. Motivated by these two observations, in this chapter, we focus on the problem of UA in dynamic systems, first discussed in [15], along with re-association of users under PSD.

In view of the poor performance of physical-layer based UA schemes, the authors in [15] consider the problem of *PSD-SUAWithoutReassociation* (presented in Chapter 5). As we discussed before, solving such a problem to optimality would be prohibitively complex as a network-centric online system because of the integrality constraints. Therefore, as a departure from this network-centric approach, the authors propose a simple *device-centric*[2] UA scheme where an association decision is taken individually for each user only at its own arrival instant. The proposed scheme is guaranteed to maximise the $\alpha-$fair throughput of

[1]Note that the (long-term) average number of users in the system will still be about $\lambda \times \hat{t}$. However, the number of users may significantly vary in different scheduling intervals.

[2]In a device-centric UA scheme, a user makes the UA decision itself with the help of its own physical-layer measurements and possibly some extra information about the state of the network broadcast by the BSs.

the network upon a user's arrival (assuming the rest of the users retain their association) and achieves the upper (and/or lower) bounds provided by *PSD-SUAWithoutReassociation*. As well as being a simple device-centric approach, the proposed UA scheme is scalable in the sense that it only requires a constant and small amount of network-information irrespective of the number of active users in the system. However, it is limited to the case where re-association of users is not allowed and, hence, still retains a large gap from the benchmark provided by *PSD-SUAWithReassociation*. In this chapter, our objective is to incorporate the idea of re-association in the UA scheme proposed by [15] and to characterise the system performance under the new scheme.

Next, we will outline a framework for all device-centric UA schemes and, then, introduce the UA scheme proposed by [15]. Throughout this chapter, we use the same notations and assumptions as in Chapter 5 unless otherwise specified.

6.2 Device-centric user association rules

UA schemes are generally divided into two paradigms: Network-centric or device-centric. In network-centric UA schemes, the association decisions are made by the network based on the channel state information provided by the users, and are reported to the users. The association decisions can be made locally at each BS or jointly among multiple BSs through cooperation, or a central controller in C-RAN. An example of a network-centric UA scheme is *PSD-SUAWithReassociation* described in Chapter 5 where the UA variables are jointly computed by all BSs. In device-centric UA schemes, the association decisions are made by the users themselves (usually individually and independently of other users) using their channel state measurements and possibly some extra information about the network broadcast by the BSs. Most of the existing UA schemes are device-centric, e.g., Max-SINR, SCF, and typically require a user to take periodic physical-layer measurements (e.g., SINR) from the nearby BSs and associate with one of them according to a rule.

The benefit of a device-centric approach is that it is often very simple and scalable (in the sense that the complexity of the UA scheme does not change with the number of users). However, most of the existing device-centric UA rules are purely based on physical-layer measurements and can lead to poor network performance (e.g., under file-download model in Chapter 5). If we want to improve a user's association decision, it is necessary that some additional information about the state of the network, e.g., per-BS load or network's fairness criterion, is available. However, providing this network-dependent information will potentially create higher signalling cost and complexity in network design. Therefore, ideally, we would want a device-centric UA to require very little information from the network and perform as well as *PSD-SUAWithReassociation*. If this is not possible, finding the right trade-off between performance and complexity is important.

6.2.1 A general framework

In order to study the trade-off between performance and complexity, we outline a common framework for all device-centric UA schemes [15]: *User u, at an association event u (e.g, an arrival instant a_u or a departure instant d_u), performs the association decision based on 1) its own link measurements (e.g., SINR), 2) network-provided information (e.g., per-BS load), and 3) a given rule. The node-specific roles in this framework are outlined below.*

BS j broadcasts a set of BS-specific information $Info_j$ periodically to assist the users to make their UA decisions. Deciding on which information should be rebroadcast (and how often) is part of the system design.

