POLITECNICO DI TORINO
SCUOLA DI DOTTORATO

Dottorato in Ingegneria Elettronica e delle Comunicazioni - XXVIII ciclo

PhD Dissertation

Resource and power management in next generation networks

Adivsor:
Prof. Carla-Fabiana Chiasserini

Candidate:
Zana Limani Fazliu

February 2017

Summary

The limits of today's cellular communication systems are constantly being tested by the exponential increase in mobile data traffic, a trend which is poised to continue well into the next decade. Densification of cellular networks, by overlaying smaller cells, i.e., micro, pico and femtocells, over the traditional macrocell, is seen as an inevitable step in enabling future networks to support the expected increases in data rate demand. Next generation networks will most certainly be more heterogeneous as services will be offered via various types of points of access (PoAs). Indeed, besides the traditional macro base station, it is expected that users will also be able to access the network through a wide range of other PoAs: WiFi access points, remote radio-heads (RRHs), small cell (i.e., micro, pico and femto) base stations or even other users, when device-to-device (D2D) communications are supported, creating thus a multi-tiered network architecture. This approach is expected to enhance the capacity of current cellular networks, while patching up potential coverage gaps. However, since available radio resources will be fully shared, the inter-cell interference as well as the interference between the different tiers will pose a significant challenge. To avoid severe degradation of network performance, properly managing the interference is essential. In particular, techniques that mitigate interference such Inter Cell Interference Coordination (ICIC) and enhanced ICIC (eICIC) have been proposed in the literature to address the issue. In this thesis, we argue that interference may be also addressed during radio resource scheduling tasks, by enabling the network to make interference-aware resource allocation decisions.

Carrier aggregation technology, which allows the simultaneous use of several component carriers, on the other hand, targets the lack of sufficiently large portions of frequency spectrum; a problem that severely limits the capacity of wireless networks. The aggregated carriers may, in general, belong to different frequency bands, and have different bandwidths, thus they also may have very different signal propagation characteristics. Integration of carrier aggregation in the network introduces additional tasks and further complicates interference management, but also opens up a range of possibilities for improving spectrum efficiency in addition to enhancing capacity, which we aim to exploit.

In this thesis, we first look at the resource allocation in problem in dense multi-tiered networks with support for advanced features such as carrier aggregation and device-to-device communications. For two-tiered networks with D2D support, we propose a centralised, near optimal algorithm, based on dynamic programming principles, that allows a central scheduler to make interference and traffic-aware scheduling decisions, while taking into consideration the short-lived nature of D2D links. As the complexity of the central scheduler increases exponentially with the number of component carriers, we further propose a distributed heuristic algorithm to tackle the resource allocation problem in carrier aggregation enabled dense networks. We show that the solutions we propose perform significantly better than standard solutions adopted in cellular networks such as eICIC coupled with Proportional Fair scheduling, in several key metrics such as user throughput, timely delivery of content and spectrum and energy efficiency, while ensuring fairness for backward compatible devices.

Next, we investigate the potentiality to enhance network performance by enabling the different nodes of the network to reduce and dynamically adjust the transmit power of the different carriers to mitigate interference. Considering that the different carriers may have different coverage areas, we propose to leverage this diversity, to obtain high-performing network configurations. Thus, we model the problem of carrier downlink transmit power setting, as a competitive game between teams of PoAs, which enables us to derive distributed dynamic power setting algorithms. Using these algorithms we reach stable configurations in the network, known as Nash equilibria, which we show perform significantly better than fixed power strategies coupled with eICIC.

Acknowledgements

First and foremost, I would like to thank my advisor Prof. Dr. Carla Fabiana Chiasserini, without whose support, encouragement and guidance, this thesis would have not been possible. As my academic advisor through these last 4 years, she has been a mentor, a role model and a friend at the same time. Her unwavering support and understanding were crucial to the successful outcome of this journey.

I am sincerely grateful to Telecom Italia, whose generous scholarship, made it possible for me to pursue this PhD, but especially I want to thank Telecom's team: Gian Michele dell'Aera, Marco Caretti and Bruno Melis for the excellent collaboration. They have significantly contributed to this thesis, through the engaging and fruitful discussions and their readiness to offer insight into practical aspects of cellular networks.

Special thanks go to everyone in the Telecommunications Networking Group at Politecnico di Torino, and especially to Francesco Malandrino for the valuable and helpful collaboration as I took my first steps in academic research, and above all, for being a friend.

I would like to also thank my external co-advisor at University of Prishtina, Prof. Dr. Enver Hamiti, who was extremely supportive during the period spent there, and everyone in the Telecommunications Department at the Faculty of Electrical and Computer Engineering.

Last but not least, I want to thank my family. My parents, for offering me unending love and support, and for their unrelenting belief in me. My sisters, for always being there, lending an ear, and reading and proofreading all my writings even when it bored them to death. My best friend, Hëna, for the incredible support she has shown, and for being an amazing godmother to my son. And finally, this thesis is dedicated to the two most important people in my life, my husband Alban and my son Mal, who have made it all worthwhile.

Contents

List of Tables

List of Figures

List of Algorithms

List of publications

(1) F. Malandrino, **Z. Limani**, C. Casetti, C.-F. Chiasserini, "Interference-Aware Downlink and Uplink Resource Allocation in HetNets with D2D Support," IEEE Transactions on Wireless Communications, vol. 14 n. 5, pp. 2729-2741, 2015.

(2) **Z. Limani Fazliu**, C.-F. Chiasserini, G. M. Dell'Aera, "Downlink Transmit Power Setting in LTE HetNets with Carrier Aggregation", IEEE WoWMoM, June 2016 (best paper award).

(3) **Z. Limani**, C.-F. Chiasserini, G. M. Dell'Aera,"Interference-Aware Resource Scheduling in LTE HetNets with Carrier Aggregation Support", IEEE ICC June 2015.

(4) F. Malandrino, C. Casetti, C.-F. Chiasserini, **Z. Limani**, "Fast Resource Scheduling in HetNets with D2D Support," IEEE INFOCOM, April 2014.

(5) F. Malandrino, C. Casetti, C.-F. Chiasserini, **Z. Limani**, "Uplink and downlink resource allocation in D2D-enabled heterogeneous networks," IEEE WCNC, April 2014.

(6) **Z. Limani Fazliu**, C.-F. Chiasserini, G. M. Dell'Aera, E. Hamiti, "Downlink Power Setting in Carrier Aggregation Enabled Dense Networks", submitted to IEEE Transactions on Wireless Networks (October 2016).

Part I

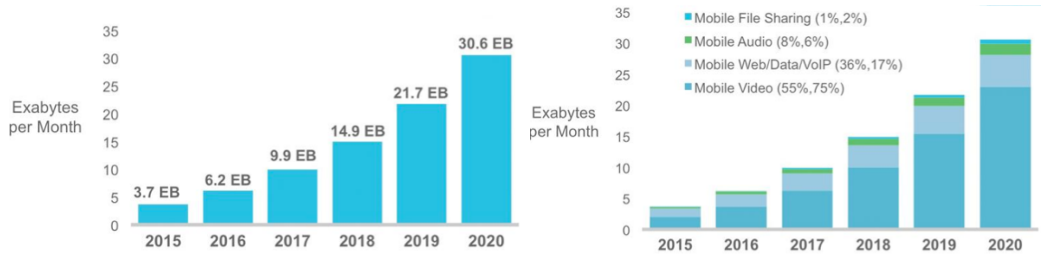
Introduction and background

Chapter 1

Introduction

The exponential increase in mobile data traffic demand in recent years, due to the proliferation of wireless gadgets and evolution of smartphones, has become a serious challenge for today's cellular communication networks. It is clear that there is a limit to what today's networks with the current cellular infrastructure can achieve both in terms of capacity and coverage. Indeed, they are not flexible enough for the ever-changing behaviour of users who now have access to a diverse range of applications on their smartphones and tablets, and who are ultimately more demanding in terms of user experience. According to Cisco Visual Networking Index [1], global mobile data traffic grew by 74 percent in 2015, and this trend is expected to continue. As shown in Fig. 1.1(a), data traffic is expected to exceed 30 exabytes by 2020 [1]. In addition, as Fig. 1.1(b) shows, video content is increasingly growing more popular; it already accounted for more than half of the traffic in 2015, while it is expected to represent more than 75% of traffic by 2020. This has far reaching implications in terms of data rates networks will be expected to offer, considering that video content usually has much higher bit rates than other mobile content types. As noted in [1] video usage, unlike other types of content usage, tends to occur during "prime time", which will result in more traffic during the, already congested, peak hours of the day.

To stave off this impending "capacity crunch", researchers from industry and academia, have proposed several strategies, introducing both new technologies and enhancing old ones. Intensive research has been ongoing especially in key enabling technologies for future 5G networks. One such strategy, already gaining traction in 4G networks, such as LTE-A, is the deployment of dense heterogeneous networks (HetNets). Indeed densification of the network, by overlaying smaller cells over the traditional macrocell, is seen as an inevitable step in enabling future networks to support the expected increase in data rate demand. Additionally, networks will become more heterogeneous as users will be able to access the services through a range of points of access (PoAs) besides the traditional macro base station, e.g., through



(a) (b)

Figure 1.1. a) Mobile data traffic forecast; b) Mobile data traffic forecast for different types of content. Figures in parentheses refer to 2015 and 2020 traffic share [1].

WiFi access points, small cell (i.e., micro, pico and femto) base stations, or even other users when device-to-device communications are supported [2]. This approach will improve both the capacity and coverage by closing the gap between the access network and the user [3]. Further heterogeneity in the network will be introduced as cellular communications advance and expand to mmWave bands. Integration of mmWave communications within the cellular network will pose significant technical challenges due to the particularly hostile propagation environment in these bands, however there is general consensus that the benefits of these technologies easily justify the additional hurdles [4].

Another important impediment in the current wireless environment is the lack of available contiguous portions of frequency spectrum, which would be able to accommodate high user data rates. Carrier aggregation (CA) is seen as one of the key features that aims at overcoming this challenge. Indeed, it enables future cellular networks to concurrently use several LTE component carriers (CC) so as to provide aggregated bandwidths as high as 100 MHz. The individual component carriers can be of different bandwidths and, in general, may belong to different frequency bands (e.g., 800 MHz, 1.8 GHz and 2.6 GHz) [5].

While densification and heterogeneity promise to greatly improve the network performance, there are several technical challenges that have to be overcome to ensure that these technologies perform to the maximum of their potentials. The most prominent is the inter-tier and inter-cell interference, which is bound to result from intense frequency reuse in neighboring or overlapping cells. Indeed as spectrum is in short-supply, it is most likely that future cellular networks, like LTE-A, will be designed to operate with a frequency reuse factor of 1, meaning that all PoAs, even those with overlapping coverage areas, will be sharing the same radio resources. The interference in such cases may be so severe as to significantly limit system throughput performance. Careful management of inter-cell and inter-tier interference is

therefore key to enable optimal use of advanced technologies. Interference mitigation techniques, such as Inter-cell Interference Coordination (ICIC) and Enhanced ICIC (eICIC) have been proposed for 4G networks such as LTE-A, and already included in the specifications. However such techniques in general do not pay special attention to advanced features such as carrier aggregation or direct device-to-device (D2D) communications. Since aggregated component carriers may belong to different frequency bands, they will have different coverage areas as well different impact with respect to interference, due to their diverse propagation characteristics. D2D links on the other hand are short-range, with great potential for spectrum spatial reuse, but fickle availabality due to user mobility.

In this thesis we argue that the diversity offered by the different component carriers, as well as the spatial diversity offered by D2D communications, can be effectively leveraged to mitigate the interference, either during the resource scheduling tasks performed by the network, or by applying solutions which adjust the downlink transmit power of the different carriers. In the first part of the thesis we focus on tackling inter-tier and inter-cell interference through interference-aware resource allocation techniques. To this end we propose a centralised interference-aware scheduling algorithm for dense networks with D2D communications, derived by applying approximate dynamic programming techniques (ADP) to reduce the complexity. The algorithm takes into account interference and availability of D2D links to reach scheduling decisions which significantly improve network performance. In particular, the approach is deft at taking advantage of user proximity to alleviate the traffic load, especially for viral and video traffic. Next, to address resource scheduling in a CA-enabled multi-tier networks, a heuristic algorithm is proposed, implementable both in centralised and distributed manner. The algorithm, operating in an interference-aware manner, is able to take full advantage of the diversity offered by the different carriers, while ensuring fairness for users without CA capabilities (legacy users).

In the second part of the thesis, aiming to further exploit the diversity in coverage offered by the different component carriers, we allow the infrastructure nodes to lower and dynamically adjust the transmit powers of the different component carriers. We model the problem of carrier downlink transmit power as a team-based competitive game, with the objective of devising distributed and dynamic solutions which would enable a wide range of network configurations which reduce power consumption, provide high throughput and ensure a high level of coverage to network users.

1.1 Outline of the thesis

The remainder of this thesis is organised as follows:

In Chapter 2 we provide a summary of some of the main features of future cellular cellular networks which are central to our thesis. This way, we describe the context in which we formulate the problems we tackle in this thesis, and provide the reader with the necessary background information to easily appreciate the discussion that follows. Here we introduce the concept of dense heterogeneous networks, direct device-to-device communications and carrier aggregation. A brief summary of the user association and resource allocation tasks, which are performed by the network, is also provided.

In Chapter 3 we provide a comprehensive overview of the related work, in particular, works that address the same or related problems in future generation networks, works that we have used as reference, and works that have otherwise contributed to our research. A section is dedicated for each topic addressed in this work.

In Chapter 4 we address the resource allocation problem in a two-tiered dense heterogeneous network with support for device-to-device links. We first provide a detailed description of the network model under study and list any assumptions we make in our scenario. The introduced model is used in the successive chapters as well. We then propose a dynamic programming approach to address the problem at hand. A dynamic programming model is used to formulate and solve the resource allocation problem. The goal of our model is to decide i) which PoA in the network (a macro, a micro or another user) should serve ii) which traffic flows and iii) which radio resources should be used for that purpose. Each action in our model, which corresponds to a resource allocation strategy, is evaluated and assigned two values, one to account for the current cost and the other to account for future costs. This way, taking into account the future evolution of the traffic demand, we may appreciate the benefits of allowing some users to receive requested traffic sooner, so they would be able to then provide the same content to other nearby users. To reduce the complexity of the centralised problem, an approximate dynamic programming approach is proposed to reduce and control the size of the solution space. An additional algorithm is provided to reduce the complexity of calculating the cost incurred by future actions. Part of the work described in this chapter has been previously published in [6, 9, 10].

In Chapter 5 we address the resource allocation problem in a carrier aggregation enabled dense network. We expand the network model described in the previous chapter, to account for multiple carriers and allow for two possible implementations of carrier aggregation. Next, a heuristic interference-aware algorithm is proposed which jointly performs resource allocation and carrier selection. The algorithm, sifts through potential resource allocation strategies, and assigns *urgency* values to active traffic flows, to account for time constraints, and *pollution* values to potential serving cell and radio resource combinations which can be used to accommodate said traffic flows, to account for the interference. The two values are use to produce a single weight value for every potential combination of serving cell, traffic flow and radio

resource. The combinations with highest weight values are added to the resource allocation strategy every iteration. Part of the work described in this chapter has been previously published in [8].

In Chapter 6 we approach the interference management problem from a new perspective. Instead of tackling interference through resource allocation, we propose to leverage the diversity offered by the availability of the different carriers, by allowing nodes in the network to dynamically change and adjust the downlink transmit power settings of each carrier. We analyse the problem through the lens of game theory, which is an excellent mathematical tool to obtain a multi-objective, distributed solution in a scenario with entities (PoAs) sharing the same pool of resources (available CCs). The downlink transmit power setting is modelled as a competitive game between teams of PoAs, which we show belongs to the class of *pseudo-potential* games. As these games are known to possess pure Nash equiibria, which can be reached through best reply dynamics, we proceed to propose a distributed power setting algorithm which is applied at the team level. Teams exchange their decisions until convergence is reached, at which they point update their power settings. The algorithm is repeated with a specific update period, or can be triggered when teams notice significant changes in the user distribution or traffic demand. Part of the work described in this chapter has been previously published in [7, 11].

Finally, in Chapter 7 we recapitulate our main results, draw final conclusions and discuss potential follow-up topics.

Chapter 2

Features of next generation cellular networks

What does the future hold for future next generation cellular networks? No doubt, it is a question that many researchers are attempting to answer. Indeed massive research efforts are currently underway which aim at defining the design and performance objectives, identifying technical and operational challenges and proposing innovative technologies and solutions which would enable 5G networks to meet the challenge of the ever-growing traffic demand and avoid the looming "capacity crunch" [12]. The authors in [3] provide an overall summary of the challenges ahead and avenues for research on potential technologies to be used in 5G. A more in-depth survey is provided in [14], offering a comprehensive view on some of the potential 5G technologies, the current status and development of related research and open issues and challenges.

In this section we will provide only a brief summary of what the current vision for 5G networks is, and concentrate more thoroughly on some of the features which are central to this thesis.

2.1 5G vision

While technical requirements for next generation networks are still under development, there seems to be a degree of consensus between the different research initiatives from the industry and academia, about what are some of the broad expectations from 5G [3, 13–16]:

- *Significant increase in supported data rates.* 5G is expected to support data rates of up to 10 Gbps or more. This would imply a 1000x increase from 4G in terms of aggregate data rate, and around 100x in terms of edge rate [3] .

- *Reduced round trip latency.* It is expected that some future applications such as virtual and enhanced reality applications, two-way gaming and "tactile internet" [17] will require round trip latencies in the range of 1 ms.

- *More bandwidth.* Without increasing the bandwidth that is supported and available, it will be impossible to achieve the expected data rates. Simply put more Hz is required, and tapping into unlicensed and mmWave bands will most likely be necessary. Carrier aggregation is an important feature that will allow component carriers from different bands to be used simultaneously to obtain more aggregate bandwidth.

- *Energy efficiency.* Future networks will most certainly be green and energy efficient. Not just because it's fashionable and environmentally friendly, but also because, as networks become ultra dense, energy-efficiency will be key to their sustainability.

As smart devices become ubiquitous, an unspoken requirement will be the expectation for service anytime and anywhere. This implies that networks will be expected to offer full coverage, while ideally maintaining backward compatibility for older generation devices. In order to achieve all the above requirements, a major overhaul of existing networks is required. In terms of network architecture, densification and heterogeneity seem to be the defining features of future networks. In additional solutions such as massive MIMO, Coordinated Multi-point transmissions (CoMP) and Carrier Aggregation are being considered for improved spectral efficiency. From a physical layer perspective, adoption of mmWave bands will represent a major paradigm shift. In this thesis we tackle on some of the challenges that future networks will face, especially in terms of interference. We focus mostly on dense heterogeneous networks with carrier aggregation support, however whenever possible we also take into account other advanced features such as CoMP.

2.2 Dense heterogeneous networks

Densifying the cellular network by overlaying low-power nodes over the traditional macrocell architecture, has been considered for a while now, as one of the most effective and cost-efficient approaches to improving the cellular capacity and coverage [18]. Adding these low-power nodes creates a tiered heterogeneous network architecture which enables very high data rates for users under the coverage of these nodes, while ensuring overall coverage by the traditional macro cell. Transmitting at low power, these nodes usually have much smaller coverage areas, hence they are often referred to as small cells, or depending on the type of the node deployed, they are also called micro, pico or femto-cells. Coverage can also be improved by

deploying the small cells to cover coverage gaps or enhance coverage near the cell edges. Further heterogeneity in the network is introduced by integrating different radio access technologies (RATs) within the cellular network and allowing direct device-to-device communications (D2D). While today user devices can access WiFi networks and WiFi-offloading is already being used to alleviate the traffic load of the cellular network [19], it is expected that in future networks the integration of the different technologies will be seamless and transparent to the end user.

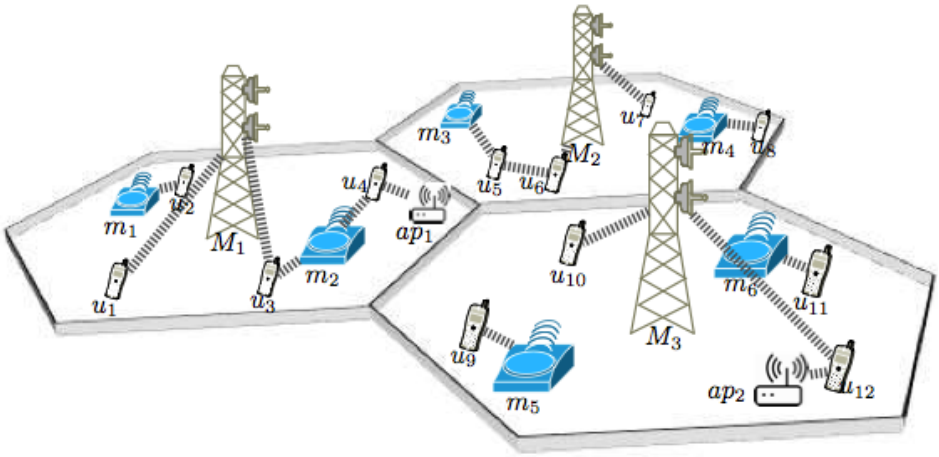


Figure 2.1. A dense heterogeneous network. Macrocells are denoted by $M_1, \cdots, M_6$, microcells by $m_1, \cdots, m_6$ and WiFi access points by ap_1, ap_2. Users are denoted by $u_1, \cdots, u_{12}$.

An example of a dense heterogeneous network is shown in Fig. 2.1. The traditional macrocell network structure is represented by the standard hexagonal-grid model while small cell base stations are deployed over the network area to complement the macro cell coverage and increase capacity. Dashed black lines represent supported types of communications between devices and network nodes. Users will be able to access the network through different types of PoAs: either through macro or small cell base stations, WiFi access points (like u_4 and u_{12} in the example) or via direct communication (D2D) between nearby devices (see u_5 and u_6). Simultaneous transmission between different base stations is another supported feature; in our example u_3 is receiving simultaneously from M_1 and m_2. For consistency reasons, throughout the text we will refer to the node providing access to the network as PoA (regardless of the type of node).