User u

1. measures channel-related information periodically, e.g., $\{\gamma_{ju}\}_{j\in\mathcal{B}\cup\{0_M,0_S\}}$,
2. uses the available BS-specific information and the channel-related measurements to decide on the best BS j_u^{ua}, based on some predefined rule $\zeta^{ua}(.)$, e.g.,

$$j_u^{ua} = \zeta^{ua}(\{Info_j\}_{j\in\mathcal{B}\cup\{0_M,0_S\}}, \{\gamma_{ju}\}_{j\in\mathcal{B}\cup\{0_M,0_S\}}),$$

3. sends the association request to BS j_u^{ua}.

We identify the following three aspects as the most important features while designing online UA schemes.

1. **Scalability**, i.e., constant amount of information per-BS irrespective of the number of active users. If the amount of the information that each BS has to broadcast increases with the number of users, the scheme is not scalable.
2. **Simplicity** in terms of computation and constant complexity of association decision with respect to the number of users in the system.

3. **Performance** with respect to the chosen network objective function, i.e., the global $\alpha-$fairness objective.

Next, we will outline how such a UA scheme with the above-mentioned features can be carried out in the device-centric framework.

6.3 System model

We use the same system model as in Section 5.1 with the same notations and assumptions unless otherwise specified. Consider the *PSD-SUAWithoutReassociation* problem (presented in Chapter 5). Let the system utility be $\Gamma(u, a_u, j) = \sum_{i\in\mathcal{U}(a_u^-)\cup\{u\}} U_\alpha(T_i(a_u))$ if user u selects BS j at her arrival time a_u, where the $T_i(a_u)$'s (user i's throughput at her arrival time) are easily computed using the $\alpha-$fair schedules given in Remark 2 of Chapter 3 since all the associations are known. The best that the user can do at time a_u is to select the BS $j^*(u, a_u) = \arg\max_{j\in\mathcal{B}\cup\{0_S,0_M\}} \Gamma(u, a_u, j)$. By selecting $j^*(i, a_i)$ upon every user i's arrival time, the system can achieve the performance bound provided by *PSD-SUAWithoutReassociation*.

With this objective to maximise the sum of the utilities at each new arrival instant, the authors in [15] propose an *optimal individual*[3] UA scheme. We refer to this UA scheme as *OptIndividualUA* and summarise it in the following theorem.

Theorem 1. *[15, Chapter 7] For $\alpha = 1$, user u at her association event a_u, i.e., her arrival time, chooses the BS based on the following simple formula to maximise the global sum of user utilities.*

$$j^*(u, a_u) = \arg\max_{j\in\mathcal{B}\cup\{0_S,0_M\}} \log(r_{ju}) + \log\left(\frac{U_j(a_u^-)^{U_j(a_u^-)}}{(U_j(a_u^-)+1)^{U_j(a_u^-)+1}}\right), \tag{6.1}$$

where $U_j(a_u^-)$ is the number of users associated with BS j just before the association event at time a_u. For $\alpha \neq 1$, let

$$O_j(\alpha, t) = \frac{1}{1-\alpha}\left(\sum_{i\in\mathcal{A}_j(t)} r_{ji}^{\frac{1-\alpha}{\alpha}}\right)^{\alpha} \tag{6.2}$$

[3] Since it is 'optimal' with respect to *PSD-SUAWithoutReassociation* and UA decisions are carried out for users 'individually'.

represent the local utility at BS j where $\mathcal{A}_j(t)$ is the set of users associated with BS j at time t. Then, the optimal decision for u at association time a_u is to select $j^*(n, a_u)$ such that

$$j^*(u, a_u) = \arg \max_{j' \in \mathcal{B} \cup \{0_S, 0_M\}} \sum_{j \in \mathcal{B} \cup \{0_S, 0_M\}} O_j^{u \to j'}(\alpha, a_u),$$

where $O_j^{u \to j'}(\alpha, a_u)$ represents the new value of local utility at BS j when user u decides to associate with BS j' at time a_u.