Interference is a major limitation in these types of networks, as radio resources are expected to be fully shared between the different tiers in the network. Even device-to-device communications are expected to take place within the cellular network bands, in what is often called "in-band underlay" mode [20], where device-to-device opportunistically accesses the same spectrum resources used by the other nodes

in the cellular network. In the downlink, user equipment (UE) connected to the small cells will experience sharp interference when macro transmissions are ongoing, especially those on the edges of the small cell. In addition, if UEs simply connect to the PoA with the downlink highest received power, most UEs will choose the macro PoA, whose transmit power is significantly higher than that of a small PoA. This typically leads to a situation in which only a small portion of UEs, usually in close proximity to the small PoA connect to it, thus underutilising the small cell potential and rendering its deployment useless. In the uplink too, the UEs are unable to take advantage of a nearby small cell, if they are simply connected to the PoA with the strongest received power in the downlink. It is clear therefore that interference must be properly addressed in order to take full advantage of the possibilities offered by dense multi-tier networks. Interference mitigation techniques have already been proposed for LTE and LTE-A. Frequency domain techniques such as Inter-Cell Interference Coordination (ICIC), sometimes referred to as *soft frequency reuse*, has been used in LTE to alleviate the interference between different macro cells [21]. But they are not sufficient enough to address the complexities of dense networks.

2.2.1 Enhanced Inter Cell Interference Coordination

A technique which focuses exclusively on the interference problems in dense heterogeneous networks, called enhanced ICIC (eICIC), was proposed for LTE-A. In broad terms, eICIC is a two-step technique. It first encourages macro UEs to offload to small cells by applying Cell Range Expansion (CRE). This technique incentivizes UEs to connect to small PoAs by applying biases during the user association step. A bias is introduced to the downlink received power for small cells to artificially expand the range of the small cell, hence forcing UEs that would normally connect to the macro PoA to connect to the small PoA instead. The UEs forced to connect to the small PoA are however exposed to severe interference during macro PoA transmissions. To alleviate this, the macro PoA is muted during some subframes, called Almost Blank Subframes (ABS), thus allowing the small PoA to transmit to its edge UEs interference-free. The biases for CRE and the percentage of ABS subframes are usually preconfigured in the network; how to properly adjust and configure them is somewhat of an open question. Algorithms for optimising these parameters have been proposed in literature [22]. Note that the independent application of either CRE or ABS is also possible. Throughout this work we usually use eICIC as a benchmark to compare the performance of the approaches we propose. As eICIC already improves the performance of dense networks, and has been included in the standardization efforts by 3GPP for LTE-A, it is a very good starting point to measure the performance of any algorithm that aims at further improving the spectrum utilisation efficiency of the network.

2.3 Device-to-device (D2D) communications

Support for device-to-device (D2D) links, that enable direct communication between two devices, bypassing thus the infrastructure nodes, is another key expected feature of future networks. Indeed D2D has already been introduced in LTE-A, 3GPP Release 12 [23], but the application is mainly limited to public safety services. However, as most of the research on D2D shows, the technology has potential to also improve spectrum utilization, overall throughput and energy consumption [20]. By exploiting the proximity between two devices, D2D communication may be significantly faster and less energy consuming as there is no need for the additional hop to the infrastructure node. In addition, spectrum may be better utilised as D2D links are usually short range, thus allowing for better spatial frequency reuse. As traffic is offloaded via D2D, alleviating congestion, other UEs not participating in D2D may benefit as well.

In general, it is envisioned that D2D communications will be undertaken under network-control. This simplifies many problematic tasks such as discovery, synchronisation and security, and also implies that traffic between UE pairs can be scheduled by the network as well. However, the management of D2D links poses a significant challenge, especially since D2D links are by nature ephemeral due to user mobility, forcing thus the network to promptly adapt to changes in the availability of such nodes. A major issue is the sharing of spectrum resources between the traditional cellular communications and D2D. Depending on the type of spectrum sharing, D2D can either be in-band or out-of-band. The out-of-band implies that D2D communications take place outside of the bands allocated for cellular communications (e.g., the ISM bands). In-band D2D, on the other hand means that D2D will be sharing the same spectrum with the traditional infrastructure-to-device communications.

The in-band D2D can be further classified into overlay and underlay spectrum sharing. In the overlay mode, D2D and cellular communications will be assigned orthogonal portions of the in-band spectrum. While this mode ensures interference-free communications, it is quite inefficient in terms of spectrum use. Conversely, in the underlay mode, D2D accesses the same spectrum resources as the cellular users in an opportunistic manner [20]. While the latter option allows for more flexibility in terms of spectrum use, interference caused by D2D links operating simultaneously in the same spectrum as cellular users, may be significantly degrading. Therefore tighter control over D2D links is necessary to manage interference. To this end, D2D links could be scheduled by the network together with the traditional cellular communications, that is, the scheduling of D2D links becomes part of the resource scheduling task. Due to its better performance in terms of achieved throughput as shown in [20], throughout the thesis we will consider the in-band underlay mode configuration for D2D.

In principle, the in-band underlay D2D links could potentially use either the

downlink or uplink radio resources to establish communication. In both scenarios, D2D can cause significant interference to normal infrastructure-to-device communications, either to nearby receiving UEs when implemented in the downlink bands, or to nearby receiving infrastructure node when deployed in the uplink bands. Without proper management of this interference, D2D communication may easily end up doing more harm than good. Currently, the use of uplink resources is widely favoured [23], mainly because, at present, the traffic load is asymmetric and significantly heavier in the downlink. However, it is expected that in the future traffic will be much less asymmetric, then the use of downlink resources will represent a viable option as we show in Chapter 4.

2.4 Carrier Aggregation

Carrier aggregation is an advanced technology, introduced in LTE-A, which allows the concurrent use of several LTE component carriers (CCs) in order to increase the amount of the aggregated bandwidth [24]. As shown in Fig. 2.2, the aggregated carriers are used as a single large data pipe which enables higher data rates and increases overall capacity. The different component carriers can be of varying bandwidths and, in general, may belong to different frequency bands. According to LTE-A specification, the width of the carrier can be any of the LTE supported bandwidths, i.e., 1.4, 3, 5, 10, 15 or 20 MHz. A maximum of 5 CCs can be aggregated to accrue up to 100 MHz of bandwidth.

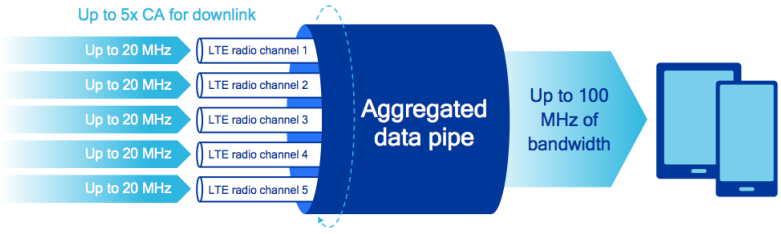


Figure 2.2. Carrier aggregation concept: several component carriers are used simultaneously to provide more bandwidth to the end user [25].

There are three possible deployment configurations: *intra-band contiguous, intra-band non-contiguous* and *inter-band non-contiguous* carrier aggregation. The three possible deployments are shown in Fig. 2.3. In the intra-band contiguous configuration, the aggregated CCs belong to the same frequency band and are contiguous. In the intra-band non-contiguous configuration, the aggregated CCs belong to the same frequency band but are not contiguous. And finally, in the inter-band non-contiguous configuration allows CCs to belong to completely different frequency bands and in general may have different bandwidths. While the first configuration

is most easily implemented, the latter two offer significantly more flexibility in terms of spectrum use, and more importantly do not require large contiguous chunks of frequency spectrum which are scarce. However the fact that CCs may belong to different frequency bands creates some additional complications due to the fact that the propagation conditions may vary greatly from one CC to another. This implies that the CCs may have varying coverage areas, and hence have different interference footprints. This affects especially the radio network planning phase and radio resource allocation algorithms, which need to take into account these differences.

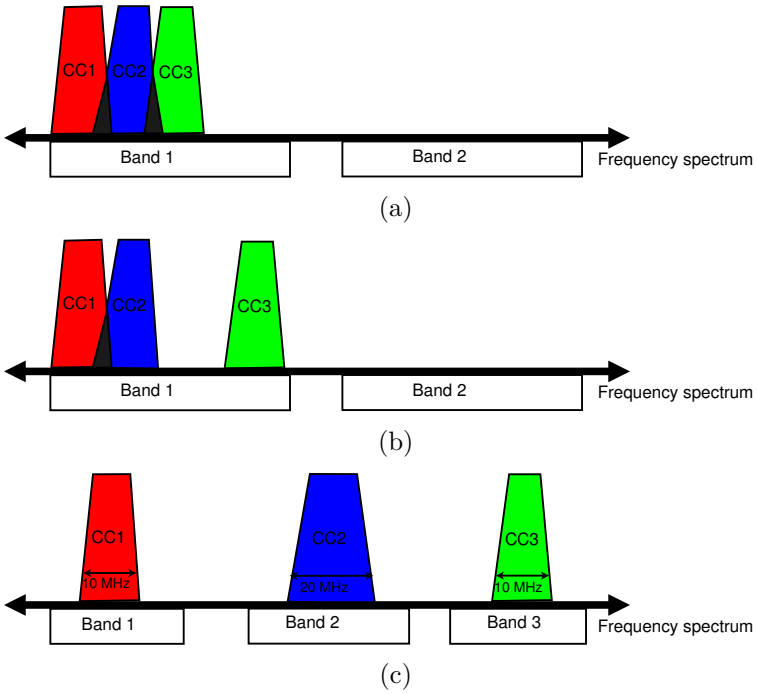


Figure 2.3. Possible deployments of carrier aggregation: a) Intra-band contiguous; b) Intra-band non-contiguous; and c) Inter-band non-contiguous.

An important consideration for carrier aggregation is backward compatibility with older legacy UEs. To achieve this, each individual CC, inherits the core physical layer design and can act as legacy LTE carrier. Indeed, each combination of PoA and CC can be considered as an individual serving cell, implying that legacy UEs can associate to any one of them based on their association rules, while CA-supporting UEs can select several carriers. For each CA-supporting UE, one CC is selected as a Primary Cell (PCell) which is used to send control information and perform radio link monitoring, with the possibility of selecting additional CCs,which are called Secondary Cells (SCell). While the PCell cannot be deactivated, the SCells can be dynamically activated and selected depending on the needs of the UE [26]. It should be noted also, that UEs under the coverage of the same PoA do not have to share

the same PCell, i.e., they can be assigned different CCs to act as PCells. This opens up a range of possibilities for UE load management and radio resource planning. Finally, two types of CA implementation can be envisioned [27]:

- *Single-flow* CA - The UE associates to a single PoA and can aggregate the CCs available at that particular PoA.
- *Multi-flow* CA - The UE can associate to different PoAs belonging to different tiers, on different CCs. This implementation, which is more advanced, enables the UEs to aggregate all the CCs available in the network, not limited to those available only at a particular PoA or tier.

2.5 User association in dense networks

In the traditional macro-based cellular networks, the user association, i.e., the cell selection process is quite straightforward. As the macros are usually deployed in a planned manner and have approximately similar coverage areas, the UEs will simply select the cell with the strongest received power (in the downlink). In LTE for example, UEs measure the received power for all the base stations and cells in the vicinity, and select the strongest one among them. Such measurements are known as the Reference Signal Received Power (RSRP) and the Reference Signal Received Quality (RSRQ). The UEs will perform cell selection according to the *S-criterion*, which, when fulfilled, implies that the RSRP and RSRQ are above a given minimum value. If it is fulfilled by more than one cell, the UE will select the cell with the strongest indicators [21].

However, as we briefly noted before, in dense heterogeneous deployments, such straightforward methods may lead to inefficient and underperforming load distributions. Indeed, the specific features of dense heterogenous networks must be accounted for when performing the user association task to ensure fair load distribution among available cells. Some expected improvements to the user association procedure are listed below:

- Cell range expansion (CRE)

  Cell range expansion, as we already mentioned, is part of the overall eICIC mechanism. The aim of CRE is to encourage offloading of UEs from macro cells to small cells by introducing a positive bias to the received power of small cells. Instead of associating to the cell with the strongest received power, the UEs associate with the cell with the strongest *biased* received power. If we denote by $P^u_{i,k}$ the received power (in dBm) from BS i, from tier k, at UE u, then the user association rule can be expressed as:

$$\arg\max_{i,k} P^u_{i,k} + J_k \tag{2.1}$$

where J_k is the biasing coefficient for tier k expressed in dB. Biasing can be also used effectively to balance the load among different available CCs, as we show in [8], especially when there is a significant number of legacy UEs in the network. In particular, biasing can be used to offload some of the UEs from low frequency carriers, that may benefit from better signal propagation conditions, and hence have stronger received power, compared to some of the high frequency CCs.

- Dual connectivity

  Dual Connectivity (DC) is viewed as another emerging solution for better user association in dense multi-tiered networks [28]. In DC mode, the user can associate simultaneously to a macrocell and a set of small cells. This way, the UE avoids performing frequent handovers, as it is already associated to the macro tier which covers a wider area, while the data traffic can be offloaded and served from the various small cells. This form of user association is implicitly assumed in carrier aggregation enabled networks with multi-flow implementation.

- Multiple association

  Proposed by Kamel et.al. [29], multiple association generalizes the idea of DC further to allow for the UE to be connected to the nearest M cells to form what the authors termed a Multicell. The traffic is distributed amongst them to overcome the backhaul and the maximum achievable rate of individual cells, thus achieving higher data rate. This form of user association scheme also fits within the framework of multi-flow CA, and may be of particular interest when advanced CoMP techniques such as Joint Transmission (JT) are also considered.

2.6 Radio resource management

Resource scheduling management in cellular networks is an important function that aims at efficiently allocating downlink and uplink channel resources to UEs, in order to meet as many of their expectations while optimizing network performance. In 4G cellular networks such as LTE-A, the underlying multiple access technique of choice is Orthogonal Frequency Division Multiple Access (OFDMA) based on the Orthogonal Frequency Division Multiplexing (OFDM) technique, which has found wide-spread application in wireless networks. OFDM is a multi-carrier modulation technique which divides the frequency-selective wideband channel into many non-frequency-selective narrowband subcarriers. These subcarriers are overlapping but orthogonal, as shown in Fig. 2.4. The orthogonality feature makes OFDM highly

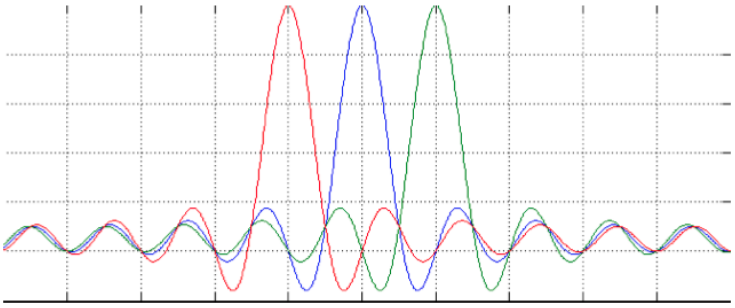


Figure 2.4. OFDM orthogonal subcarriers.

spectrally efficient as it is no longer necessary to separate the subcarriers with guard bands, as is the case with traditional multi-carrier techniques. The spacing between the subcarriers in OFDM allows for perfect separation at the receiver. In addition, the low complexity receiver implementation using FFT, makes OFDM attractive for high-rate mobile data transmission such as the LTE downlink [21]. OFDMA is an extension of OFDM applied for multi-user communication systems. OFDMA allows individual or contiguous groups of subcarriers to be assigned to different users, so several users can simultaneously receive data on the same channel. Thus, OFDMA can also take advantage of the multi-user diversity, as different users may experience different channel conditions on different subcarriers. As users report give feedback on their experienced channel conditions, this can be used to properly match subcarriers and users to enhance the total system spectral efficiency.

While non-orthogonal multiple access techniques are also being investigated, it is most likely that OFDMA will be adopted also for 5G [3]. As a consequence, it is anticipated that the radio resource structure will remain relatively the same as networks evolve to 5G. In general, throughout this thesis, we assume the radio resources to be organised as they are in LTE and LTE-A, which is described in the following subsection.

2.6.1 Radio resource structure

Radio resources in LTE are organised in units of Resource Elements (REs), which are further grouped into Resource Blocks (RBs). The resource units are defined both in time and frequency: the RE is composed of one subcarrier in the frequency domain for the duration of 1 OFDM symbol. The RB is a group of REs, composed of 12 subcarriers in frequency (or a total of 180 kHz with 15 kHz subcarrier spacing) and assigned for the duration of 1 time slot, i.e., 0.5 ms in time. Timewise, the resources are grouped into frames, which are 10 ms long and divided into 10 subframes, each 1 ms long. Each subframe is composed of two 0.5 ms slots, each equivalent to the

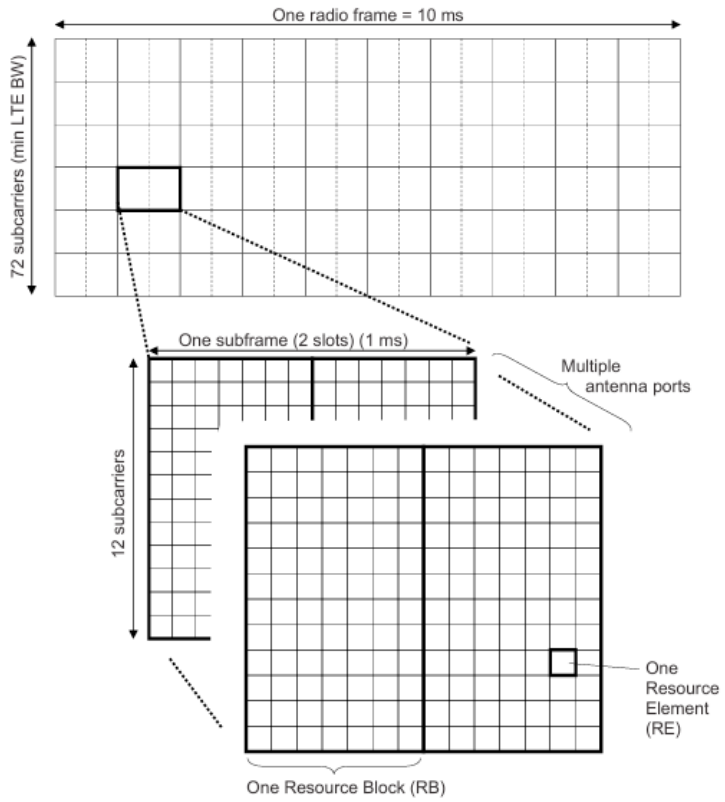


Figure 2.5. LTE resource structure [21].

duration of 7 OFDM symbols.[1] Accordingly, an RB is composed of 84 REs, and is the smallest unit of resources which can be assigned during scheduling. The resource structure described herein, is shown in Fig. 2.5.

2.6.2 Resource scheduling

Resource scheduling in LTE is performed every subframe. The task of this procedure is to designate at the BS[2] level which RBs should be allocated to which UEs, both in the downlink and uplink direction. Once the resource allocation is decided and UEs are informed of this decision on the appropriate control channel, the data destined for each UE can be transmitted on the scheduled RBs.

Resource scheduling algorithms are the procedures used by the network to reach

[1]Each slot comprises 7 OFDM symbols with normal Cyclic Prefix (CP), or 6 OFDM symbols with extended CP.

[2]Base stations are referred to as eNodeBs in LTE.

resource allocation decisions. Generally speaking, such algorithms aim at matching RBs and UEs in a way that optimises specific network performance metrics such as: average per-user or total throughput, average/maximum/minimum delay or per-user or total spectral efficiency. To reach informed allocation decisions they can make use of two types of information: channel state information (CSI) and traffic measurements such as volume and priority. This information can be collected either by PoA measurements or obtained through UE reports and feedback (such as RSRP and RSRQ). The scheduler must ensure that all UEs under the coverage of a cell are served and strive to fulfil as much as possible all UE requirements in terms of latency and QoS. The scheduling algorithms are usually internal to the PoA and are generally not standardised, leaving ample room for proprietary and vendor-specific algorithms to be developed, tailored to optimise specific network performance metrics. In dense network deployments special attention needs to be paid to inter-cell and inter-tier interference, which can impose significant limitations on the network performance. ‘ That being said, there are a few scheduling algorithms which have found wide-spread application, which we will briefly describe here.

- Maximum rate scheduling

  This scheduling algorithm aims at maximising the overall throughput in the system, and its performance relies heavily on the multi-user diversity in the network. In other words, as the different users are expected to experience different channel conditions on the different frequencies due to multipath fading and UE mobility, this type of scheduler allocates the RB at each given time instant to the UE experiencing the best channel conditions on the particular RB. This is a variation of the water-filling algorithm adapted for scenarios without power adaptation. Therefore, this kind of scheduler invariably leads to situations in which UEs with good channel conditions are scheduled disproportionally more often than those experiencing bad channel conditions, such as cell edge users. So while this type of scheduling can indeed deliver optimal performance in terms of overall data rate, it has no mechanisms to ensure any kind of fairness, or even basic coverage for all the UEs under the coverage of the network.

- Proportional fair scheduling

  It is clear therefore that in dense multi-tiered networks fairness among UEs needs to be imposed by the network, through the scheduling algorithm. The Proportional Fair (PF) scheduling is a widely popular algorithm, precisely because it aims at ensuring fairness in terms of service offered to the UEs in the network. Specifically, the PF scheduler will allocate UE u on RB r on a specific time instant k, if the instantaneous channel quality of the user i

relatively high, with respect to its own average channel conditions over time. This can be expressed as [21]:

$$\hat{u}_r = \arg\max_{u=1,\ldots,U} \frac{\delta_u(r,k)}{\bar{\delta_u}(k)} \tag{2.2}$$

where $\delta_u(r,k)$ is the rate obtained by user u on RB r in time instant k, and $\bar{\delta_u}(k)$ is the long-term average throughput of user u, calculated at time instant k. The long-term average is calculated recursively using:

$$\bar{\delta_u}(k) = \left(1 - \frac{1}{t_c}\right)\bar{\delta_u}(k-1) + \frac{1}{t_c}\sum_{r=1}^{R}\delta_u(r,k)\mathcal{I}\{\hat{u}_r = u\} \tag{2.3}$$

where t_c is the time window over which fairness is imposed, and $\mathcal{I}\{\hat{u}_r = u\}$ is the indicator function, equal to 1 when $\hat{u}_r = u$ and 0 otherwise. A very large averaging window will tend to maximise the total average throughput and the PF algorithm converges to the maximum rate algorithm as $t_c \to \infty$. When the t_c is very short, on the other hand, PF tends to behave like a round-robin scheduler, i.e., scheduling the same number of RBs for each UE, irrespective or channel conditions. In CA-enabled networks, cross-CC PF has been proposed [30], which takes into account the throughput obtained over all carriers when calculating the long-term average. This ensures some fairness between CA-enabled and legacy UEs.

In the context of carrier aggregation, there are as we said, further complications to the resource scheduling task. Indeed, in addition to the RB allocation there is an additional step which needs to be performed which is the carrier selection. While this step could be performed independently from the resource scheduling task, the load over the different carriers as well as the coverage properties of each carrier significantly impact the performance of the resource allocation strategy chosen. Indeed in this thesis we show that joint carrier selection and resource allocation in CA-enabled dense networks, greatly improves networks performance when comparing to standard implantation using CRE during user association, biasing for carrier selection and PF as the underlying resource scheduling algorithm [8]. It should be noted that it is expected that future networks will also support *cross-carrier scheduling* which allows BSs to transmit the scheduling information for all CCs on a single CC (designated as the PCell).

Chapter 3

Related work

3.1 Interference management and resource allocation in dense multi-tier networks

The deployment of a multi-tier network where cells use the same radio resources is highly beneficial since it allows traffic offloading from macrocells to smaller cells [31, 32]. However, such networks are particularly vulnerable to severe interference. As a result, interference mitigation and management in dense networks has attracted a lot of attention as a research topic. A comprehensive survey which provides a summary of interference control techniques in emerging dense multi-tier networks is provided in [33].

A good survey on inter-cell interference coordination techniques (ICIC) can be found in [34]. ICIC has been proposed mainly to tackle the inter-cell interference, however it is not adequate enough to address the complexity of multi-tier networks. In addition, enhanced ICIC specifications in 3GPP Rel. 10 [35] foresee the use of the Cell Range Expansion (CRE). Such a technique involves adding a positive range expansion bias to the pilot downlink signal strength received from microcells so that more users connect to them. Then, in order to mitigate the interference between overlaying macro- and microcells, macrocells may periodically mute their downlink transmissions in certain subframes (called almost blank subframes - ABSs). By using ABSs for edge users, microcells can significantly improve their performance [22].

3.1.1 Dense multi-tier networks with D2D support

The integration of D2D communication in cellular networks and its applications are investigated in [36]. This work presents a conceptual framework for the formulation of problems such as peer discovery, scheduling, and resource allocation.

Another good survey on D2D proximity services as foreseen by 3GPP can be

found in [37]. This work provides a comprehensive overview on some of the key functionalities of D2D communication and design challenges concerning the integration of such communication within cellular networks. The authors of [37] focus on two main use cases of D2D, as envisioned by 3GPP, namely for public safety and traffic offloading. In the latter use case, which is the one we also focus on in our work, it is expected that D2D will be established with support from infrastructure nodes, simplifying tasks such as synchronization and discovery, and enabling dynamic resource allocation for D2D links [38].

The problem of resource allocation is also studied in [39–43], although only macrocells and D2D mode are considered therein. Additionally, in [39] the D2D pairs wishing to exchange data are given at the outset (i.e., unlike our work, [39] does not address the endpoint association problem). In [40] the authors seek an optimal resource allocation scheme for cellular networks with D2D support, comparing orthogonal and non-orthogonal resource sharing modes. As numerical results show, D2D coupled with non-orthogonal resource sharing mode ensures the highest gains in terms of sum rate and per-cell throughput. Authors of [39, 41] formulate resource allocation as a mixed integer optimization problem, which is notoriously hard to solve, with [39] also presenting a greedy heuristic.

The work in [44] further compounds the problem by investigating the selection of the most suitable communication mode, still in a single-tier scenario with D2D. There, an analytical model is proposed, based on the assumption that the positions of BSs and users can be modeled as a Poisson point process.

Beside the different methodology and scope of the studies above, we stress that our work addresses dense networks including macrocells, microcells and D2D. While [44] derives an optimal factor of spectrum partition between cellular and D2D communication, we aim at determining the endpoint that should serve each user and an efficient data scheduling on a single radio resource basis.

3.1.2 CA-enabled dense networks

LTE multi-tier networks with carrier aggregation capabilities have been investigated in works such as [5, 26, 27, 30, 33, 45].

In [5] and, particularly, in [45], the authors propose an autonomous carrier selection algorithm which ultimately serves as an interference coordination technique between low-power and macro cells. However, neither of these two works addresses the issue of resource allocation once the carrier selection is performed. In [27], the authors propose a load-aware model for LTE multi-tier networks with carrier aggregation support. The work focuses on biasing, which is a simple load balancing approach that allows small cells to increase the strength of their pilot signal so as to expand their coverage area. In this scenario, the study analyzes the impact of biasing in carrier aggregation-enabled networks with different band deployment

configurations. Again, the resource allocation problem is not particularly addressed.

Similarly in [46], the authors present a comparison between several carrier deployment configurations for the macrocells and microcells, and then analyze the benefits of applying cooperation techniques between cells for each configuration. The authors address the extreme configuration cases where dedicated carriers are assigned to macrocells and microcells, and where all carriers are available at all cells (the co-channel configuration), and two other hybrid configurations. They also consider two cell cooperation techniques, the eICIC and the "inter/intra site CA" which allows users to connect to two different basestations on distinct carriers (also referred to as *multi-flow* CA in this text).

Resource allocation in single-tier networks with carrier aggregation support has been studied in [47–49]. In [47], the authors propose a two-step procedure where load balancing among the different carriers is performed before resources are allocated according to a proportional-fair based scheduler. Two approaches are proposed for load balancing among legacy users, a round-robin scheme which allocates new users to the carrier with the lowest load, and a mobile hashing scheme, which assigns new users randomly over the carriers, which aims at ensuring balanced load across the carriers in the long term. The CA-enabled users are automatically assigned on all available CCs. Furthermore, two versions of the proportional fair scheduling algorithm are proposed: the independent scheduling scheme, where users on each CC are scheduled independently from other CCs, and the cross-CC scheduling, where scheduling is performed taking into consideration scheduling in other CCs. The latter version aims at enhancing fairness for users that do not support carrier aggregation.

An interesting take on the problem is provided in [48], where the authors address resource allocation in a scenario where users are assigned only a subset of the available carriers for energy saving purposes. The innovative aspect of the work is focused around the optimal selection of the secondary cell set while the scheduling is performed based on a proportional fair metric.

A greedy algorithm to tackle the resource allocation problem in cellular networks with carrier aggregation support is presented in [49], which aims at maximizing system throughput and ensure fairness among CA-enabled and legacy terminals. However, none of the above works tackle the interference aspects, as the authors consider only a single-tier [47] or a single-cell [48, 49] LTE network.

An interference and traffic aware algorithm for resource allocation in LTE heterogeneous networks is presented in [9]. The network under consideration, however, does not support carrier aggregation, instead it allows device-to-device communications, which is another advanced LTE feature. The algorithm presented takes a dynamic programming approach and applies approximate dynamic programming principles to devise a resource allocation strategy which best matches basestations/

serving UEs with receiving UEs who are downloading content, while keeping interference at a minimum and taking into account traffic characteristics. The algorithm presented performs almost optimally which makes it a good candidate to use as a comparison reference for heuristic algorithms.

In this thesis, we address the resource allocation problem in dense multi-tier networks that support carrier aggregation with backward compatibility for legacy user terminals. Unlike previous work, we look at the resource allocation problem by tackling the two main problems afflicting multi-tier networks: inter-cell/inter-tier interference, and the complexity imposed by the availability of multiple carriers with potentially very different propagation and coverage characteristics. Furthermore, we propose a solution that jointly addresses carrier selection and resource allocation, while taking into account interference, in order to fairly serve CA-enabled and legacy user terminals.

3.2 Downlink carrier transmit power setting

While many papers have appeared in the literature on uplink power control, fewer exist on downlink power setting in traditional cellular systems. Among these, [50] and [51] are of particular interest. In particular, [50] shows that uplink power control can be modelled as a game of weighted strategic substitutes and complements, which is a type of game belonging to the class of pseudo-potential games [52, 53]. Instead, [51] proposes a distributed uplink power control scheme that leverages the use of a sigmoid utility function (first convex and then concave). The use of such function, which we will exploit too in our study, is a natural choice considering that a number of important performance measures (e.g., cell throughput) have a sigmoid-like shape as a function of signal to interference ratio.

In heterogeneous networks, game theory has been adopted to address downlink and uplink resource allocation in [54] where the authors investigate the economic incentive for a cellular operator to add femtocell service on top of its existing macrocell service, and model the interactions between a cellular operator and users as a two-stage Stackelberg game. The goal is to address both the user association problem as well as the orthogonal division of resources between macrocell and femtocell users.

The works [55] and [56] uses coalitional games to investigate power and resource allocation in heterogeneous networks where cooperation between players is allowed. Downlink power allocation in cellular networks with underlaid femtocells is modeled in [57] as a Stackelberg game, with macro and femto base stations competing to maximise their individual capacities under power constraints. Resource allocation in heterogeneous networks is also addressed in [58] where the authors propose two possible solutions, a heuristic approach using simulated annealing and geometric

optimization, while taking into account both the geometry of the network and load fluctuations. Interference in densely deployed femtocell networks is addressed in [60] through proper power adjustment and user scheduling. The authors propose a heuristic distributed algorithm that adjusts the coverage radius of the femtocells and then schedules the users in a fair manner. However, the algorithm applies only to femtocells, thereby missing out on many possible solutions offering both better energy efficiency and network throughput. A backhaul-aware approach is taken by the authors in [61] where they propose an optimal user association scheme to mitigate interference, which takes into account the base station load, the backhaul load as well as backhaul topology.

An energy efficient approach is instead proposed in [62]. There, BSs do not select transmit power levels as we do in our work, rather they can only choose between on and off states. The authors propose a Nash bargaining solution to solve the problem. Maximising energy efficiency is also the goal of [63], which however is limited to the study of resource allocation and downlink transmit power in a two-tier LTE single cell. In [64] in order to improve the energy efficiency of ultra-dense networks, the authors frame the problem of joint power control and user scheduling as a mean-field game and solve it using the drift plus penalty (DPP) approach in the framework of Lyapunov optimization. Mean-field games are also used in [65] where the interference problem (both inter-tier and inter-cell interference) is formulated as a two-nested problem: an overlay problem at the macrocell level and an underlay problem at the small-cel level. In the overlay problem, the macrocell selects the optimal action first, to provide minimum service, while the underlay problem is then formulated as a non-cooperative game among the small cells. The mean-field theory is exploited to help decouple a complex large-scale optimization problem into a family of localized optimization problems.

Multi-cell network with inter-cell interference is considered in [66], where energy efficiency is optimised by applying resource allocation and discrete transmit power levels. The authors propose a suboptimal distributed two step algorithm, to solve the optimisation problem they formulate.

We remark that the above papers address heterogeneous dense networks but, unlike our work, they do not consider CA support, which will be a fundamental feature of future cellular networks and significantly changes the problem settings. Also, [56, 57, 63] formulate a resource allocation problem that aims at distributing the transmit power among the available resources under overall power constraints. In our work, instead, we foresee the option of varying the overall transmit power at carrier level, which is constrained by an upper bound, but which is not necessarily treated as budgeted resource to be allocated among the different carriers. In short, we do not formulate the problem as a downlink power allocation problem, rather as a power setting problem at carrier level, assuming *each carrier has an independent power budget*. Additionally, while most of the previous work [58,60,62,63,66] focus on

the heterogeneous network interference problem only, using game theory concepts we jointly address interference mitigation, power consumption and user coverage by taking advantage of the diversity and flexibility provided by the availability of multiple component carriers. Finally, we propose a solution that enables the PoAs to dynamically change their power strategies based on user distribution, propagation conditions and traffic patterns.

To our knowledge, the only existing work that investigates downlink power setting in cellular networks with CA support is [67]. There, Yu et al. formulate an optimisation problem that aims at maximising the system energy efficiency by optimising power allocation and user association. However, interference issues, which are one of the main challenges we address, are largely ignored in [67] as the authors consider a non-heterogeneous single cell scenario.

Part II

Interference-aware resource allocation in dense networks

Chapter 4

Interference-aware resource scheduling in D2D-enabled networks

In this chapter we look closely at the resource allocation problem in a two-tier dense network where in-band D2D communications are supported under network control. As all PoAs, including devices involved in D2D communications use the same set of radio resources, the may heavily interfere with each-other. We devise a dynamic programming approach to efficiently schedule download and upload traffic, by (i) efficiently matching communicating endpoints and (ii) assigning radio resources in an interference-aware manner while accounting for the characteristics of the traffic content to be delivered.

We consider that D2D will take place within the LTE bands, in the “in-band underlay" mode [44], where D2D opportunistically accesses the same spectrum resources used by the other nodes in the cellular network.Currently, it is widely accepted that uplink resources should be used [68], since, at present, traffic is significantly heavier in downlink than in uplink. However, it is expected that in the future traffic will be much less asymmetric, then the use of downlink resources will represent a viable option.

We therefore address and compare both D2D scenarios, and propose a resource allocation procedure based on approximate dynamic-programming. The procedure itself is adaptable to both downlink and uplink D2D scenarios, it is updated every subframe and is efficient enough to be applied to large-scale scenarios. The performance of our approach is numerically evaluated and compared to standard resource scheduling algorithms adopted in today’s cellular networks, employing interference mitigation techniques. Results highlight that the proposed approach is apt at fully exploiting the potential of both the heterogeneity of the network and D2D support, while consuming far less energy. Results further reveal that D2D interactions act

inherently as an additional layer in the heterogeneous network, thereby potentially reducing the need for deploying more microcells. Finally, while the uplink and downlink scenarios provide similar performance in current traffic load conditions, the downlink will become preferable as the upload and the download traffic tend to even out.

4.1 Network model and scenario assumptions

In this subsection we detail the system scenario and any assumptions we make about the D2D-enabled dense network under study. As was previously mentioned, the network model that we use is based on a two-tier architecture. The network is made up of two-tiers of infrastructure-based PoAs: a set of first tier macro base stations (BS) that control the macrocells and a set of micro BSs deployed within their coverage area that control the second tier microcells. We define the coverage of an infrastructure PoA (either macro or micro), as the area where the received strength of the pilot signal is higher than a given threshold [69]. We denote the collective set of infrastructure PoAs by $\mathcal{B}$. A UE under the coverage of both a macrocell and a microcell can be served by either of them. In addition, we consider a third tier of device-based PoAs, which are in fact the UEs participating in D2D communications and serving content to other UEs in their vicinity, under network control. The set of UEs is denoted by $\mathcal{U}$.

In general, we assume that our dense network supports Frequency Division Duplexing (FDD), with uplink and downlink using two different portions of the spectrum, however our formulation can be easily extended to the Time Division Duplexing (TDD) case. Radio resource allocation is performed and updated every subframe by the Area Controller (AC) in the core network, which assists PoAs in radio resource allocation and traffic scheduling. The AC collects information on the channel quality from the PoAs and on content items that users wish to upload/download. Thus, PoAs are only concerned with propagation and spectrum aspects, while they are oblivious to higher-layer demands.

From the collected information, the AC determines:

(1) which PoA (among the possible ones: macro, micro, or UE) should serve each user,

(2) which RB(s) to employ for such communication.

Decisions taken by the AC are issued to the PoAs, which forward them to UEs. The fact that the AC performs the resource allocation task in a centralised manner is in contrast with the distributed schemes that are adopted in today's cellular networks, where eNodes B are in charge of resource scheduling. However, given the

expected complexity of future cellular networks, it is possible that a more centralised structure will be adopted [70]. An example scenario of such a network is shown in Fig. 4.4, while Fig. 4.2 shows the roles of ACs, PoA, and users.

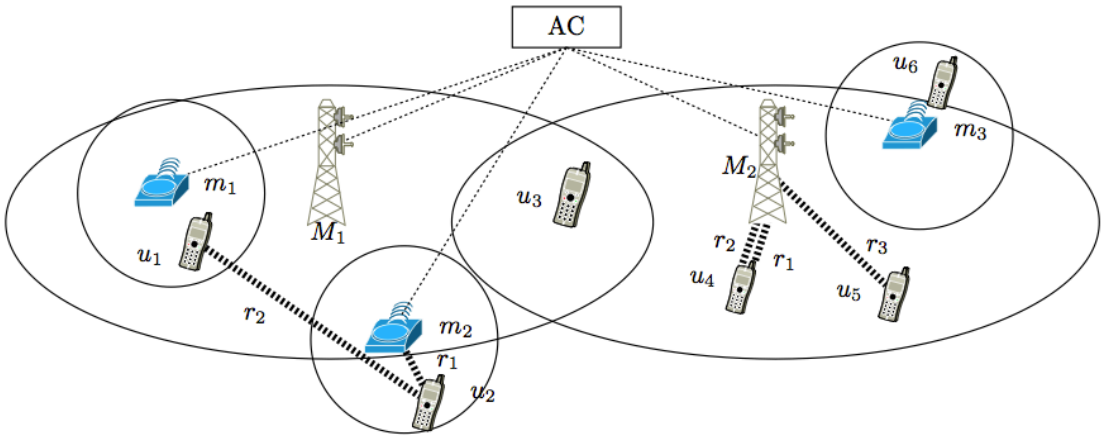


Figure 4.1. An example scenario. UEs are denoted by $u_1, \ldots, u_6$, macro PoAs by M_1, M_2 and micro PoAs by m_1, m_2, m_3. Solid lines denote coverage areas. Dotted lines correspond to RBs used by a pair of endpoints.

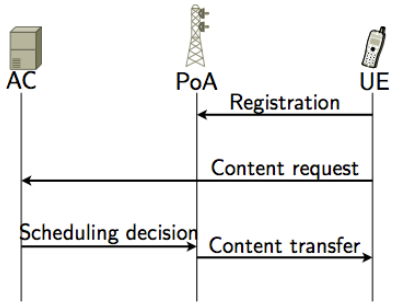


Figure 4.2. The role of the area controller in our system model, in the case of content download. Users register with infrastructure PoAs, as in current LTE networks. Content requests, however, are directed to the AC, which makes content-aware scheduling decisions. Such decisions are pushed to PoAs, which are then in charge of enacting them.

We assume that macro and micro PoAs have optical fiber connectivity to the core network, as envisioned by operators and network manufacturers [71, 72]. This assumption is reasonable given the new, complex tasks and the ever-increasing amount of traffic that the cellular infrastructure is expected to handle.

As envisioned by recent trends and standardization activities, we consider network-controlled D2D communication [38, 40, 73]. This implies that, not only can synchronization and security issues be easily solved, but also UE pairs can be efficiently

scheduled so as to use cellular resources even at high traffic load.

We assume in-band, underlay, deployment of D2D communication, that is, spectrum resources are shared between devices using the D2D communication paradigm and the cellular infrastructure. However, we consider two possible scenarios: one in which D2D interactions leverage the downlink radio resources (downlink scenario), and the other in which D2D takes place in the uplink spectrum (uplink scenario). To this end, and in order to ensure efficient allocation of all radio resources, in either scenario, we consider both download and upload traffic scheduling.

We mainly focus on unicast data transfers and assume that UEs can be served download traffic by only one PoA at a time, and similarly they can transmit upload data to only one PoA at a time. Considering the most popular types of terminals, we also assume UEs to be half-duplex, i.e., they cannot transmit and receive at the same time. This implies that, e.g., in the downlink scenario a UE receiving information from the cellular infrastructure cannot simultaneously serve another UE.

4.1.1 Interactions, traffic flows and resource blocks

Within the network the following interactions are supported: interactions between infrastructure PoAs and UEs and vice-versa, and interactions between two UEs or D2D communications.

PoAs and UEs interact with each other using assigned radio resources. RBs allocated to each interaction can be selected from two distinct sets of RBs: the set of downlink radio resources $\mathcal{R}_d$, and the set of uplink resources $\mathcal{R}_u$. The RB allocation is valid for one subframe, therefore the time is divided into a set $\mathcal{K}$ of 1 ms time steps.

The goal of the communications is to transfer the content requested by the UEs. Each data request made by an UE will initiate a traffic flow f. We denote the total set of traffic flows as $\mathcal{F}$. For each traffic flow f, we record the time step at which flow f is initiated as $e(f) \in \mathcal{K}$. Also, we define a parameter $l(f)$ indicating the amount of data requested in bits, and a completion deadline $d(f)$, which denotes the time-window available for the traffic flow to be completed. For consistency reasons, we denote the UE requesting traffic flow f by $u(f)$[1]. Both traffic flow directions are considered: upload or download. When there is a download request, one of the infrastructure PoAs, or, when D2D is enabled, another UE caching the requested content item, can serve the traffic flow. Similarly, a UE may wish to upload a specific content item to the Internet, through the cellular infrastructure (i.e., macro or micro base stations). We denote by f_d traffic flows in the download direction, and by f_u

[1]Note that by requesting UE we denote the UE requesting a traffic download or upload, therefore $u(f)$ may be either a transmitting endpoint, or a receiving one.

those in the upload direction.