Theorem 1 shows that, *OptIndividualUA* scheme can be carried out if users have the following information: 1) One scaler quantity $\sum_{i \in \mathcal{A}_j(t)} r_{ji}^{\frac{1-\alpha}{\alpha}}$ per BS, 2) the RA parameter k, and 3) parameter α and rate-function $f(.)$, indicating that the device-centric *OptIndividualUA* scheme satisfies all the three desirable features listed before.

Although *OptIndividualUA* is a simple and device-centric UA scheme which yields the best system performance with respect to *PSD-SUAWithoutReassociation*, it still retains a large gap with respect to *PSD-SUAWithReassociation* as we showed in Chapter 5. In the following, we propose a simple scheme based on Theorem 1 where we allow each user to periodically change her association independent of other users in the system.

6.4 Proposed UA scheme

Let τ_u represent the set of periodic association instants for user u, i.e., $\tau_u = \{a_u,\ a_u + t_u,\ a_u + 2t_u, ...\}$, where t_u is the periodicity of user u's associations. Then, assuming that t_u is known to the user and BSs a priori, if each user u at each $\tilde{t} \in \tau_u$ selects $j^*(u, \tilde{t}) = \arg\max_{j \in \mathcal{B} \cup \{0_S, 0_M\}} \Gamma(u, \tilde{t}, j)$, we expect that the system performance will improve since periodic re-association can improve on the myopic decisions as the network evolves. Based on this, we propose the following UA scheme, a.k.a. *OptIndividualUA(reassoc)*.

Each user u at her association event $\tilde{t} \in \tau_u$ chooses the BS $j^*(u, \tilde{t})$ based on the following simple formula

$$j^*(u, \tilde{t}) = \begin{cases} \arg\max_{j \in \mathcal{B} \cup \{0_S, 0_M\}} \log(r_{ju}) + \log\left(\frac{U_j(a_u^-)^{U_j(a_u^-)}}{(U_j(a_u^-)+1)^{U_j(a_u^-)+1}}\right), & \text{if } \alpha = 1 \\ \arg\max_{j' \in \mathcal{B} \cup \{0_S, 0_M\}} \sum_{j \in \mathcal{B} \cup \{0_S, 0_M\}} O_j^{u \to j'}(\alpha, a_u), & \text{if } \alpha \neq 1. \end{cases} \tag{6.4}$$

Note that our proposed algorithm still allows each user to individually change her association as subsequent association events are still independent from other users (unlike in the network-centric approach where user association for a number of users are computed jointly).

Although the above-mentioned UA scheme might not achieve the benchmark provided by *PSD-SUAWithReassociation*, we expect that it will help improve the performance with respect to the benchmark at the cost of a (slight) increase in the handover signalling depending on the frequency of the users' re-associations.

Next, we present the average delay performance of *OptIndividualUA(reassoc)* for the file-download scenario introduced in Section 5.3.3.

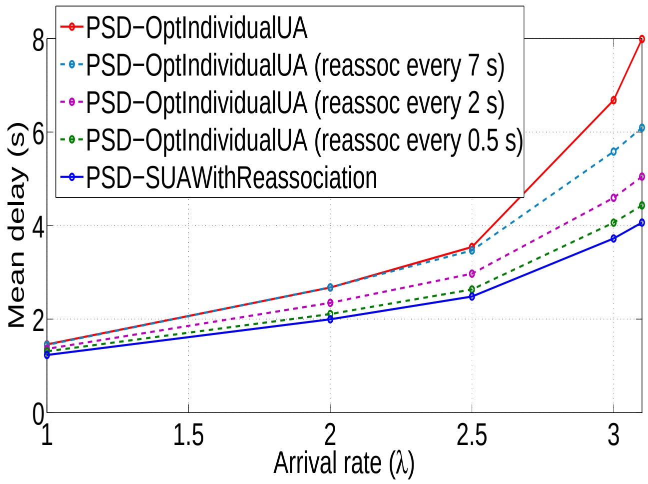
6.5 Numerical results

In this section, we compare the average delay performance of *OptIndividualUA(reassoc)* to that of *PSD-SUAWithReassociation* (the benchmark) for the file-download scenario (presented in Section 5.3.4). The parameter settings are assumed to be the same as the ones in Section 5.3 unless otherwise specified. We set t_u to a fixed value from the set $\{0.5, 2, 7\}$ *sec* for all users.