The allocation of RBs depends on the type of the communicating endpoints since the different RB sets may be used only for specific types of interactions. Namely, infrastructure-to-device (I2D) interactions may be allocated RBs only from the $\mathcal{R}_d$ set, while device-to-infrastrucutre (D2I) interactions may be allocated RBs only from the $\mathcal{R}_u$ set. For D2D communications, the RB set to be used shall depend on whether D2D is established in the uplink or downlink band.

Indeed, since we assume that D2D links are used for download traffic offloading, i.e., they may serve to provide UEs with requested content items if available in the vicinity, the choice of the RB set is constrained by the traffic direction.

During each interaction a certain amount of data is transferred between the two communicating endpoints. We define the variable $X^k(s, f)$ to indicate the amount of traffic flow f downloaded from the serving PoA s[2] at time step k. Then, for each initiated traffic flow $f \in \mathcal{F}$ a user is interested in, we define a variable $h^k(f)$, which denotes the total amount of traffic flow f downloaded/uploaded by the requesting user $u(f)$ until time step k. Note that $h^k(f)$ is a non-decreasing quantity bounded by the size of the content item, i.e., $0 \leq h^k(f) \leq h^{k+1}(f) \leq l(f), \forall k \geq 0, \forall f \in \mathcal{F}$.

The relationship between the two quantities is the following:

$$\begin{aligned} h^{k+1}(f) &\leftarrow h^k(f) + \sum_{s \in \mathcal{B} \cup \mathcal{U}} X^k(s, f)), \forall d \in \mathcal{F}_d \\ h^{k+1}(f) &\leftarrow h^k(f) + \sum_{s \in \mathcal{B}} X^k(s, f), \forall f \in \mathcal{F}_u \,. \end{aligned} \tag{4.1}$$

The serving PoA(s) must use allocated RBs to accommodate the traffic flow and deliver the requested amount of data within the delivery deadline. Otherwise, the UE will lose interest in the specific flow.

4.1.2 Propagation and interference model

For each communication, we denote the power with which PoA s transmits to UE u by $P(s, u)$. The power transmitted by infrastructure PoAs does not depend on the position of the receiving UE, but rather on the type of PoA (macro or micro) and is considered to be fixed. The transmit power is equally distributed over all RBs. The attenuation experienced by the signal, however, depends on both the user position and the central frequency of the cell carrier. The transmit power of UEs on the other hand is power controlled, regardless of whether they are participating in a D2D link or uploading to the network. The total attenuation experienced by

[2]Note that since we also consider D2D communications as well as upload traffic, a serving PoA can be either an infrastructure-based PoA or another UE.

the signal transmitted from serving PoA s to a UE u is denoted as $A(s,u)$. To pre-calculate these values, we used the urban propagation models specified in [69], which are described in this section.

Infrastructure to device (I2D) links

Macro PoA - UE links. For macro PoA-UE links, the attenuation is given by the following expression:

$$A(s,u)|_{dB} = G_T + AP(\theta_{s,u}) - PL(s,u)|_{dB} \qquad s \in \mathcal{B}_M, u \in \mathcal{U} \tag{4.2}$$

where G_T is the antenna gain of macro BS b controlling cell s, and $AP(\theta_{s,u})$ is the antenna pattern factor, which depends on the angle $\theta(s,u)$ between s's antenna maximum direction and the direction between the antenna and UE u . $PL(s,u)$ is the path loss experienced between the serving and the receiving endpoint. As mentioned before, the path loss in general will depends on the central frequency of the carrier c_f used by the cell, the distance between the two endpoints, on the Line of Sight (LoS) conditions between them as well as on the transmitting and receiving antenna heights. For macroPoA-UE links, it is calculated according to the Urban Macro (UMa) propagation model in [69].

Specifically in LoS conditions, it is given by the following expression:

$$PL(s,u)|_{dB} = 22 \cdot \log_{10}(d_{s,u}) + 28 + 22 \cdot \log_{10}(c_f) \qquad 10m < d_{s,u} < d'_{BP} \tag{4.3}$$

$$PL(s,u)|_{dB} = 40 \cdot \log_{10}(d_{s,u}) + 7.8 - 18 \cdot \log_{10}(h'_{bs}) - 18 \cdot \log_{10}(h'_{ms}) + 2 \cdot \log_{10}(c_f) \qquad d'_{BP} < d_{s,u} < 5\ km \tag{4.4}$$

In NLoS conditions:

$$\begin{aligned} PL(s,u)|_{dB} =& 161.04 - 7.1 \cdot \log_{10}(W) + 7.5 \cdot \log_{10}(h) - \\ & - (24.37 - 3.7(h/h_{bs})^2) \cdot \log_{10}(h_{bs}) \\ & + (43.42 - 3.1 \cdot \log_{10}(h_{bs}))(\log_{10}(d_{s,u}) - 3) + \\ & 20 \cdot \log_{10}(c_f) - (3.2(\log_{10}(11.75 \cdot h_{ms}))^2 - 4.97) \end{aligned} \tag{4.5}$$

where $d_{s,u}$ is the distance between the serving PoA s and UE u, h_{bs} and h_{ms} are the respective antenna heights of the macro PoA and the receiving UE and, d_{BP} is the breakpoint distance and c_f is the central frequency of the carrier of PoA s.

Micro PoA-UE links. For micro PoA-UE links, assuming omnidirectional antennas with 0 dBi gain, the attenuation between two endpoints depends solely on the path loss value, i.e., $A(s,u) = PL(s,u), \forall s \in \mathcal{B}_m$, which is precomputed using the Urban Micro (UMi) propagation model in [69]. Specifically in LoS conditions, it is the same as the UMa model given by Eq. 4.4-4.5.

In NLoS conditions it is:

$$PL(s,u)|_{dB} = 36.7 \log_{10} d_{s,u} + 22.7 + 26 \log_{10}(c_f) \tag{4.6}$$

Device to device (D2D) links

In general, we assume that the transmit power of an UE is subject to a power control scheme, so that its value may depend on such factors as propagation conditions and positions of either endpoints. Specifically, we use the following power control formula, from [74], to calculate the power of a transmitting UE u towards endpoint e, which can be another UE or an infrastructure PoA, at each time step k:

$$\begin{aligned} P(u,e)|_{dB} = &\min\{P_{max}|_{dB}, 10 \log_{10}(M) + P_0|_{dB} \\ &+ \rho \cdot PL(u,e)|_{dB} + \Delta_{TF} + f(k)\}. \end{aligned}$$

P_{max} indicates the maximum power at which a UE can transmit, while M indicates the number of RBs allocated to u at time step k. P_0, ρ, Δ_{TF} and $f(k)$ are cell and user specific configuration parameters, indicating respectively the spectral power density required at the receiver, the path-loss compensation factor, a UE-specific parameter depending on the applied Modulation Coding Scheme (MCS) and a higher-layer closed-loop command to increase/decrease power level.

$PL(u,e)$ indicates the path loss experienced between the two endpoints. For D2D links, path loss is modeled using the UMi propagation model specified in [75] and a correction factor that compensates for the low antenna height of the transmitter. Specifically, in LoS conditions it is given by the expression:

$$\begin{aligned} PL(u,e)|_{dB} = 22.7 \cdot \log_{10}(d_{u,e}) + 27 + 22 \cdot \log_{10}(c_f) + correction\_factor \\ 10m < d_{u,e} < d'_{BP} \end{aligned} \tag{4.7}$$

$$\begin{aligned} PL(u,e)|_{dB} = &\, 40 \cdot \log_{10}(d_{u,e}) + 7.56 - 17.3 \cdot \log_{10}(h'_{bs}) \\ &- 17.3 \cdot \log_{10}(h'_{ms}) + 2.7 \cdot \log_{10}(c_f) + correction\_factor \\ &d'_{BP} < d_{u,e} < 5\ km \end{aligned} \tag{4.8}$$

While, in NLoS conditions it is given by the expression:

$$PL(u,e)|_{dB} = (44.9 - 6.55\log_{10}(h_{ms}))\log_{10}(d_{u,e}) + 5.83\log_{10}(h_{bs}) + 14.78 + 34.97\log_{10}(c_f) + correction\_factor \tag{4.9}$$

The LoS probability as a function of distance between two endpoints is given by:

$$\mathbf{Pr}_{LoS} = \min\left(\frac{18}{d},1\right)\cdot\left(1 - e^{-\frac{d}{36}}\right) + e^{-\frac{d}{36}} \tag{4.10}$$

In all cases, from the viewpoint of our model, power and attenuation are input values. Thus, any assumption about propagation conditions and power control algorithms can be accommodated with no change to the model itself.

Given the transmit power and the attenuation factor, the useful power received at endpoint e_2 from serving endpoint e_1 is $P(e_1,e_2)/A(e_1,e_2)$. Every node pair (e,u) communicating on the same RB where e_2 is receiving, causes a certain amount of interference to e_2 given by $P(e,u)/A(e,e_2)$. Thus, the total amount of interference experienced by e_2 on RB r is:

$$I_r^k(e_2) = \sum_{\substack{(e,u)\text{ use } r \text{ at } k \wedge \\ e:\, A(e,e_2)>0}} P(e,u)/A(e,e_2)\,,$$

while the signal to noise plus interference ratio (SINR) is yielded by

$$\gamma_r^k(e_1,e_2) = \frac{P(e_1,e_2)/A(e_1,e_2)}{N + I_r^k(e_2)}. \tag{4.11}$$

where N denotes the noise power. We can finally map the SINR onto the amount of data that can be transferred from e_1 to e_2 using RB r during step k. We indicate this amount by $\delta_r^k(e_1,e_2)$, and we determine its value based on experimental measurements [76].

Clearly, the value $\delta_r^k(e_1,e_2)$ places a strict limitation on the amount of data that can be transferred for a certain traffic flow f between the two endpoints, which we defined earlier as $\chi_r^k(s,f)$, where s indicates the serving PoA. In this context, if $f \in \mathcal{F}_d$, e_2 is the requesting UE, i.e., $e_2 \leftarrow u(f)$ and e_1 is the serving PoA, i.e., $e_1 \leftarrow s$. Otherwise, if $f \in \mathcal{F}_u$ then $e_1 \leftarrow u(f)$ and $s \leftarrow e_2$.

The relationship between the two can be described by the following inequalities:

$$\sum_{f\in\mathcal{F}_{e_2}\wedge f\in\mathcal{F}_d} \chi_r^k(e_1,f) \le \sum_{r\in\mathcal{R}_d\cup\mathcal{R}_u} \delta_r^k(e_1,e_2) \qquad e_1 \in \mathcal{B}\cup\mathcal{U} \tag{4.12}$$

$$\sum_{f \in \mathcal{F}_{e_1} \wedge f \in \mathcal{F}_u} \chi_r^k(e_2, f) \leq \sum_{r \in \mathcal{R}_u} \delta_r^k(e_1, e_2) \qquad (4.13)$$
$$e_2 \in \mathcal{B}$$

where $\mathcal{F}_{e_1}, \mathcal{F}_{e_2}$ are the respective sets of traffic flows requested for download/upload by UEs e_1 and e_2. In (4.12), strict inequality holds when e_1 is a serving UE and the total amount of data it is caching for f is smaller than what could be transferred over the link between the two nodes.

A list of symbols and definitions used in this chapter can be found in Table 4.1.

Table 4.1: List of symbols

Symbol	Description	Symbol	Description
$\mathcal{B}$	Collective set of infrastructure PoAs b	$\mathcal{U}$	Collective set of UEs u
$\mathcal{F}$	Comprehensive set of traffic flows f	$\mathcal{F}_d, \mathcal{F}_u$	Set of download flows f_d and upload flows f_u, respectively
$e(f)$	Time step at which flow f is initiated	$l(f)$	Amount of data requested in traffic flow f (in bits)
$d(f)$	Completion deadline for traffic flow f	$\mathcal{R}$	Total set of RBs r
$\mathcal{R}_d, \mathcal{R}_u$	The set of downlink and uplink resources, respectively	$\delta_r^k(s, e)$	The potential amount of data that can be transferred from serving endpoint s, to receiving endpoint e, on RB r, at time step k
c_f	Central frequency of the carrier	$\chi_r^k(s, f)$	Total amount data transferred from serving endpoint s, related to traffic flow f, on the allocated RB r at time step k
$h^k(f)$	The total amount of traffic flow f downloaded/uploaded by the requesting user $u(f)$ until time step k	$\mathcal{K}$	Set of 1-ms time steps k

Table 4.1: List of symbols

Symbol	Description	Symbol	Description
$X^k(f)$	The amount of traffic flow f downloaded/uploaded from/to serving endpoint s	$P(s,u)$	The power with which serving endpoint s transmits to UE u
$A(s,u)$	The total attenuation experienced by the signal transmitted from serving endpoint s to a UE u	G_T	The antenna gain of the macro base station
$AP(\theta(b,u))$	The antenna pattern factor, which depends on the angle $\theta(b,u)$ between b's antenna maximum direction and the direction between the antenna and UE u	$PL(s,u)$	The path loss experienced by the signal between serving endpoint s and UE u
$I_r^k(e)$	Total amount of interference experienced by endpoint e on RB r	$\gamma_r^k(s,u)$	The signal to noise plus interference ratio (SINR) experienced by u on RB r when receiving from serving endpoint s at time step k

4.2 A dynamic programming approach to resource allocation

In this section we present the model we develop using standard dynamic programming methodology, in order to tackle the resource allocation problem formulated in the previous section. As shown by previous work [39, 41], the problem of radio resource allocation in LTE-based systems is NP-hard, even when less complex scenarios than ours are considered. Thus, we resort to approximate dynamic programming in order to solve the model in realistic, large-scale scenarios.

4.2.1 The dynamic programming model

Dynamic programming is an optimization technique based on breaking a complex problem into simpler, typically time-related, subproblems. Since scheduling in LTE

systems occurs every subframe, we solve the resource allocation problem every time step k. A dynamic programming model consists of the following elements (denoted by bold-face Latin letters) [77]:

- the *state variable*, $\mathbf{s}^k$, which describes the state of the system at time k;
- the *action* set, $\mathbf{A}^k = \{\mathbf{a}^k\}$, i.e., all possible decisions that can be taken at time k;
- an exogenous (and potentially stochastic) *information process*, accounting for information on the system becoming available at time k;
- the *cost* of an action, $\mathbf{C}(\mathbf{s}^k, \mathbf{a}^k)$, i.e., the immediate cost due to the selected action, given the current state;
- the *value*, $\mathbf{V}(\mathbf{s}^k, \mathbf{a}^k)$, of ending up at a new state $\mathbf{s}^{k+1}$, determined by the current state and action; such value is given by the cost associated with the optimal system evolution from $\mathbf{s}^{k+1}$.

Table 4.2 summarizes these quantities, their meaning in our system and the symbols we use for them. Fig. 4.3 shows how each of them is used in the model.

In particular, in our case the system state at generic time k is given by the set of duplets: $\mathbf{s}^k = \{h^k(f), e(f)\}$. Each duplet refers to a different traffic flow f, and includes (i) the amount $h^k(f)$ of the traffic flow transferred to or by the UE

and (ii) the flow initiation time $e(f)$. Clearly, at time k we only know those initiation times $e(f) \leq k$.

An action is a set of triplets, each defining PoA s should serve which traffic flow f, and using which RB r, i.e., $\mathbf{a}^k = \{(s, f, r)\}$. In simpler terms, an action is a realization of resource allocation. Note that, since we are considering both download and upload traffic, $s \in \mathcal{B} \cup \mathcal{U}$.

The dynamic programming model works as shown in Fig. 4.3 (a): for each time step we enumerate and evaluate the possible actions, select (and enact) the best one, and move to the next time step. At this point, we become aware of which content items have been recently requested, hence we can determine the next system state.

Fig. 4.3 (b) offers a more detailed view. The starting point is given by the current state $\mathbf{s}^k$ and the set of actions describing the possible resource allocations (steps 1 and 2 in the figure). The latter step is further elaborated in the next section. For each action, we compute the potential (δ) and, then, the actual (X) amount of data that can be transferred between every pair of endpoints (steps 3–4), using the algorithms we provide below. Given the variables X, we can update the total amount of data that each requesting UE u can download/upload by the beginning of the next time step using (4.1).

Table 4.2. List of symbols used in the dynamic programming model

Quantity and symbol	Description
Current state $\mathbf{s}^k$	Set of duplets, each referring to a different UE-content pair. A duplet includes the amount of traffic flow f already transferred to/by u, $h^k(f)$, and the want-time $e(f)$ if no greater than k
Action to take $\mathbf{a}^k$	Set of triplets indicating which PoA s, should serve which traffic flow f on which RB, i.e., (s, f, r)
Exogenous information	Flow initiation-times $e(f)$
Cost $\mathbf{C}(\mathbf{s}^k, \mathbf{a}^k)$	Ratio of the amount of content still to be retrieved to the remaining time before the deadline for content delivery expires
Value $\mathbf{V}(\mathbf{s}^k, \mathbf{a}^k)$	Total (expected) costs due to the system future evolution
$\tilde{\mathbf{A}}^k$	Auxiliary action space, i.e., set of values expressing the level of preference associated to each type of endpoint

For each action $\mathbf{a}^k$, we can then evaluate the cost $\mathbf{C}(\mathbf{s}^k, \mathbf{a}^k)$ the system incurs if $\mathbf{a}^k$ is selected (step 5 in Fig. 4.3 (b)). We define such cost as the sum over all active traffic flows of the ratio of the amount of data still to be transferred to the time before the delivery deadline expires, i.e.,

$$\mathbf{C}(\mathbf{s}^k, \mathbf{a}^k) \quad = \quad \sum_{f \in \mathcal{F}: e(f) \leq k} \frac{l(f) - \left(h^k(f) + \sum_{s \in \mathcal{B} \cup \mathcal{U}} \chi^k(s, f)\right)}{e(f) + d(f) - k}$$

By the above definition, a lower cost is therefore obtained for those allocation strategies, $\mathbf{a}^k$, assigning more resources to transfers that are closer to their completion deadline.

The value $\mathbf{V}(\mathbf{s}^k, \mathbf{a}^k)$ (step 6 in Fig. 4.3 (b)) is yielded by the sum of the costs $\mathbf{C}(\mathbf{s}^{k+1}, \mathbf{a}^{k+1}) + \mathbf{C}(\mathbf{s}^{k+2}, \mathbf{a}^{k+2}) + \ldots$. In other words, it is the cost that will be paid in the future, after the system has reached state $\mathbf{s}^{k+1}$. State values do not normally admit a closed-form expression. In standard dynamic programming [77, Ch. 3], they are computed by accounting for all possible states and actions, typically leading to an exceedingly high complexity in non-toy scenarios. We address such an issue by applying approximate

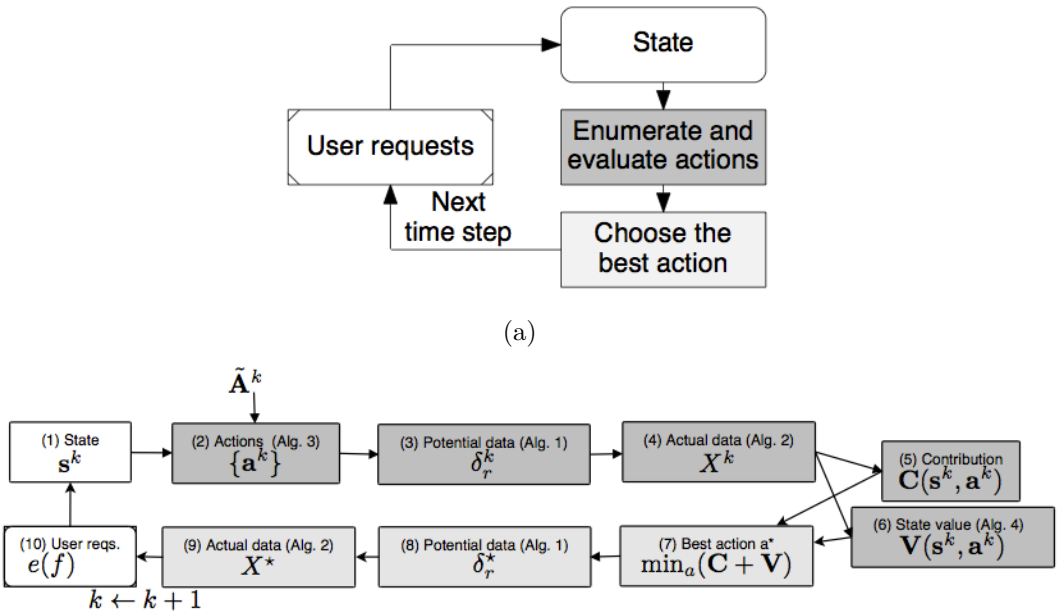


Figure 4.3. Dynamic programming: (a) Main steps involved; (b) Detailed view. Given the current state (1), the set of possible actions can be determined (2). For each action, the potential (3) and actual (4) amount of content transferred between the pairs of endpoints can be computed. These values are further used to compute the cost (5) of an action, and to estimate the value of the state it leads to (6). The latter two figures are used (7) to select the best action. The resulting data transfers (8-9), along with the users that just became interested in a content, define the next state. The description of the notations appearing in the flow diagram can be found in Tables 4.1 and 4.2.

dynamic programming as described in the following section.

Once $\mathbf{C}(\mathbf{s}^k, \mathbf{a}^k)$ and $\mathbf{V}(\mathbf{s}^k, \mathbf{a}^k)$ have been computed for all actions, the action $\mathbf{a}^*$ minimizing the cost $\mathbf{C}(\mathbf{s}^k, \mathbf{a}^k) + \mathbf{V}(\mathbf{s}^k, \mathbf{a}^k)$ is selected (step 7 in Fig. 4.3 (b)). Given $\mathbf{a}^*$, the corresponding actual amount of transferred data can be calculated (steps 8-9). This, along with fresh information on user requests (step 10), leads to the next state $\mathbf{s}^{k+1}$.