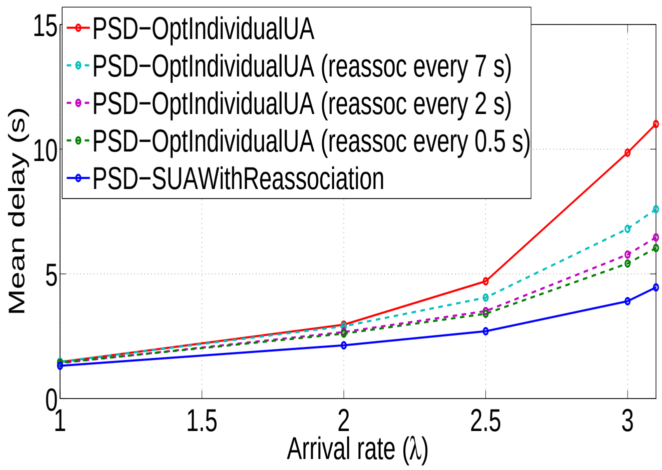
Fig. 6.1 shows the average delay performance of *OptIndividualUA(reassoc)* (labelled as *OptIndividualUA (reassoc every* t_u *s)* in the figures) with respect to the benchmark. As we increase the periodicity, the performance metric significantly improves, achieving the smallest gap with respect to the benchmark with periodicity of 500 ms. This shows that with some increase in the re-association frequency the proposed OptIndividualUA in [15] can perform very well. Similar observations can be seen for OD as shown in Fig. 6.2.

6.6 Conclusions

In conclusion, we studied the problem of association under PSD in a dynamic setting and proposed a simple device-centric re-association rule where users are periodically given a chance to re-associate to another BS. We showed that the proposed scheme outperforms a very good physical-layer based scheme and helps reduce the loss in performance with respect to the network-centric approach.

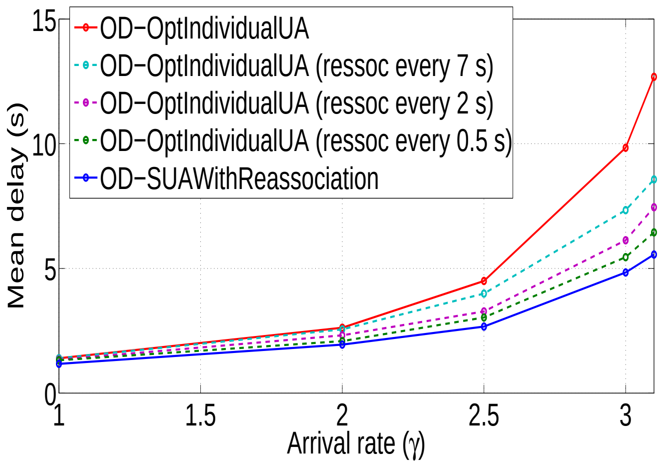


(a) PSD - $\alpha = 1$

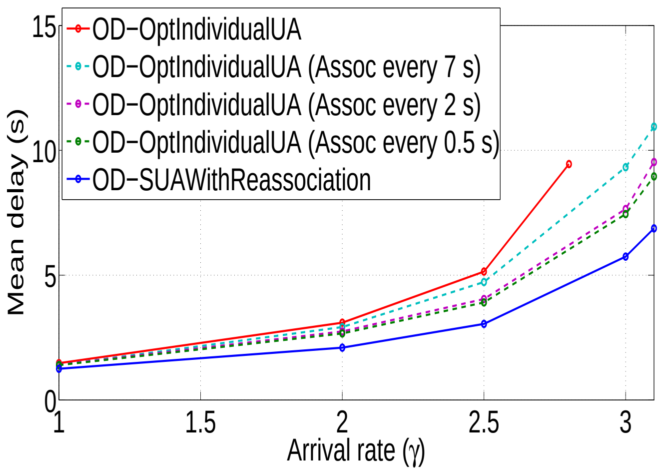


(b) PSD - $\alpha = 2$

Figure 6.1: Mean delay as a function of λ. Settings: $M = 33$, $F = 10\ Mbit$.



(a) OD - $\alpha = 1$



(b) OD - $\alpha = 2$

Figure 6.2: Mean delay as a function of λ. Settings: $M = 33$, $F = 10$ $Mbit$.

Chapter 7

Conclusions

7.1 Summary

In summary, we studied the performance of an OFDM-based HetNet on the downlink. We saw that the network performance depends on three coupled network processes, namely, RA, UA and US. Studying these processes under a unified framework was one of the main objectives of this thesis and gave us a platform to compare the system performance under different deployment schemes and HetNet configurations. In the first part of the thesis (Chapters 3-4), we revisited a snapshot-based unified framework (which jointly optimised the three network processes under PSD) in a static setting and made two amendments to it in terms of channel modelling and user scheduling. We, then, revisited a similar framework but under ABS and showed that $\alpha \neq 1-$fair scheduling under ABS can be much harder than that under PSD. We, then, proceeded to obtain upper-bounds for the throughput performance of the HetNet under the two RA schemes and showed the upper-hand of PSD over ABS under different HetNet configurations. Furthermore, we showed that SCF, a simple physical-layer based UA scheme, performs very well if the RA and biasing parameters are fine-tuned.

Although our static framework gave us some valuable insights in terms of the performance of the two RA schemes and different UA heuristics, it was limited to an offline-static study and did not necessarily reflect the behaviour of an online system including all of its dynamics, e.g., users arrival and departure. This was the motivation behind developing the second part of this thesis.

In the second part of our thesis (Chapters 5-6), we further extended our static framework to a dynamic one where we studied the performance of the two RA schemes under two service-time scenarios: 1) fixed-file and 2) file-download. Using extensive simulation results, we showed that under both of the service-time scenarios PSD outperforms ABS, reverifying our observations from the static framework. Furthermore, the dynamic setting provided a number of important insights, most notably, the fact that SCF does not work well under the file-download scenario and that re-association can potentially help improve the system performance. Motivated by these two observations, we further investigated the problem of UA for PSD under the dynamic framework. Building upon a previous work in the literature, we proposed a computationally simple and scalable device-centric UA schemes where users were periodically given a chance to re-associate to another BS and, so, significantly improved the system performance.

7.2 Future research work

In our study, we made a number of restrictive assumptions, most importantly, in terms of users association (SUA) and transmission link (downlink). As we saw in Chapter 6, periodic re-association of users can lead to significant performance gains. However, such periodic re-associations can potentially cause excessive handover signalling and possible create complexities in the system design. Recently, with an increased interest in the development of C-RAN in future HetNets, a potential future research path to performance enhancement would be to allow multi-BS user association (MUA) where a given user can associate to multiple BSs simultaneously. Such an association deployment can potentially lift the need for periodic re-association as a user now, served by multiple BSs at the same time, would likely leave the system before a re-association is deemed necessary. Hence, the need for unnecessary handover signalling may be eliminated.

In this study, we also restricted ourselves to the downlink transmissions only. An equally important problem would be to consider uplink and investigate the performance gains of different RA schemes and UA heuristics. However, unlike the downlink model where we managed to decouple the global $\alpha-$fair objective into independent local (per-BS) problems, decoupling the global problem for the uplink may be very difficult as, on one hand, because of the limited power budget of user equipments pure time-domain scheduling (i.e., allowing users to transmit on all available channels) may not be a feasible option and, so, sub-channels should be assigned to users from a global perspective. On the other hand, even with a fixed sub-channel assignment, the interference (and, hence, SINR) at one BS will depend on the schedulers at the other BSs and vice versa. Hence, decoupling the global scheduling problem on the uplink may not be as easy as the one on the downlink.