Next, we detail how to compute the potential δ_r^k (Algorithm 4.1) and actual X^k (Algorithm 4.2) amount of data, while taking into account the interference due to the spatial reuse of radio resources. It is worth stressing that, the processes we describe below have a very low computational complexity, namely $O(|\mathcal{U}|)$, while maintaining a high level of realism.

Algorithm 4.1 is used in steps 3 and 8 in Fig. 4.3 (b). In line 8, we account for the fact that every active endpoint pair may create interference at other users on the

particular used RB. All interference values are computed within the first loop. The second loop computes the SINR (line 14) and maps it onto the amount of data that can be transferred on RB r during time step k (line 8). We perform such mapping by using the experimental values in [76].

Algorithm 4.1 Computing the amount δ of data that can be potentially transferred

Require: $\mathbf{a}^k$

1: $I_r^k(u) \leftarrow 0, \forall u \in \mathcal{U}, \forall r \in \mathcal{R}_d \cup \mathcal{R}_u$
2: **for all** $(s, f, r) \in \mathbf{a}^k$ **do**
3: **if** $f \in \mathcal{F}_d$ **then**
4: $e_1 \leftarrow s$, $e_2 \leftarrow u(f)$
5: **else if** $f \in \mathcal{F}_u$ **then**
6: $e_1 \leftarrow u(f)$, $e_2 \leftarrow s$
7: **for all** $u \in \mathcal{U} \setminus \{e_1, e_2\}$ **do**
8: $I_r^k(u) \leftarrow I_r^k(u) + \mathbb{1}_{A(e_1,u)>0} P(e_1, e_2)/A(e_1, u)$
9: **for all** $(s, f, r) \in \mathbf{a}^k$ **do**
10: **if** $f \in \mathcal{F}_d$ **then**
11: $e_1 \leftarrow s$, $e_2 \leftarrow u(f)$
12: **else if** $f \in \mathcal{F}_u$ **then**
13: $e_1 \leftarrow u(f)$, $e_2 \leftarrow s$
14: $\gamma_r^k(e_1, e_2) \leftarrow \frac{P(e_1,e_2)}{A(e_1,e_2)(N+I_r^k(e_2))}$
15: $\delta_r^k(e_1, e_2) \leftarrow$ **sinr_to_delta**$(\gamma_r^k(e_1, e_2))$
16: **return** $\delta_r^k(e_1, e_2)$

Algorithm 4.2 instead refers to steps 4 and 9 in Fig. 4.3 (b). The algorithm takes as input the action $\mathbf{a}^k$ and the amount of data $\delta_r^k(e_1, e_2)$ that can be potentially transferred as a consequence of this action (computed in the previous step through Algorithm 4.1). Then, for each triplet in $\mathbf{a}^k$,the amount transferred on RB r, $\chi_r^k(s, f)$, is determined. This amount is given by the minimum between the amount of data that serving PoA s still has or awaits from the requesting UE u and the amount of data that can be accommodated in the RB[3]. Finally, the X-value is obtained by summing the χ values over all RBs (line 9).

Notwithstanding the low complexity implied by the computation of the δ and χ quantities, standard dynamic programming itself is affected by the well-known "curse of dimensionality" [77], which makes it impractical for all but very small scenarios.

[3]The computation of this amount assumes that content is downloaded in order, i.e., from the first to the last byte. It does not hold for p2p applications, however file transfers and multimedia streaming do behave this way.

In our scenario, this problem is caused mainly by the exceedingly large set of possible actions and the aforementioned complexity in the evaluation of the future cost $\mathbf{V}$. As

Algorithm 4.2 Computing the amount X of data being actually transferred

Require: $\mathbf{a}^k$, δ_r^k
1: $X^k(s,f) \leftarrow 0$, $\chi_r^k(s,f) \leftarrow 0$,
2: **for all** $(s,f,r) \in \mathbf{a}^k$ **do**
3: **if** $f \in \mathcal{F}_d$ **then**
4: $e_1 \leftarrow s$, $e_2 \leftarrow u(f)$
5: $\chi_r^k(s,f) \leftarrow \min\{l(f) - h^k(f), \quad \delta_r^k(e_1,e_2)\}$
6: **else if** $f \in \mathcal{F}_u$ **then**
7: $e_1 \leftarrow u(f)$, $e_2 \leftarrow s$
8: $\chi_r^k(s,f) \leftarrow \min\{l(f) - h^k(f), \quad \delta_r^k(e_1,e_2)\}$
9: $X^k(s,f) \leftarrow X^k(s,f) + \chi_r^k(s,f)$
10: **return** $X^k(s,f), \chi_r^k(s,f)$

an example, consider the set $\mathbf{A}^k$ of possible actions that can be taken at time step k, which includes all possible sets of (s,f,r) triplets. There are $|\mathcal{B}\cup\mathcal{U}||\mathcal{B}\cup\mathcal{U}||\mathcal{R}_d\cup\mathcal{R}_u|$ such tuples and, thus, a total of $2^{|\mathcal{B}\cup\mathcal{U}||\mathcal{B}\cup\mathcal{U}||\mathcal{R}_d\cup\mathcal{R}_u|}$ possible actions $\mathbf{a}^k \in \mathbf{A}^k$. Some of these actions can be discarded as meaningless, e.g., allocating RBs to a UE that has already completed its transfer. Others, e.g., having a UE receive from more than one endpoint in the same time step, or receiving a content while transmitting, are ruled out by technology constraints [21]. Furthermore in 4.1, we laid down scenario-imposed rules regarding which links can use which sets of RBs, which also eliminate a great deal of triplet combinations. However, the very fact that the size of $\mathbf{A}^k$ grows exponentially with the number of UEs, PoAs and RBs makes a standard dynamic programming model not scalable. For a similar reason, the evaluation of $\mathbf{V}$ stemming from $\mathbf{A}^k$ is exceedingly cumbersome. Indeed, one should consider all possible system evolutions starting from the current state, by selecting at each future time step the optimal action. Thus, we resort to ADP and propose the algorithms below so as to efficiently generate and rank actions, hence find a solution with low computational complexity.

4.2.2 The ADP solution

Recall that the immediate cost $\mathbf{C}$ of each action can be evaluated with very low complexity, thanks to Algorithms 4.1 and 4.2. Thus, in order to ensure scalability, it is sufficient to act along two directions: (i) making the number of actions to be evaluated at each time step smaller and independent of the number of UEs and PoAs, and (ii) reducing the complexity of evaluating the future cost $\mathbf{V}$ of an action.

Of course, it is not possible to achieve such a result while keeping the optimality guarantee. However, such an approach has been shown to be very effective [77, Ch. 1], as also confirmed by our performance evaluation in Sec. 4.3. Below, we describe how we tackle the two issues.

Reducing the action space

The procedure described here is performed in the second step of the dynamic programming model in Fig. 4.3 (b), where we define the set of all possible actions $\mathbf{A}^k$. Considering that the subsequent steps in the model have to be performed for every action $\mathbf{a}^k$ in the set, reducing this set implies reduction in the complexity of the procedure as a whole.

To do so, we define an auxiliary action space $\tilde{\mathbf{A}}^k$, whose size is much smaller than the original action space $\mathbf{A}^k$ and, more importantly, does not grow with the number of UEs or BSs. Then, we show a deterministic (and computationally efficient) way to map an action $\tilde{\mathbf{a}}^k \in \tilde{\mathbf{A}}^k$ of the auxiliary action space into an action $\mathbf{a}^k \in \mathbf{A}^k$. It follows that the actions we evaluate (steps 5–7 in Fig. 4.3 (b)) are only those $\mathbf{a}^k \in \mathbf{A}^k$ that have a correspondence in $\tilde{\mathbf{A}}^k$.

To determine the auxiliary action space, we proceed as follows: we ask ourselves what kind of choice has the highest relevance in a system such as ours. The most significant one is to rank transfer paradigms, i.e., using macro PoAs, micro PoAs or D2D – and test which combination of them yields the highest throughput and carries the least interference. We thus rank the "importance" of each paradigm by a triplet of real values $\alpha_M, \alpha_m, \alpha_D \in [0,1]$. These values indicate which endpoints should be preferably used, as shown in Algorithm 4.3, and each triplet represents an auxiliary action $\tilde{\mathbf{a}}^k$. For the set of auxiliary actions to be manageable, we need to discretize each value in the α triplet. The set $\tilde{\mathbf{A}}^k$ is thus finite and we can control its size by choosing the granularity of each α. This is our tuning knob for scalability purposes.

Algorithm 4.3 takes as input an action $\tilde{\mathbf{a}}^k$ and maps it onto an action $\mathbf{a}^k$ (line 11). Its logic is straightforward: we serve traffic flows, starting from the neediest ones, selecting the most effective serving PoA.

More specifically, in line 1, we identify the set $\mathcal{F}_a^k \subseteq \mathcal{F}$ of active traffic flows, i.e., flows with an incomplete transfer. This set is sorted (line 2) by the initiation time $e(f)$, so the earliest requests are given higher priority.

Next, the set of devices which can act as potential servers, S_d^k, is defined to ensure that no devices which have current active requests can simultaneously act as servers to other devices (line 3). Then, for each flow $f \in \mathcal{F}_a^k$, we loop over the potential servers s and RBs r that may be used to transfer content (line 3), and asses all potential combinations of s, f and r that make sense and are allowed in the scenario. Note that we have explicitly reflected all the scenario-imposed constraints

Algorithm 4.3 Mapping α-triplets into actions

Require: $\tilde{\mathbf{a}}^k = (\alpha_M, \alpha_m, \alpha_D)$

1: $\mathcal{F}_a^k \leftarrow \{f \in \mathcal{F}\colon e(f) < k \wedge h^k(f) < l(f) \wedge e(f) + d(f) \geq k\}$
2: **sort** $\mathcal{F}_a^k$ **by** $e(f)$
3: $S_d^k \leftarrow \{u \in \mathcal{U} \setminus \{u(f)\}$ s.t $u(f)\colon f \in \mathcal{F}_a^k\}$
4: **for all** $f \in \mathcal{F}_a^k$ **do**
5: **for all** s, r **do**
6: $\sigma(s, f, r) \leftarrow 0$
7: **if** $f \in \mathcal{F}_d \wedge s \in \mathcal{B} \wedge r \in \mathcal{R}_d$ **then**
8: **compute** $\chi_r^k(s, f)$ (Algorithm 4.2)
9: $\sigma(s, f, r) \leftarrow \chi_r^k(s, f)$
10: **if** $f \in \mathcal{F}_d \wedge s \in S_d^k \wedge r \in \mathcal{R}_{D2D}$ **then**
11: **compute** $\chi_r^k(s, f)$ (Algorithm 4.2)
12: $\sigma(s, f, r) \leftarrow \chi_r^k(s, f)$
13: **if** $f \in \mathcal{F}_u \wedge s \in \mathcal{B} \wedge r \in \mathcal{R}_u$ **then**
14: **compute** $\chi_r^k(s, f)$ (Algorithm 4.2)
15: $\sigma(s, f, r) \leftarrow \chi_r^k(s, f)$
16: **if** $s \in \mathcal{B}_M$ **then**
17: $\sigma(s, f, r) \leftarrow \sigma(s, f, r) \cdot \alpha_M$
18: **else if** $s \in \mathcal{B}_m$ **then**
19: $\sigma(s, f, r) \leftarrow \sigma(s, f, r) \cdot \alpha_m$
20: **else if** $s \in \mathcal{U}$ **then**
21: $\sigma(s, f, r) \leftarrow \sigma(s, f, r) \cdot \alpha_D$
22: $s^\star, r^\star \leftarrow \arg\max_{s,r} \sigma(s, f, r)$
23: $t_{curr} \leftarrow 0$, $t_{new} \leftarrow 0$
24: **for all** $(s, \phi, r) \in \mathbf{a}^k$ **do**
25: **compute** $\chi_r^k(s, \phi)$ (Algorithm 4.2)
26: $t_{curr} \leftarrow t_{curr} + \chi_r^k(s, \phi)$
27: **for all** $(s, \phi, r) \in \mathbf{a}^k \cup (s^\star, f, r^\star)$ **do**
28: **compute** $\chi_r^k(s, \phi)$ (Algorithm 4.2)
29: $t_{new} \leftarrow t_{new} + \chi_r^k(s, \phi)$
30: **if** $t_{new} > t_{curr}$ **then**
31: $\mathbf{a}^k \leftarrow \mathbf{a} \cup (s^\star, f, r^\star)$
32: **return** $\mathbf{a}^k$

in lines 7–13. Most importantly in line 13, we set the constraint on which set of RBs may be used for D2D communications. Namely, since we consider two different scenarios, this can be either the uplink ($\mathcal{R}_u$) or the downlink ($\mathcal{R}_d$) RB set. We can

switch between the two scenarios by simply defining $\mathcal{R}_{D2D}$ beforehand.

Specifically, for each triplet (s, f, r), allowed in our scenario, we compute a score σ, initialised to zero in line 6, which is equal to to the amount of data (computed by Algorithm 4.2) that s may transfer to requesting UE of flow f. Each score σ is then weighted by the α-coefficient corresponding to the type of server s, setting priorities over the different possible data transfer paradigms. As an example, the α-coefficients give us leverage to encourage D2D transfers by setting a high value for α_D, or to limit the usage of macroBSs to users that have no other means to be served by setting α_M to a low value. In line 2, we select the pair (s, r) corresponding to the highest score over all possibles RBs. Notice that by selecting only one pair in line 2, we honor the technology constraint by which each user can either download, upload or serve data from at most one source and to at most one destination in a given time step.

However, before conclusively including the selected triplet $(s^\star, f, r^\star)$ in the allocation strategy yielded by $\mathbf{a}^k$, we check whether it increases the total amount of data transferred in the network or not (lines 24–30). While verifying that, we resort again to the interference-aware Algorithms 4.1 and 4.2 to compute the δ and X values. If the amount of data grows, the triplet is added to action $\mathbf{a}^k$ (line 6). In conclusion, we stress that the size of the auxiliary action space $\tilde{\mathbf{A}}$ is small and it is independent of the number of UEs and BSs. Thus, we have achieved our scalability goal.

Evaluating the state values

To evaluate an action, it is important to compute the value of the state $\mathbf{s}^{k+1}$ the action leads to. As already stated, the value of a state corresponds to the sum of the costs we will pay due to future actions, if these are chosen optimally. Clearly, if we set $\mathbf{V}(\mathbf{s}^k, \mathbf{a}^k) = 0$ for all actions, i.e., we select the action that seems more profitable at the current step, we end up adopting a greedy strategy. However, in network scenarios where D2D is allowed, accounting for future actions may be of particular relevance: e.g., transmitting to some users at a faster pace, so that they can act as serving UEs later, may be beneficial to the whole network.

It follows that we need to compute the value function $\mathbf{V}$ accurately enough, while keeping the complexity low. To do so, we resort to the methodology typically used in ADP. Such methodology [77, Ch. 9] implies that, at each step k, we fix the sequence of future actions, starting from state $\mathbf{s}^{k+1}$. We apply this procedure to our problem as described in Algorithm 4.4.

The algorithm takes as input: (i) the current state $\mathbf{s}^k$ and the current action to be evaluated $\mathbf{a}^k$ (i.e., the two elements determining next step $\mathbf{s}^{k+1}$), and (ii) the future actions that we expect will be taken. In order to compute the latter, we start by assuming that the conditions experienced by a user do not change during

the transfer time. This is a fair assumption since, as shown by Figs. 4.5(b),(c) and 4.6(b),(c) in Sec. 4.3, users complete their transfer in hundreds of ms, hence the movement of pedestrian users during content transfer is negligible.

Also, note that the procedure for computing the value function $\mathbf{V}$ is repeated at every time step k. We feed such information to a machine learning model, so as to compute future actions $\{\mathbf{a}^{k+1}, \ldots, \mathbf{a}^{K}\}$ [77, Ch. 9].

Algorithm 4.4 Estimating the value of a state

Require: $\mathbf{s}^k$, $\mathbf{a}^k$, $\{\mathbf{a}^{k+1}, \ldots, \mathbf{a}^{K}\}$

1: $v \leftarrow 0$
2: **for** $q = k+1 \rightarrow K$ **do**
3: **for all** $(s, f, r) \in \mathbf{a}^q$ **do**
4: **if** $e(f) \leq k \wedge h^q(f) < l(f) \wedge e(f) + d(f) \geq q$ **then**
5: **compute** $X^q(s, f)$ **using** Algorithm 4.2
6: $\hat{h}^{q+1}(f) \leftarrow \hat{h}^q(f) + X^q(s, f)$
7: **compute** $\mathbf{C}(\mathbf{s}^q, \mathbf{a}^q)$
8: $v \leftarrow v + \mathbf{C}(\mathbf{s}^q, \mathbf{a}^q)$
9: **return** $\mathbf{V}(\mathbf{s}^k, \mathbf{a}^k) = v$

Next, we exploit the estimated information on the system to compute, at each future time step $q > k$, the Xi values for each communication foreseen by action $\mathbf{a}^q$ are calculated 5) using the algorithms presented in Sec. 4.2.1.

In line 6 for each step $q > k$, given the previous state and the χ values, we apply (4.1) and update the amount of data of flow f, $h^q(f)$, that each downloader/uploader can retrieve/transfer until step q. Then, we use the quantities X and h to evaluate the cost of action $\mathbf{a}^q$. Note that we cannot predict future user requests, however, due to the short time span before a transfer completion, their number is limited. Additionally, their deadline will be further away in time,[4] hence their impact is minimal (see (4.14)). At last, $\mathbf{V}(\mathbf{s}^k, \mathbf{a}^k)$ is calculated by summing all future cost contributions (line 9).

4.2.3 Solution complexity

As mentioned before, to meet the scalability requirements, the algorithms must be of sufficient low-complexity. Applying ADP, this requirement is indeed met. With reference to Fig. 4.3 (b), and assuming that the dominant factor is the number of users, the complexity is as follows: step (2), $O(2^{|\mathcal{U}|})$ with plain dynamic programming, which reduces to $O(|\mathcal{U}|)$ using Algorithm 4.3. Steps (3) and (4), $O(|\mathcal{U}|)$. Step

[4]Recall that Algorithm 4.4 is repeated at every time step k.

(5), $O(1)$. Step (6), $O(|\mathbf{A}|^k)$ with plain dynamic programming, which reduces to $O(|\mathcal{U}|)$ with Algorithm 4.4.

4.3 Performance evaluation

4.3.1 Simulation scenario

We evaluate the performance of our approach in the two-tier network that is typically used within 3GPP for LTE network evaluation [78]. The scenario comprises a service network area of 12.34 km^2, covered by 57 macrocells and, unless otherwise specified, 228 microcells. Macrocells are controlled by 19 three-sector base stations; the inter-site distance is set to 500 m. Microcells are deployed over the network area, so that there are 4 non-overlapping microcells per macrocell. The network topology is shown in Fig. 4.4. A total of 3420 users are present in the area. In particular, in order to have a higher user density where microcells are deployed, 10 users are uniformly distributed within 50 m from each micro PoA. The rest of the users are uniformly distributed over the remaining network area. Users move according to the cave-man model [79], with average speed of 1 m/s.

In line with [69, 78, 80] we assume a transmitting power of 43 dBm for macro PoAs, and 30 dBm for micro PoAs, and antenna height values of 25 m and 10 m, respectively. For the macro PoA antenna we further assume the antenna gain to be 14 dBi and the maximum attenuation 20 dB, while the micro PoA antennas are omnidirectional with 0 dBi gain. UE transmitting power is controlled, using the following values for the cell and user configuration parameters: $P_{max}|_{dB} = 23$ dBm, $P_o|_{dB} = -70$ dBm, $\rho = 0.7$ and $\Delta_{TF} = 0$. Closed loop control is disabled. We assume the antenna height of the UE to be 1.5 m.

All network nodes operate over a 10 MHz band at 2.6 GHz for downlink and at 2.5 GHz for uplink, hence we have $|\mathcal{R}_d| = |\mathcal{R}_u| = 50$ RBs.

As already mentioned, the signal propagation for infrastructure-to-device links is modelled according to ITU specifications as described in Sec. 4.1.2, while the SINR is mapped onto per-RB throughput values using the experimental measurements in [76]. The noise power level is set at -174 dBm/Hz, according to [69]. The energy consumption of the network nodes is computed according to [80].

Users may require three different types of content for download or upload:e-books, videos, or viral content.

Their characteristics and intervals between user requests are summarized in Table 4.3. We highlight that video and viral items have stricter constraints on delivery time. Content items from the e-book and video category may be requested either for download or upload. The viral content item, on the other hand, is modeled as being in high demand during a narrow time interval to mimic content becoming suddenly

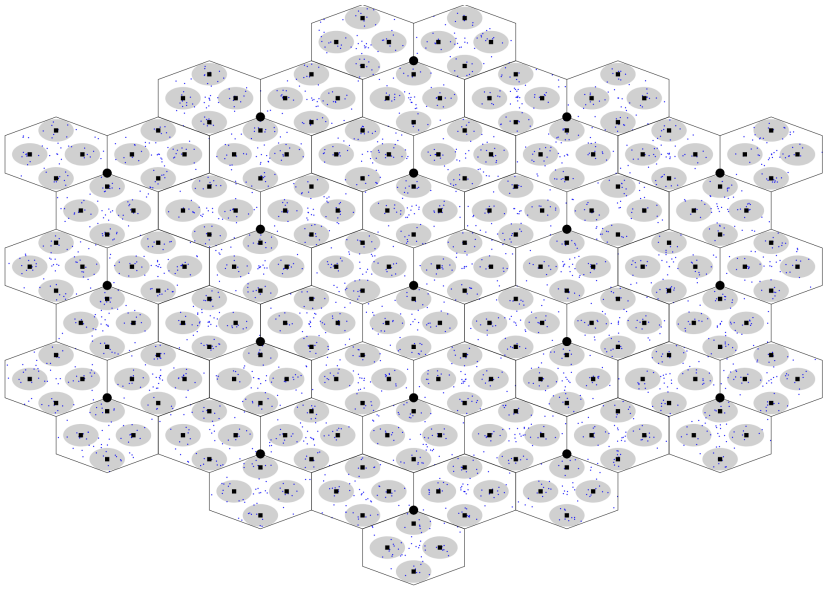


Figure 4.4. Simulation scenario.

Table 4.3. Content types

Feature	eBook	Video	Viral
Dowload request rate [items/s]	$1e-3$	$1e-3$	$5e-2$
Upload request rate [items/s]	$0.5e-3$	$0.5e-3$	–
Size [Mbit]	12	3	3
Deadline [steps]	4000	1000	1000
Request interval [steps]	1–1000	1–1000	41–60

popular, hence users may request it only for download. We consider the traffic load to be asymmetric with upload traffic being half the download one.

The scheduling decisions issued by the AC are valid for one subframe, therefore the resource allocation algorithm is performed every 1 ms. While applying our ADP approach, we consider that the values of the $\alpha_M, \alpha_m, \alpha_D$ parameters, are discretized as $\{0.1, 0.2, \ldots, 1\}$. Additional experiments with values exhibiting finer granularity have shown negligible improvement.

We will additionally consider two different D2D deployment scenarios, depending on which of the two RB sets (downlink or uplink) D2D communications are allowed to share with the cellular infrastructure. We shall refer to them as the DL scenario, when D2D operates in the downlink portion of the spectrum, and the UL scenario when it operates in the uplink portion.

4.3.2 Baseline solution

We compare our approach against a system implementing the 3GPP eICIC with a microcell bias of 15 dB and the ABS model where macro PoAs are silent in 1 out of every 2 subframes [81]. In the latter, D2D mode is not supported and UEs connect to the PoA from which they receive the strongest pilot signal. At the PoA, traffic is scheduled according to the proportional fairness (PF) algorithm, which is standard in today's LTE networks [21], as described in Sec. 2.6.2. In the following, we will refer to this benchmark scenario as PF.

4.3.3 Numerical results

The comparison between ADP and PF for both scenarios is shown in Figs. 4.5–4.7. In particular, Fig. 4.5(a) and Fig. 4.6(a) show that ADP allows the transfer of more data than the state-of-the-art (around 7%), while using over 40% less energy. Such a gain can be attributed to the lower usage of macrocells (characterized by very high transmit power), in favour of microcells and D2D. In the plot, the possible endpoints are differentiated by using different colors: black for macro PoAs, gray for micro PoAs and red for UEs. Note that the energy consumption due to D2D mode is negligible and can be barely seen in the plot. Also, under both ADP and PF, transmissions from micro PoAs are more efficient than those from macro PoAs, as the former carry a higher amount of data at a much lower energy cost.

Figs. 4.5(b)–(c) and Figs. 4.6(b)–(c) depict the CDF of the completion time of successful downloads (b) and uploads (c), for the different content categories (differentiated by the different colors). A download/upload is successful if it can be completed by the corresponding deadline. Comparing ADP (solid lines) to PF (dotted lines), we notice that in general ADP outperforms PF in terms of ensuring faster content delivery both for uploads and downloads, regardless of the D2D scenario. This is especially true for viral and video content, which have stricter deadlines. Indeed, in ADP, the cost $\mathbf{C}$ in (4.14) accounts for content deadlines, giving higher priority to those content transfers that are closer to their completion deadline. In particular, for video content, ADP is able to provide a far lower completion time for at least 90% of the successful downloads, than PF. These results are closely related to the percentage of failed transfers shown in Fig. 4.5(d) and Fig. 4.6(d), where the performance of ADP and PF are differentiated by using the orange and blue color, respectively. Clearly, ADP guarantees higher success rates than PF for all content categories and in both traffic directions, but the contrast is most dramatic for viral content. This can also be attributed to the fact that D2D is heavily used by ADP to deliver this type of content.

Fig. 4.7 highlights the improvement that ADP offers in terms of usage of radio resources, both downlink and uplink, compared to PF. Observe that, on average, in

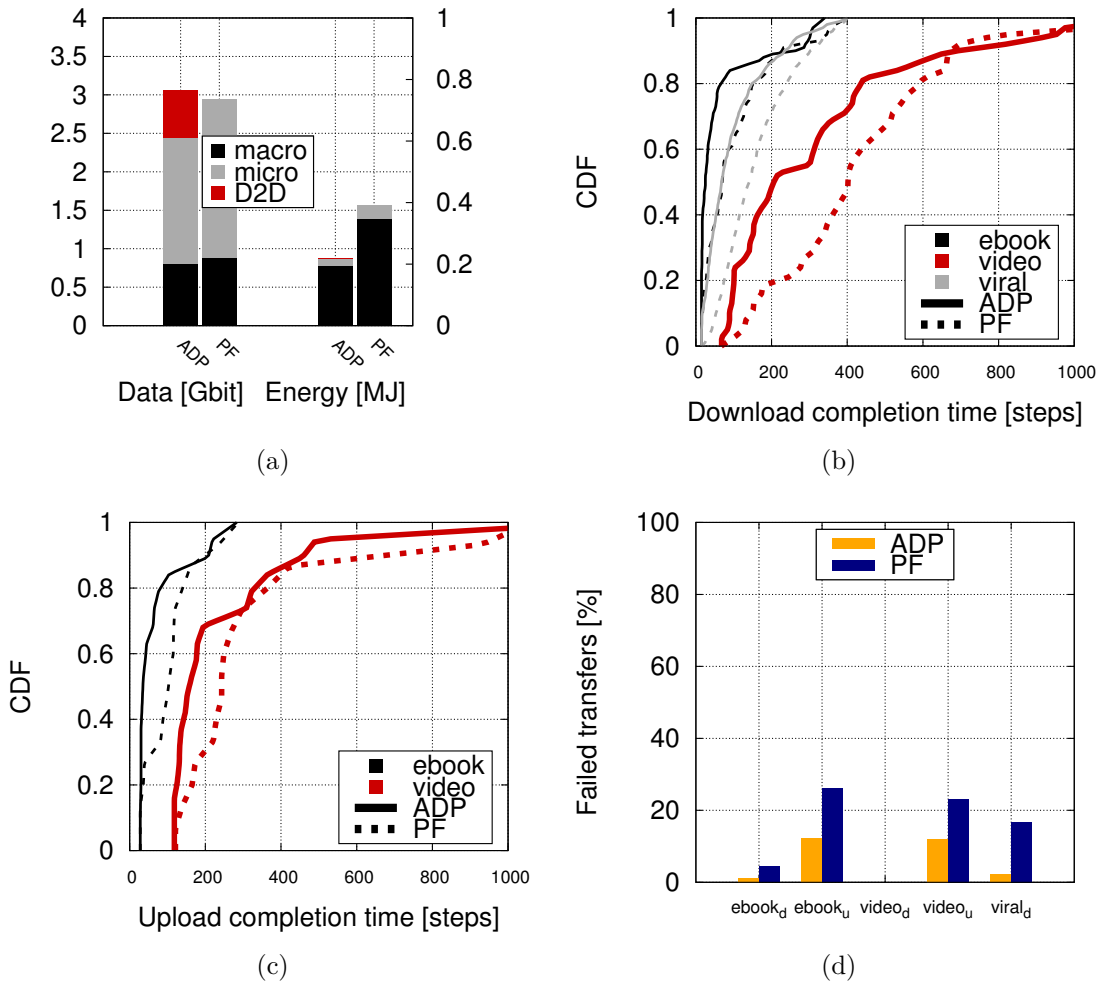


Figure 4.5. DL scenario. ADP vs. PF: (a) total amount of transferred data and consumed energy, (b) CDF of the download completion time, (c) CDF of the upload completion time, (d) failed transfers.

downlink ADP can transmit much more data per RB than PF. We observe a gain of around 35-40% in RB usage efficiency for macrocells and around 50% gain for microcells. The reason for such behavior is that our interference-aware approach is far more efficient in matching potential endpoints than the PF based system. In other words, ADP scheduling yields higher values of SINR at the receiving endpoints, hence higher data rates per RB. In the DL scenario, the amount of data per RB is especially high for D2D links, which is remarkable considering that UEs transmit at significantly lower power than microBSs or macroBSs.

By looking at Figs. 4.5–4.7, we also notice the differences in performance between the DL and UL scenario. In terms of RB usage, the values of data transferred per

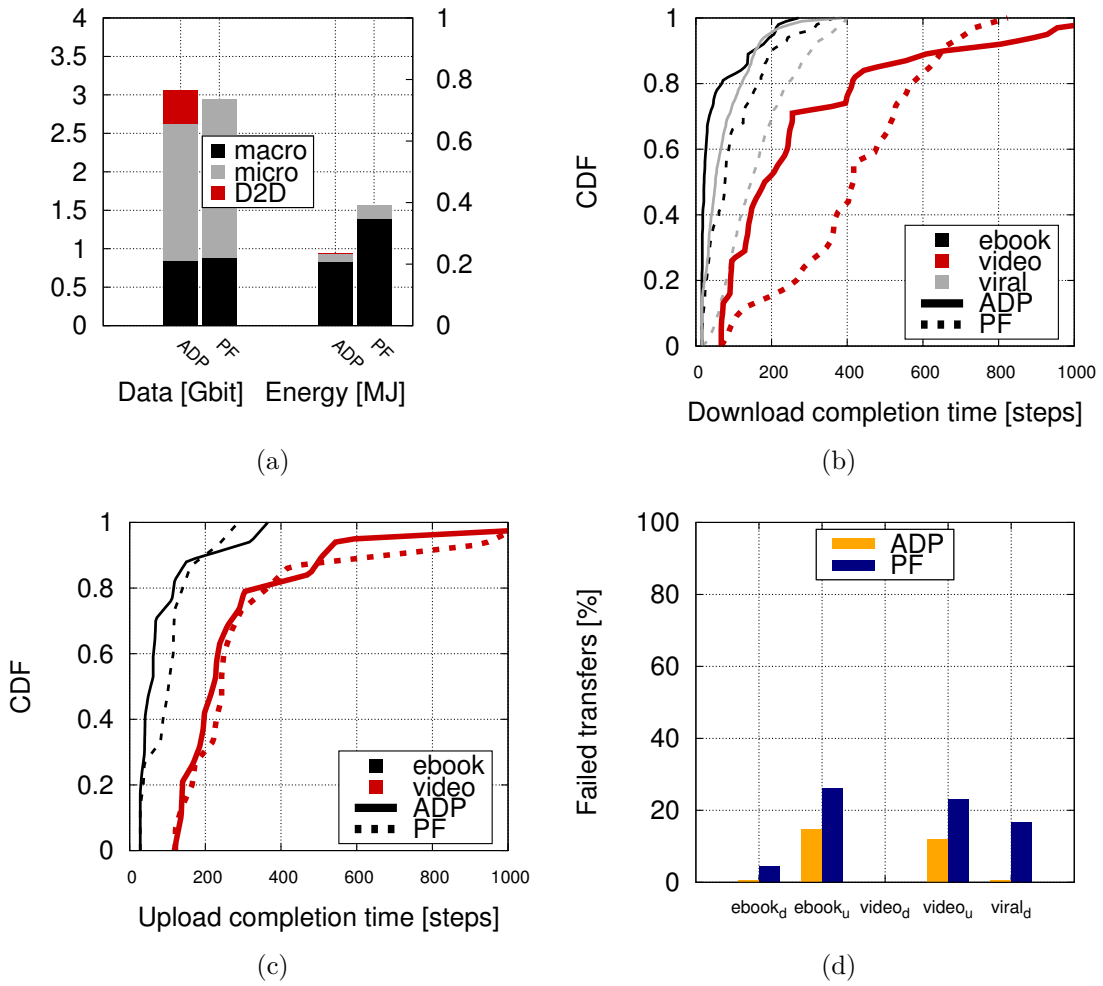


Figure 4.6. UL scenario. ADP vs. PF: (a) total amount of transferred data and consumed energy, (b) CDF of the download completion time, (c) CDF of the upload completion time, (d) failed transfers.

RB are in general higher in the DL scenario than in the UL scenario. This is mainly due to the fact that the overall achievable data rates for a certain value of SINR are higher in the downlink than the uplink, according to the experimental measurements used in our evaluation. Nonetheless, D2D in the UL scenario is significantly more efficient in using RBs compared to UE-macro links, and comparable to UE-micro links. This impacts also the overall amount of data that ADP is able to transfer through D2D in the UL scenario compared to the DL scenario, as can be noticed by comparing Fig. 4.5(a) and Fig. 4.6(a). While the overall amount of transferred data is similar in both scenarios, the amount transferred by D2D is slightly higher in the DL. In the UL scenario, the slack is picked up by microcells, which causes a

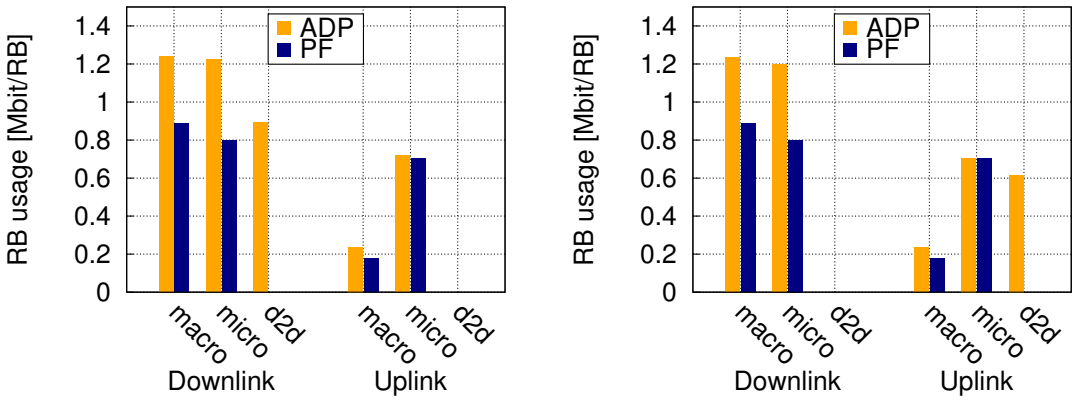


Figure 4.7. RB usage: (a) DL scenario, (b) UL scenario.

slight increase in energy consumption. We therefore conclude that the UL and DL scenarios provide similar performance in current traffic load conditions, however the DL scenario will become preferable as the upload and the download traffic tend to even.

In the same scenarios as above, we now halve the number of microcells from 228 to 114, i.e., 2 microcells per macrocell. The most noticeable effect is that, with ADP, D2D communication steps up to compensate for the missing microcells, as shown in Figs. 4.8(a) and (b). Instead, PF falls considerably short of providing the same throughput as before. Indeed, comparing these plots to Figs. 4.5(a) and 4.6(a), ADP exhibits a mere 10% drop in transferred data in the DL scenario, and around 6% drop in the UL scenario, with respect to 28% for PF. Energy consumption increases for both approaches, achieving similar levels for both. As expected, ADP tends to favour content with stricter time constraints (viral and video), at the expense of e-books. For sake of brevity, we omit plots comparing other metrics, which however confirm the above observations.

Summary. Thanks to a lesser usage of macrocells, our proposed scheme enables the transfer of 7% more data than PF, at an energy cost that is reduced by over 40%. ADP also provides a completion time that is significantly lower than PF, for most of the data transfers. Particularly striking is the success rate of viral content delivery: thanks to D2D communications, ADP exhibits 0-2% failures versus 18% of PF. As for the efficiency in RB usage, the interference-aware scheduling performed by ADP leads to an improvement of 35-40% over PF for macrocells and of around 50% for microcells. Another interesting finding is the advantage of accommodating D2D communications in the uplink or in the downlink bandwidth. Under current traffic load conditions, the two options are equally effective. However, as upload and

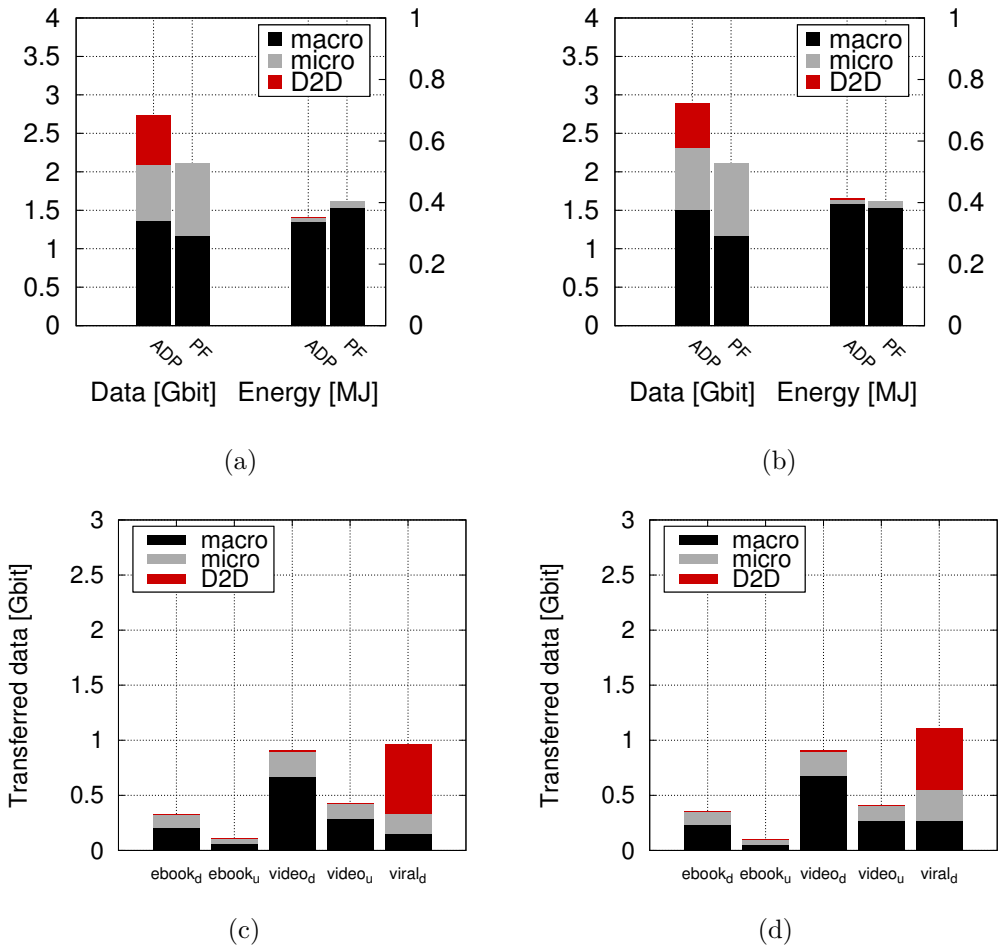


Figure 4.8. Halving the number of microcells: amount of transferred data and consumed energy, in the DL scenario (a), and in the UL scenario (b); amount of transferred data by ADP, in the DL scenario (c), and in the UL scenario (d).

download traffic tend to even out, using the downlink bandwidth will be preferable. Finally, D2D is found to be an effective, low-energy and low-cost replacement for microcell deployment.

4.4 Conclusions

In this chapter we addressed the resource allocation task in a two-tier dense network, with support for D2D communications. We devised an interference-aware centralised solution to the problem of uplink and downlink radio resource allocation, to efficiently accommodate the download and upload traffic in such a complex

network. For each traffic request, our algorithm selects which PoA should serve a user, and allocates the radio resources for such communication, in an energy-aware and spectrum-efficient manner. To reduce the complexity of the centralised problem, we presented approximate dynamic programming algorithms to generate and rank possible resource allocation decisions. This way, we obtained a low-complexity solution that can deal with realistic, large-scale scenarios. In addition, we evaluate two possible approaches to in-band, network-controlled D2D implementation, and assessed the performance of our solution for both cases. Results show that our solution combined with D2D outperforms the state-of-the-art used in today's networks both in terms of overall throughput and user experience. Furthermore, we highlight that D2D mode can be a valid, low-cost alternative to microcells in supporting traffic with little energy consumption.

Chapter 5

Interference-aware joint CC selection and resource scheduling in CA-enabled dense networks

As we showed in the previous chapter optimal resource allocation in cellular networks is known to be a hard problem; it is further exacerbated when support for advanced features such as heterogeneity and carrier aggregation are also considered. In particular, in dense multi-layered networks where radio resources are shared between different layers of base stations as well direct D2D communications, interference management can be a daunting task. Carrier aggregation, which allows the simultaneous use of several LTE component carriers to achieve high user data rates, also adds to the complexity. Indeed, the complexity of the centralised problem increases exponentially with the number of carriers available throughout the network, rendering even algorithms with reduced complexity such as the one we present in the previous chapter, unfeasible for application in realistic large-scale networks.

To adapt to a CA-enabled network, in this chapter, we propose an interference-aware heuristic distributed algorithm that jointly performs carrier selection and resource allocation to serve a mix of users with CA-enabled and legacy terminals. As reported in in Sec. 3.1.2, most of the related work treat interference management and carrier selection separately from the resource scheduling task. In general, techniques that mitigate interference such as eICIC and its modifications are usually applied for interference management. In networks with advanced features such as CA, some authors also propose the use of different carriers at different tiers to tackle the interference problem in the network.

Instead we propose to embed interference mitigation in the resource allocation procedure, by enabling the scheduler to make interference-aware decisions. The resource allocation problem is formulated by tackling the two main problems afflicting

dense multi-tier networks: inter-cell and inter-tier interference, and the complexity imposed by the availability of multiple carriers with potentially very different propagation and coverage characteristics. As a result, we propose a solution that jointly addresses carrier selection and resource allocation, while taking into account interference, in order to fairly serve CA-enabled and legacy user terminals.

We evaluate the performance of our approach in a large-scale scenario and compare it with other widely used heuristic algorithms such as Proportional-Fair scheduling and eICIC techniques. Simulation results show that the solution we propose increases system throughput, minimises energy consumption and improves spectrum utilisation, while also ensuring better fairness between CA-enabled and legacy user terminals.