Bibliography

[1] R1-100701 3GPP. Importance of serving cell association in HetNets. January 2010.

[2] 3GPP Tech. Rep. TR 36.814. Evolved Universal Terrestrial Radio Access (E-UTRA); Further Advancements for E-UTRA Physical Layer Aspects. 2010.

[3] 3rd Generation Partnership Project. Evolved universal terrestrial radio access (e-utra), carrier aggregation, base station (bs) radio transmission and reception. *Sophia Antipolis, France, Tech. Rep. 3GPP TR 36.808*, 2013.

[4] 3rd Generation Partnership Project. New work item description: Dual connectivity for lte, 3gpp tsg-ran meeting 62. 2013.

[5] J.G. Andrews, H. Claussen, M. Dohler, S. Rangan, and M.C. Reed. Femtocells: Past, present, and future. *Selected Areas in Communications, IEEE Journal on*, 30(3):497–508, April 2012.

[6] A. Bedekar and R. Agrawal. Optimal muting and load balancing for eicic. pages 280–287, May 2013.

[7] G. Boudreau, J. Panicker, Ning Guo, Rui Chang, Neng Wang, and S. Vrzic. Interference coordination and cancellation for 4g networks. *Communications Magazine, IEEE*, 47(4):74–81, April 2009.

[8] T. Bu, Li Li, and R. Ramjee. Generalized proportional fair scheduling in third generation wireless data networks. pages 1–12, April 2006.

[9] V. Chandrasekhar, J.G. Andrews, and Alan Gatherer. Femtocell networks: a survey. *Communications Magazine, IEEE*, 46(9):59–67, September 2008.

[10] A. Damnjanovic, J. Montojo, Yongbin Wei, Tingfang Ji, Tao Luo, M. Vajapeyam, Taesang Yoo, Osok Song, and D. Malladi. A survey on 3gpp heterogeneous networks. *Wireless Communications, IEEE*, 18(3):10–21, June 2011.

[11] S. Deb, P. Monogioudis, J. Miernik, and J. P. Seymour. Algorithms for enhanced inter-cell interference coordination (eicic) in lte hetnets. *IEEE/ACM Transactions on Networking*, 22(1):137–150, Feb 2014.

[12] E. Dahlman, S. Parkvall and J. Sköld. *4G LTE/LTE-Advanced for Mobile Broadband*. Elsevier Ltd, Oxford, UK, 2011.

[13] D. Fooladivanda. *Comparison between Static and Dynamic Modeling Approaches for Heterogeneous Cellular Networks*. Phd thesis, University of Waterloo, Waterloo, Canada, August 2014.

[14] D. Fooladivanda and C. Rosenberg. Joint Resource Allocation and User Association for Heterogeneous Wireless Cellular Networks. *IEEE Transactions on Wireless Communications*, 12(1):248–257, January 2013.

[15] J. Ghimire. *Heterogeneous Cellular Networks: from Resource Allocation to User Association*. PhD thesis, University of Waterloo, 2015.

[16] J. Ghimire and C. Rosenberg. Impact of limited backhaul capacity on user scheduling in heterogeneous networks. pages 2480–2485, April 2014.

[17] J. M. Jornet I. F. Akyildiz and C. Hana. Terahertz band: Next frontier for wireless communications. *Physical Communication*, 2014.

[18] Informa. Small cell market status, issue i. 2013.

[19] Y. Jin and L. Qiu. Joint User Association and Interference Coordination in Heterogeneous Cellular Networks. *IEEE Communications Letters*, 17(12):2296–2299, December 2013.