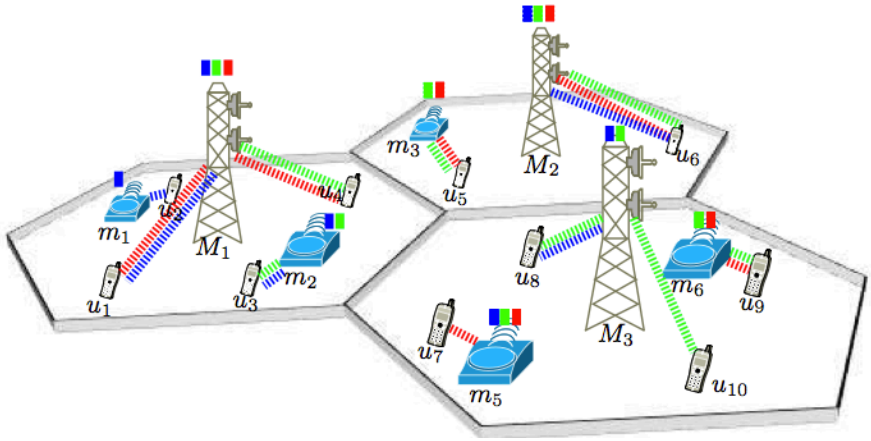
5.1 Network model

In this chapter too, we consider a two-tiered dense network, composed of macro PoAs which control macrocells, and micro PoAs controlling microcells. The topology of the network under study as well as the notation is the same as in the previous chapter. To better analyse the impact of carrier aggregation in the network, we do not consider D2D links during our analysis. However, the extension to such a scenario will be straightforward, since as we show in the previous chapter, network controlled D2D can be considered as an additional tier in the network.

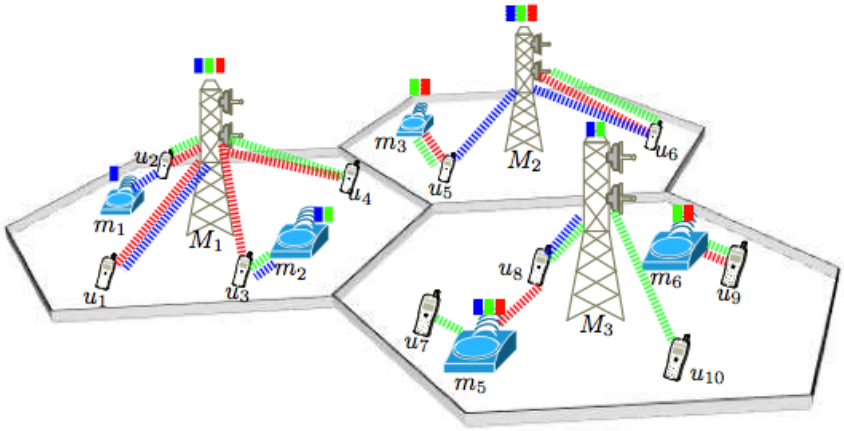
5.1.1 Carrier aggregation and serving cells

In addition to what we have already defined in Sec. 4.1, we define the comprehensive set of component carriers (CC) which is available to the infrastructure PoA, denoted by $\mathcal{C}$. Each PoA b may have at its disposal a subset of CCs, indicated by $\mathcal{C}_b$. Each CC is defined by a central frequency c_f and a certain bandwidth c_w. The c_f affects the carrier's coverage area, as the propagation conditions deteriorate greatly with increasing frequency. The component carriers may be of varying bandwidths, supported by LTE, ranging from 1.4 MHz to 20 MHz. Each carrier will contain a certain number of RBs; this number depends on the bandwidth of the carrier.

To ensure backward compatibility in the network, each PoA-CC combination is defined as a separate serving cell [78]. We denote the set of serving cells, containing all possible combinations of PoAs and CCs, by $\mathcal{S}$. Non-CA users, i.e., UEs which do not support carrier-aggregation, can connect to the selected serving cell using legacy procedures, while for CA-enabled users, additional carriers may be aggregated to provide more bandwidth. For each user, we further define a Primary Cell, $\mathbf{PCell}(u)$ and, if applicable, a set of Secondary Cells, $\mathbf{SCell}(u)$, which indicates the carriers that can be used in the carrier aggregation procedure. The Primary Cell is assumed



(a)



(b)

Figure 5.1. An example scenario with carrier aggregation: (a) *single-flow* implementation and (b) *multi-flow* implementation. UEs are denoted by $u_1, \ldots, u_{10}$, macro PoAs by $M_1, \ldots, M_3$ and micro PoAs by $m_1, \ldots, m_6$. Dotted lines with different colours (blue, green and red) correspond to the different CCs used by a pair of endpoints. The colour bars over each PoA indicate which CCs are available at the particular PoA.

to be fixed, while Secondary Cells can be activated and deactivated dynamically. Deactivating secondary cells can significantly impact power consumption at the UE terminal. The Primary Cell can however be changed, similar to a handover.

We consider two possible carrier aggregation implementations [27]. The *single flow* CA implementation allows a CA UE to be served by only one PoA at a time, using all the carriers available at that particular PoA. *Multi-flow* implementation, instead, allows a CA UE to be served simultaneously by more than one PoA with

different component carriers. The two implementations are depicted in Fig. 5.1, in an example scenario.

Table 5.1: List of additional symbols

Symbol	Description	Symbol	Description
$\mathcal{C}$	Comprehensive set of CCs c	$\mathcal{C}_b$	Subset of CCs available at BS b
c_f	Central frequency of CC c	c_w	Width of CC c
$\mathcal{S}$	Set of serving cells s	$\mathbf{t}^k(f)$	Total amount of data related to traffic flow f downloaded up to time k
PCell	Primary cell	**SCell**	Set of secondary cells

We recall that, during each time step, a certain amount of traffic,$X^k(s, f)$, flows between two communicating endpoints. In addition, for each traffic flow $f \in \mathcal{F}$, we define a status variable $\mathbf{t}^k(f)$, which denotes the total amount of data related to traffic flow f downloaded up to time k. That is, it tells us what portion of the traffic flow has already been completed:

$$\mathbf{t}^{k+1}(f) = \mathbf{t}^k(f) + \sum_{s \in \mathcal{S}} X^k(s, f), \forall f \in \mathcal{F} \tag{5.1}$$

Note that, at any given timestep k, we consider that a cell can accommodate only one traffic flow on the same RB, while it is possible to allocate several RBs to the same traffic flow.

5.2 The interference-aware joint carrier and resource scheduler

In this section we introduce a heuristic algorithm for constructing an interference and traffic-aware resource allocation strategy, so as to efficiently schedule download traffic flows while ensuring backward compatibility and fairness for legacy users.

The aim of the algorithm is to enable the controller to decide at each time step k: (i) which PoA should serve each traffic request, i.e., flow, (ii) on which available carriers and (iii) which RB(s) to employ for such communication. To achieve this, the algorithm assigns an *urgency* value to every active flow, which depends on the characteristics of the requested traffic, such as data size and delivery deadline. Additionally, for each potential resource allocation decision, it calculates a *pollution* value that accounts for the interfering impact a specific decision may have on the system. Using these parameters, the algorithm produces a resource

allocation strategy, denoted by $\mathbf{a}^k$. The strategy $\mathbf{a}^k$, we recall, is a set of triplets (s, f, r) indicating which serving cell (i.e., which PoA-CC combination) is chosen to serve which traffic flow f on which RB r. It is assumed that the algorithm is repeated every subframe and performed either in a centralised manner, with one AC making scheduling decision for all the cells under its control or in a distributed manner by grouping PoAs into clusters.

The main steps of the algorithm are provided in Alg. 5.1. Four auxiliary algorithms are used within the main algorithm; they are detailed in Alg. 5.2, Alg. 5.3, Alg. 5.4 and Alg. 5.5, respectively.

5.2.1 Building the allocation strategy

At each time step k, the algorithm is provided with updated and fresh information regarding incoming flow requests and status information, $\mathbf{t}^k$, regarding all other active flows. Additional input to the algorithm may be the UEs' predetermined **PCell** and **SCell** sets. If these sets are not provided, the algorithm will dynamically configure them during the allocation procedure.

Algorithm 5.1 Constructing the allocation strategy

Require: $\mathbf{t}^k$, $e(f)$, l_f, d_f, $\mathcal{S}$, $\mathcal{R}$, **PCell**, **SCell**
1: $\mathbf{a}^k \leftarrow \emptyset$ and $\mathbf{t}_a^k \leftarrow 0$
2: $count = 0$
3: **repeat**
4: $\quad \mathcal{F}_a^k \leftarrow \{f \in \mathcal{F} \colon e(f) \leq k \wedge \mathbf{t}^k(f) < l_f \wedge e(f) + d_f \geq k\}$
5: $\quad$ **compute** $urgency(f)$ using Eq. (5.2), $\forall f \in \mathcal{F}_a^k$
6: $\quad$ **sort** $\mathcal{F}_a^k$ **by** *urgency*
7: $\quad$ **compute** $\mathbf{W}^k$ (Alg. 5.2)
8: $\quad$ **construct** $\mathbf{a}^k$ (Alg. 5.3)
9: $\quad$ **update** $\mathbf{t}_a^k$ using Eq. (5.3)
10: $\quad count \leftarrow count + 1$
11: **until** $count \geq |\mathcal{R}|$
12: **return** $\mathbf{a}^k$

Initially, the allocation strategy $\mathbf{a}^k$ is an empty set, which is then iteratively filled with resource allocation triplets (s, f, r). A temporary status variable $\mathbf{t}_a^k$, initialised at 0, is also introduced, to store the evolving status information for the scheduled active flows.

The following steps in the main loop are repeated at most $|\mathcal{R}|$ times to ensure that all available RBs are evaluated at least once.The first step of the main loop identifies the set of active traffic flows $\mathcal{F}_a^k$. These are flows that have started before or at the current time step k, have not been completed yet, and aren't past the

deadline, that is, flows that fulfil the following conditions: $e(f) \leq k$, $\mathbf{t}^k(f) < l_f$ and $e(f) + d_f \geq k$. To each active flow, the algorithm assigns an *urgency* value, defined as the ratio between the amount of data still left to complete the flow and the time before the delivery deadline expires:

$$urgency(f) = \frac{l_f - \mathbf{t}_a^k(f)}{e(f) + d_f - k}, \forall f \in \mathcal{F}_a^k \,. \tag{5.2}$$

This value gives us a sense of *urgency* for scheduling a particular flow. The flows are then sorted according to such value.

Algorithm 5.2 Calculating the weight matrix

Require: $\mathbf{a}^k$, $\mathbf{t}^k$, $\mathcal{F}_a^k$, $\mathcal{S}$, $\mathcal{R}$, **PCell**, **SCell**
1: **for all** $f \in \mathcal{F}_a^k$ **do**
2: **define** $\mathcal{S}_a^k(u_f)$ **given PCell**(u_f), **SCell**(u_f) (Alg. 5.4 or Alg. 5.5)
3: **for all** $s \in \mathcal{S}_a^k$ **do**
4: $\mathcal{R}_c^s \leftarrow \{r \in \mathcal{R}_c \colon (s, r) \notin \mathbf{a}^k\}$
5: **for all** $r \in \mathcal{R}_c^s$ **do**
6: $\mathbf{a}^\star \leftarrow (s, f, r_c)$
7: **compute** $\mathrm{SINR}_{\mathbf{r_c}}(\mathbf{s}, \mathbf{f} | \mathbf{a^k} \cup \mathbf{a}^\star)$
8: $\delta_{r_c}^k(s, f) \leftarrow$ **sinr_to_delta**$(\mathrm{SINR}_{r_c}^k(s, f))$ (Alg. 4.1)
9: **compute** $\chi_{r_c}^k(s, f | \mathbf{a}^k \cup \mathbf{a}^\star)$ (Alg. 4.2)
10: **compute** $pollution(s, f, r | \mathbf{a}^k \cup \mathbf{a}^\star)$ (Eq. 5.4)
11: $\mathbf{W}^k \leftarrow \boldsymbol{\chi}./\mathbf{pollution}$
12: **return** $\mathbf{W}^k$

Algorithm 5.3 Selecting the allocation triplets

Require: $\mathbf{a}^k$, $\mathbf{t}_a^k$, $\mathcal{F}_a^k$, $\mathbf{W}^k$, $\mathcal{S}$, $\mathcal{R}$, **PCell**, **SCell**
1: **for all** $f \in \mathcal{F}_a^k$ **do**
2: $s^\star, r^\star \leftarrow \arg\max_{s,r} \mathbf{W}^k$
3: $\mathbf{a}^\star \leftarrow (s^\star, f, r^\star)$
4: $t^\star \leftarrow \sum \chi_r(s, f)$, $\forall (s, f, r) \in \mathbf{a}^k \cup \mathbf{a}^\star$
5: **if** $t^\star > \mathbf{t}_a^k$ **then**
6: $\mathbf{a}^k \leftarrow \mathbf{a} \cup \mathbf{a}^\star$
7: **if isempty PCell**(u) **then**
8: **PCell**$(u) \leftarrow s^\star$
9: **else if** $u \in \mathcal{U}_{CA}$ **and** $s^\star \notin$ **SCell**(u) **then**
10: **SCell**$(u) \leftarrow$ **SCell**$(u) \cup s^\star$
11: **return** $\mathbf{a}^k$

Next a weight matrix $\mathbf{W}^k$ is calculated using Alg. 5.2. The weight matrix $\mathbf{W}^k$ stores a weight value for every potential combination of (s, f, r). Using $\mathbf{W}^k$ and Algorithm 3, we select the resource allocation triplets that will finally be admitted to the resource allocation strategy $\mathbf{a}^k$. After choosing the allocation triplets, the temporary status values $\mathbf{t}_a^k$ are updated using:

$$\mathbf{t}_a^k(f) = \sum_{s \in \mathcal{S}} \sum_{r \in \mathcal{R}} \chi_{r_c}^k(s, f), \forall f \in \mathcal{F}_a^k \tag{5.3}$$

and then used to recompute the *urgency* values in the next iteration.

During each iteration of the main loop, triplets are added to the allocation strategy, thus gradually constructing the whole strategy. Once the loop is finished, we obtain the final allocation strategy $\mathbf{a}^k$. With the final allocation decision at hand, the status values $\mathbf{t}^{k+1}$ for the next time step can be calculated using Eq. (5.1). These values are used as an input to the algorithm in the next time step, together with fresh traffic requests.

5.2.2 Calculating the weight matrix

As already mentioned, Alg. 5.2 is used to calculate a weight value for each potential resource allocation decision, i.e., potential combination of (s, f, r). To do so, for each active flow $f \in \mathcal{F}_a^k$, first we identify the set of potential serving cells $\mathcal{S}_a^k(f)$ using Alg. 5.4 or Alg. 5.5, as explained in more detail in Sec. 5.2.4. This set includes all serving cells that are eligible to serve the particular flow.

Next, for each potential serving cell s, and each available RB r at s, we calculate the potential amount of data that can be transferred over r, $\chi_r^k(s, f)$, using Alg. 4.1 and Alg. 4.2. For the same combination of (s, r), we also calculate the *pollution* value, to account, as already mentioned, for the interference each potential allocation triplet can cause. The *pollution* value is defined as the sum of interference caused to all other active UEs in the network, if the particular combination under consideration, i.e., (s, f, r), is to be admitted to the final strategy. Specifically,

$$pollution(s, f, r) = \sum I_r^k(u_{f_a}) \tag{5.4}$$

where u_{f_a} are all the UEs associated with other active traffic flows f_a, with $f_a \neq f$.

After both loops are completed, the *pollution* values are normalized and then the weight matrix $\mathbf{W}^k$ is constructed by obtaining the weight value for each (s, f, r) combination as the ratio $\chi_r^k(s, f)/pollution(s, f, r)$.

5.2.3 Selecting the allocation triplets

Finally, Alg. 5.3 selects the allocation triplets that will be admitted to the final allocation strategy $\mathbf{a}^k$. To do so, for each active flow, it selects the $(s^\star, r^\star)$ combination that maximises its weight value. Then, it evaluates whether the addition of

the identified triplet $(s^\star, f, r^\star)$ improves the overall amount of data transferred over the network, i.e., whether it increases $\sum_{\mathcal{S}} \sum_{\mathcal{F}_a^k} \sum_{\mathcal{R}} \chi_r^k(s, f)$. If it does, the selected triplet is permanently included in the allocation strategy $\mathbf{a}^k$.

In case the **PCell** and **SCell** sets are not already provided, the selected serving cell $s^\star$ will be added to **PCell**, if empty. If **PCell** is already configured, the algorithm proceeds to populate the **SCell** set by adding $s^\star$.

5.2.4 Defining the set of potential serving cells

The definition of the set of potential serving cell, at each time instant k, is an important step in the algorithm, which for each UE defines the set of serving cells over which the algorithm performs the search for resources. Depending on the implementation of the algorithm, this set can be predetermined by the network, if it chooses to preconfigure the **PCell** and **SCell** sets for each UE. In that case however, the carrier selection option of the algorithm is disabled. In general, however we consider that such sets are defined within the algorithm at every subframe k, or at least, in the case of **PCell**, reset with a certain update period, which can be longer than the scheduling frequency. The procedure also depends on the supported implementation of CA. If the implementation of CA is *single-flow* then Alg. 5.4 is used, while if *multi-flow* is supported Alg. 5.5 is used instead.

Algorithm 5.4 Defining the set of potential serving cells - Single-flow

Require: **PCell**(u),**SCell**(u)
1: **if isempty PCell(u) then**
2: **for all** $s \in \mathcal{S}$ **do**
3: **if** $10 \log \frac{P(s,u)}{A_c(s,u)}) \geq P_{thr}$ **then**
4: $\mathcal{S}_a^k(u) \leftarrow \mathcal{S}_a^k(u) \cup s$
5: **else**
6: **if** $u \in \mathcal{U}_{noca}$ **then**
7: $\mathcal{S}_a(u) \leftarrow$ **PCell**(u)
8: **else if** $u \in \mathcal{U}_{ca}$ **then**
9: $\mathcal{S}_a(u) \leftarrow$ **PCell**$(u) \cup$ **SCell**(u)
10: $b^\star \leftarrow b \in$ **PCell**(u)
11: **for all** $c \in \mathcal{C}_{b^\star} : c \notin$ **PCell**$(u) \wedge c \notin$ **SCell**(u) **do**
12: $s \leftarrow (b^\star, c)$
13: **if** $10 \log \frac{P(s,u)}{A(s,u)} \geq P_{thr}$ **then**
14: $\mathcal{S}_a^k(u) \leftarrow \mathcal{S}_a^k(u) \cup s$

Single-flow implementation

In the single-flow implementation, the set of potential serving cells, $\mathcal{S}_a^k(u)$ for UE u is defined as shown in Alg.5.4. First, the contents of **PCell** are checked. If **PCell**(u) is empty, then all serving cells satisfying the coverage criteria, will be included in $\mathcal{S}_a^k(u)$. If **PCell**(u) is not empty[1] and u is a legacy terminal, the $\mathcal{S}_a^k(u)$ is limited to the **PCell**(u) only as legacy UEs do not support aggregation of further cells. If u is a CA-enabled terminal, both **PCell**(u) and **SCell**(u) (if not empty), are included in $\mathcal{S}_a^k(u)$. In addition, the algorithm checks over the all the CCs available at the PoA $b^{\star 2}$ whether u is under their coverage area, i.e., the received power is over the threshold P_{thr}. If there are any such cells, they are included in $\mathcal{S}_a^k(u)$.

Multi-flow implementation

Algorithm 5.5 Defining the set of potential serving cells - Multi-flow

Require: **PCell**(u),**SCell**(u)

if isempty PCell(u) then
  for all $s \in \mathcal{S}$ **do**
    if $10 \log \frac{P(s,u)}{A(s,u)}) \geq P_{thr}$ **then**
      $\mathcal{S}_a^k(u) \leftarrow \mathcal{S}_a^k(u) \cup s$
else
  if $u \in \mathcal{U}_{noca}$ **then**
    $\mathcal{S}_a^k(u) \leftarrow$ **PCell**(u)
  else if $u \in \mathcal{U}_{ca}$ **then**
    $\mathcal{S}_a^k(u) \leftarrow$ **PCell**$(u) \cup$ **SCell**(u)
    for all $s \in \mathcal{S}$ **and** $c \in s$: $c \notin$ **PCell**$(\mathbf{u}) \wedge c \notin$ **SCell**$(\mathbf{u})$ **do**
      if $10 \log \frac{P(s,u)}{A(s,u)} \geq P_{thr}$ **then**
        $\mathcal{S}_a^k(u) \leftarrow \mathcal{S}_a^k(u) \cup s$

For the multi-flow CA, Alg. 5.5 is applied. The first few steps are identical: if **PCell**(u) is empty then all serving cells satisfying $10 \log \frac{P(s,u)}{A(s,u)}) \geq P_{thr}$ will be included in $\mathcal{S}_a^k(u)$. Also, if **PCell**(u) is not empty and u is non-CA UE, $\mathcal{S}_a^k(u)$ is limited to the **PCell**(u). For CA users on the other hand, in addition to any cells which are already in **PCell**(u) and **SCell**(u), the algorithm will search over all serving cells which are not associated to any of the carriers already in these two sets. All serving cells which satisfy the coverage criteria are admitted to $\mathcal{S}_a^k(u)$. Note that in this case, these serving cells can be associated to different PoAs, which implies that the CA-enabled UE can take advantage of all the carriers available in

[1] **PCell** will not be empty if it was either i) preconfigured by the network; or ii) populated in a previous iteration of the algorithm.

[2] $b^\star$ is the PoA controlling serving cell selected in **PCell**(u).

the network, and is not limited to only those carriers available at one PoA.