[20] F. Kelly. Charging and rate control for elastic traffic. *European transactions on Telecommunications*, pages 33–37, 1997.

[21] A. Khandekar, N. Bhushan, Ji Tingfang, and V. Vanghi. Lte-advanced: Heterogeneous networks. pages 978–982, April 2010.

[22] Li Li, M. Pal, and Y.R. Yang. Proportional fairness in multi-rate wireless lans. April 2008.

[23] D. Lopez-Perez, Ming Ding, H. Claussen, and A.H. Jafari. Towards 1 gbps/ue in cellular systems: Understanding ultra-dense small cell deployments. *Communications Surveys Tutorials, IEEE*, 17(4):2078–2101, Fourthquarter 2015.

[24] D. Lopez-Perez, I. Guvenc, and X. Chu. Mobility management challenges in 3gpp heterogeneous networks. *Communications Magazine, IEEE*, 50(12):70–78, December 2012.

[25] D. Lopez-Perez, I. Guvenc, and X. Chu. Mobility management challenges in 3gpp heterogeneous networks. *IEEE Communications Magazine*, 50(12):70–78, December 2012.

[26] C. Mehlfuhrer, M. Wrulich, J. Ikuno, D. Bosanska, and M. Rupp. Simulating the long term evolution physical layer. August 2009.

[27] J. Mietzner, R. Schober, L. Lampe, W. H. Gerstacker, and P. A. Hoeher. Multiple-antenna techniques for wireless communications - a comprehensive literature survey. *Communications Surveys Tutorials, IEEE*, 11(2):87–105, Second 2009.

[28] J. Mo and J. Walrand. Fair end-to-end window-based congestion control. *IEEE/ACM Transactions on Networking*, 8(5):556–567, Oct 2000.

[29] D.P. Palomar and Mung Chiang. A tutorial on decomposition methods for network utility maximization. *Selected Areas in Communications, IEEE Journal on*, 24(8):1439–1451, Aug 2006.

[30] K. I. Pedersen, P. H. Michaelsen, C. Rosa, and S. Barbera. Mobility enhancements for lte-advanced multilayer networks with inter-site carrier aggregation. *IEEE Communications Magazine*, 51(5):64–71, May 2013.

[31] N. Prasad, M. Arslan, and S. Rangarajan. Exploiting Cell Dormancy and Load Balancing in LTE HetNets: Optimizing the Proportional Fairness Utility. *IEEE Transactions on Communications*, 62(10):3706–3722, 2014.

[32] Qualcom. The 1000x Mobile Data Challenge: More Spectrum, More Small Cells, More Indoor Cells and Higher Efficiency. 2012.

[33] RAN1 Chairman (R1-105779). Way Forward on time-domain extension of Rel 8/9 backhaul-based ICIC. 2010.

[34] K. Somasundaram. Proportional Fairness in LTE-Advanced Heterogeneous Networks with eICIC. pages 1–6, Sept 2013.

[35] K. Son, S. Chong, and G. D. Veciana. Dynamic association for load balancing and interference avoidance in multi-cell networks. *IEEE Transactions on Wireless Communications*, 8(7):3566–3576, July 2009.

[36] R. Srikant. *The mathematics of Internet congestion control*. Springer, 2004.

[37] P. Wang, H. Jiang, W. Zhuang, and H. V. Poor. Redefinition of max-min fairness in multi-hop wireless networks. *IEEE Transactions on Wireless Communications*, 7(12):4786–4791, December 2008.

[38] W. Webb. *Wireless Communications: The Future*. John Wiley and Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex PO19 8SQ, England, 2007.

[39] Q. Ye, M. Al-Shalash, C. Caramanis, and J. G. Andrews. On/off macrocells and load balancing in heterogeneous cellular networks. pages 3814–3819, 2013.

[40] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. G. Andrews. User association for load balancing in heterogeneous cellular networks. *IEEE Transactions on Wireless Communications*, 12(6):2706–2716, 2013.