5.2.5 Solution complexity

In Alg.5.1 we iterate twice over the number of active flows $|\mathcal{F}_a^k|$. This number depends heavily on the traffic load, however in the worst case scenario it is $|\mathcal{U}|$. Within the first loop we have a nested loop which iterates over the potential servers $\mathcal{S}_a$ and RBs $\mathcal{R}$. While the size of $\mathcal{S}_a$ is small compared to $|\mathcal{U}|$, $|\mathcal{R}|$ may be significant. Note that as allocation decisions are added to the strategy the size of both these sets get smaller. Hence, we can say that the complexity of Alg.5.1 is $O(|\mathcal{U}||\mathcal{R}|)$. We noted that in order to ensure that all resources have been considered at least once for inclusion in the strategy, Alg.5.1 needs to be repeated several times. The number of iterations will depend on many factors. In the simplest case when we have only one active traffic flow, the maximum number of iterations will be $|\mathcal{R}|$. On the other hand, when there is a significant number of active traffic flows, the number of iterations on average will be $\frac{|\mathcal{B}||\mathcal{R}|}{|\mathcal{F}_a^k|}$. It is worth noting that after each iteration, the sets considered within Alg.5.1 shrink as resources become unavailable and the sets of potential serving cells are reduced to the sets **PCell** and **SCell**.

5.3 Performance Evaluation

5.3.1 Simulation scenario

We evaluate the performance of the proposed approach in a realistic scenario, with a two-layer dense network comprising of 57 macrocells and 228 microcells, shown in Fig. 4.4. Macrocells are controlled by three-sector base stations which are located at 19 sites. The macro PoAs inter-site distance is set to 500 m, while micro PoAs are deployed over the network area, so that there are 4 non-overlapping microcells per macrocell. A total of 3420 users are present in the area, with a higher density where microcells are deployed. 10 users are uniformly distributed within 50 m from each micro PoA. The rest of the users are uniformly distributed over the remaining network area. Thus, on average, there are 20 users per macrocell. Users move according to the cave-man model, with average speed of 1 m/s. In line with [69, 78], we assume a transmitting power of 43 dBm for macroBSs, and 30 dBm for micro PoAs, equally distributed over all RBs, and antenna height values of 25 m and 10 m, respectively. For the macro PoA antenna we further assume the antenna gain to be $G_T = 14$ dBi and the maximum attenuation 20 dB, while the micro PoA antennas are omnidirectional with 0 dBi gain.

We consider a set of three carriers at the following frequency bands: 2.6 GHz (CC1), 1.8 GHz (CC2) and 800 MHz (CC3). All carriers have a bandwidth of 10

MHz band, hence each CC has $|R_c| = 50$ RBs. All CCs are simultaneously available at all BSs, while carrier aggregation is implemented in the *multi-flow* mode.

Users request content for download from two categories: ebook and video. Video items have stricter constraints on delivery time and larger sizes, while e-books are smaller and have much longer deadlines. During the simulations, for the video category we assumed a 1 Mb item size and a deadline of 200 time steps, while for the ebook category we assumed a 500 Kb size and a 300 time step deadline.

We evaluated the performance of the interference-aware algorithm (labelled as IAW in the plots) in two possible implementations: the centralised implementation (IAW-C) and the distributed implementation (IAW-D). The former implies that the decision-making process is performed at one centralised controller (AC) for all the cells in the network, while in the latter the algorithm is performed independently for each site (consisting of three macrocells and 12 microcells).

5.3.2 Baseline solution

To assess the effectiveness of the algorithm we propose, we compare its performance to a baseline solution which implements eICIC, the de facto standard interference mitigation technique in LTE, and applies the Proportional Fair scheduling algorithm. This baseline solution is briefly described in this section and denoted as PF in the numerical results.

User association and carrier selection

To mitigate the inter-cell interference and balance the load between the different cell biasing is applied to artificially expand the range of the micro cells that have much lower transmitting powers and hence smaller coverage areas. Furthermore, since lower frequency carriers experience more favourable propagation conditions, biasing is also used to increase the coverage area of higher frequency carriers[3].

Consequently, the UEs associate with the cell from which they receive the strongest *biased* pilot signal. All non-CA users will select the best serving cell as their **PCell**. In the single flow implementation, if the UE has carrier aggregation capabilities it may add the other carriers available at that PoA as secondary cells, provided it is within their coverage area. In the multi flow implementation, the UE may connect to the PoA with the strongest signal on every available carrier, hence could be connected to several PoA simultaneously.

[3]Our simulation results showed that adding bias factors for different CCs improved the performance of the baseline solution

In both implementations, each user u selects the best serving cell as its primary cell:

$$\mathbf{PCell}(u) = \max_{s \in \mathcal{S}} \frac{B_b B_c P(s,u)}{A(s,u)}, \forall s \in \mathcal{S} \tag{5.5}$$

where B_b and B_c are the biasing factors for PoA b and carrier c, respectively, associated to serving cell s.

In the **single-flow** implementation, the set of secondary cells is defined by considering only those carriers available at the PoA b, controlling $\mathbf{PCell}(u)$. Therefore, all serving cells associated to PoA b, which fulfil the coverage criteria will be included in $\mathbf{SCell}(u)$:

$$\mathbf{SCell}(u) = \{s \in \mathcal{S} : b \in s \wedge c \notin s \wedge \frac{P(s,u)}{A(s,u)} \geq P_{thr}\} \tag{5.6}$$

In the **multi-flow** implementation on the other hand, a user u will select the best serving cells on all available carriers in the network, and is not necessarily limited to those available at the PoA controlling the selected primary cell. If we denote the carrier of the selected primary cell as c, than we can define the $\mathbf{SCell}$(u) set as':

$$\mathbf{SCell}(u) = \{\max_{s \in \mathcal{S}} \frac{B_b B'_c P(s,u)}{A(s,u)}, \forall c' \in \mathcal{C} \wedge c' \neq c\} \tag{5.7}$$

Namely, to populate its **SCell** set, each UE selects the PoA with the strongest signal on each component carrier available in the network, different from the component carrier of the **PCell**.

During numerical simulations the following biasing coefficients were used: 8 dB for microcells, 5 dB for CC2 and 8 dB for CC1. The selected values showed the best performance, after evaluating several configurations.

Resource scheduling

Proportional fair (PF) scheduling in is performed in a distributed manner, meaning that each serving cell makes individual decisions on the allocation of their resources. Note that the baseline solution does not require a central entity, like an area controller, which makes decisions based on higher layer information.

Therefore, each serving cell s will apply PF scheduling at each time step k to choose the flow it will allocate its resources to. For each RB, r, the flows will be chosen using the criteria:

$$\hat{f}_r^k = \max_{f \in \mathcal{F}} \frac{\chi_r^k(s,f)}{\bar{\chi}^k(s,u)}$$

where f is the flow initiated by user u associated to the cell under consideration, and $\bar{\chi}^k(s,u)$ is the historic average rate of user u on serving cell s.

5.3.3 Numerical results

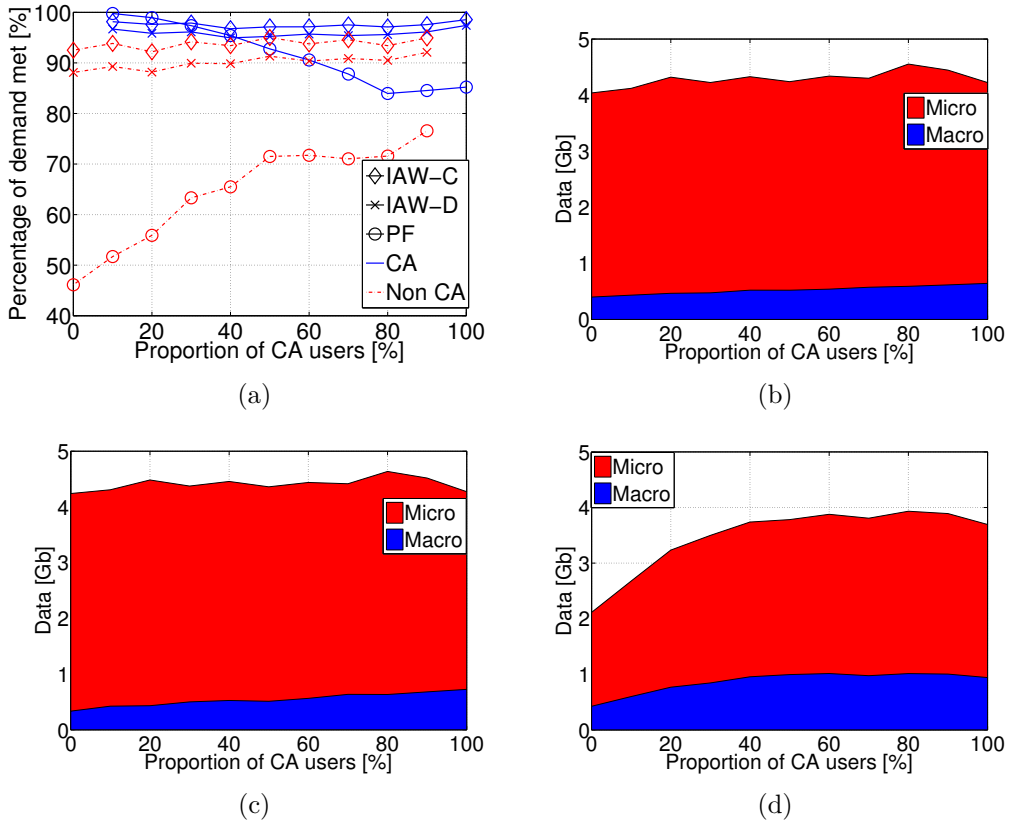


Figure 5.2. (a) Percentage of the demand met by the network for CA users (blue/solid line) and legacy users (red/dashed line); Overall amount of data downloaded during the simulation period via Macro (blue) and Micro (red) PoAs: (b) IAW-C; (c) IAW-D; (d) PF.

The simulation results for a number of performance metrics are shown in Figs. 5.2-5.7.

Meeting the traffic demand. Fig. 5.2(a) depicts the network ability to meet the data demand of the UEs during the simulation period, as the proportion of CA UEs in the network is increased. We see that IAW in both its implementations outperforms PF by a large margin in meeting the demand of the non-CA UEs (red), while consistently meeting over 90% of the demand of both types of UEs. PF in turn is able to meet all of CA UEs' need (blue, circle marker) when the presence of

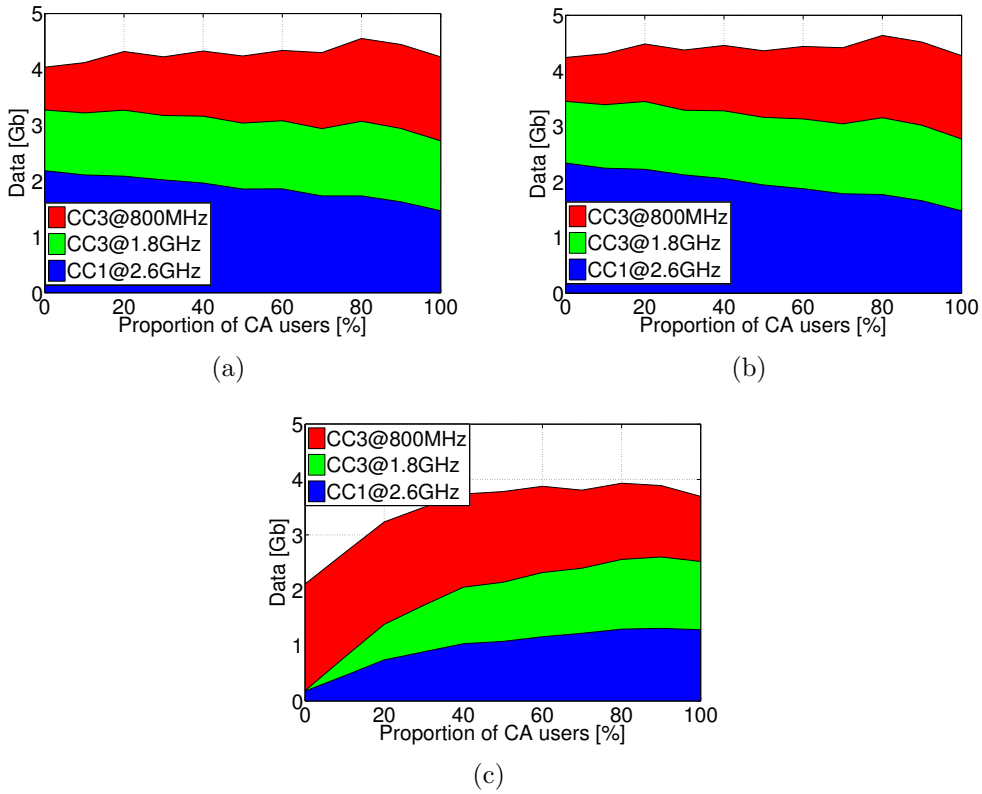


Figure 5.3. Overall amount of data downloaded during the simulation period over the different CCs: (a) IAW-C; (b) IAW-D; (c) PF.

such UEs in the network is small, and drops sharply when the proportion increases. In turn, the amount of data downloaded over the network for the duration of the simulation, shown in subfigures (b), (c) and (d) for centralised IAW, distributed IAW and PF respectively, is significantly higher for IAW implementations, especially when the number of CA UEs is small. It is also evident that micro PoAs are responsible for delivering the bulk of the data, and their contribution is even more pronounced in the IAW implementation. With IAW-C and IAW-D, they deliver 80%-90% of the overall transferred data, while with PF they deliver 75%-80%. Since, the traffic load is the same regardless of the number of CA UEs in the network, both IAW implementations deliver constant amount of data as the number of CA UEs grows. PF on the other hand struggles to deliver the data when number of CA UEs in the network is small. This indicates that IAW is more apt at using the available carriers to efficiently allocate non CA users, while PF fails to do so. It is clear that implementing biasing alone to offload traffic on the different CCs is not sufficient. Since, with PF, non-CA UEs connect to the serving cell with the strongest signal (after biasing), and CA UEs are allocated on all available CCs, this does not ensure

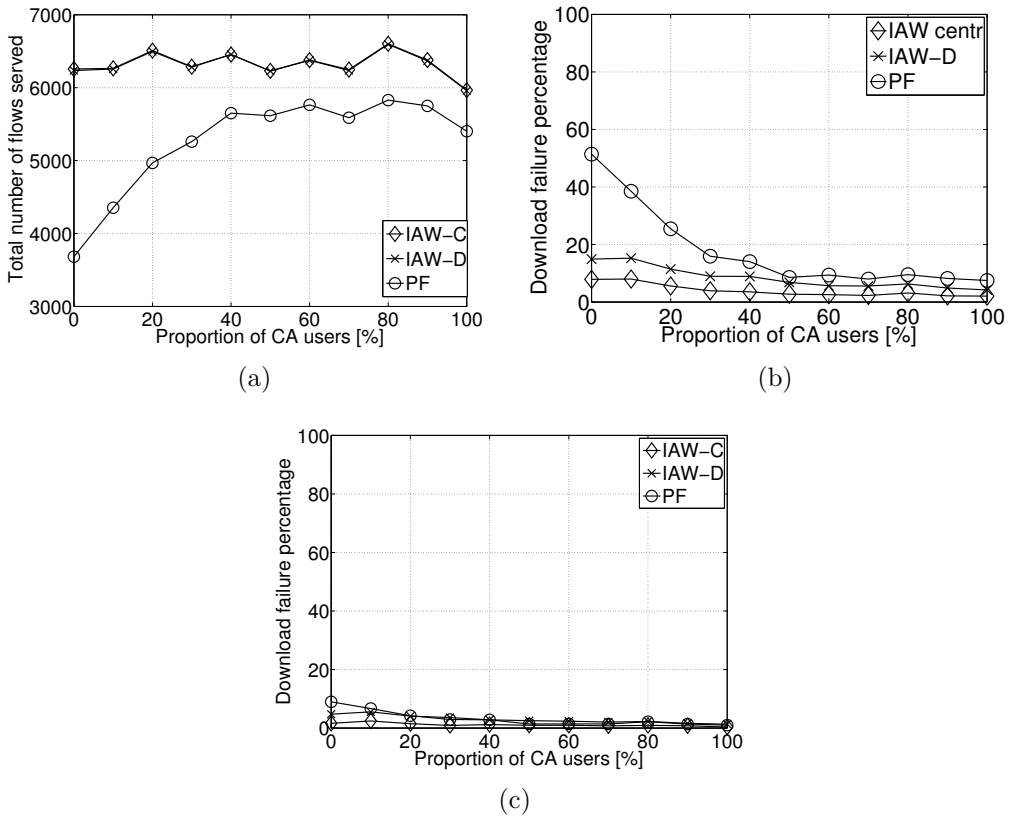


Figure 5.4. (a) Total number of traffic flows served; (b) failure rate of video downloads; (c) failure rate of ebook downloads.

a well-spread user distribution among the CCs. In fact most non-CA UEs will overcrowd the serving cell with the lowest frequency CC (CC 3), leaving the other available CCs under-utilised and interference-free for the CA UEs. IAW on the other hand is able to distribute the users more evenly among the CCs, thus is more prepared to handle the increase of CA UEs. Fig. 5.3 depicts the amount of data delivered over the different carriers using the proposed solution in (a) and (b) and the baseline solution in (c). It is clear from the figure the figure that the high frequency carrier CC (CC 1) carries most of the data for both IAW implementation, but the load in nevertheless relatively well distributed. PF on the other hand, for the same reasons listed above, is unable to offload the data from CC3 to the other CCs when the number of CA UEs is small, leaving the two higher frequency CCs heavily under-utilised.

In terms of timely delivery of requested content, and number of successfully served traffic flows, IAW performs significantly better than PF as shown in Fig. 5.4.

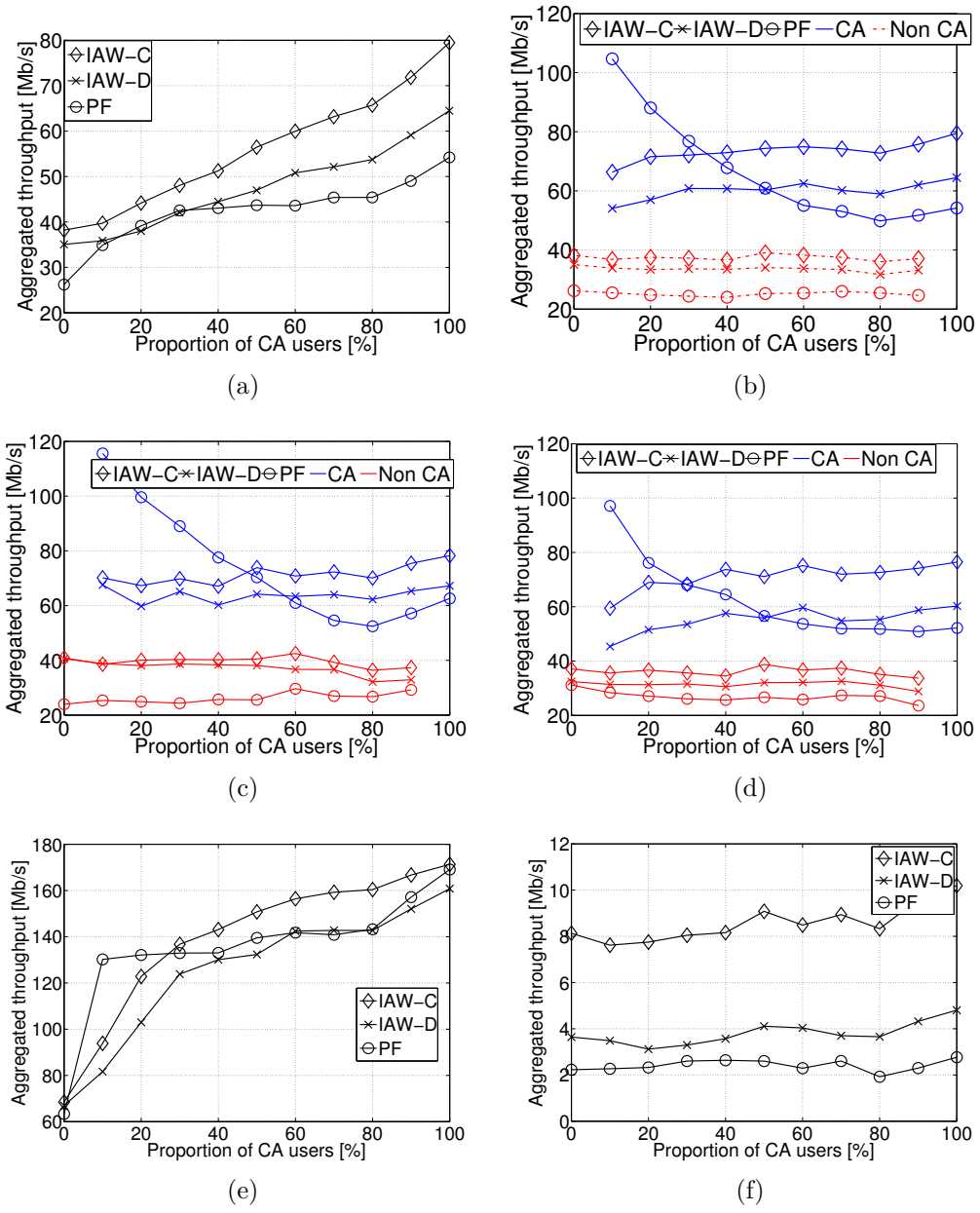


Figure 5.5. Average aggregated user throughput in Mb/s: (a) overall average of all UEs; (b) overall average for CA users (blue/solid line) and legacy users (red/dashed line); (c) average for inner CA users and legacy users; (d) average for edge CA and non-CA UEs; (e) average for top 5^{th} percentile UEs; (f) average for bottom 5^{th} percentile UEs.

In terms of flows served, the two implementations of IAW perform identically, however in terms of failed downloads (flows not completed within the deadline) the distributed implementation trails slightly behind. Indeed, for ebook downloads which

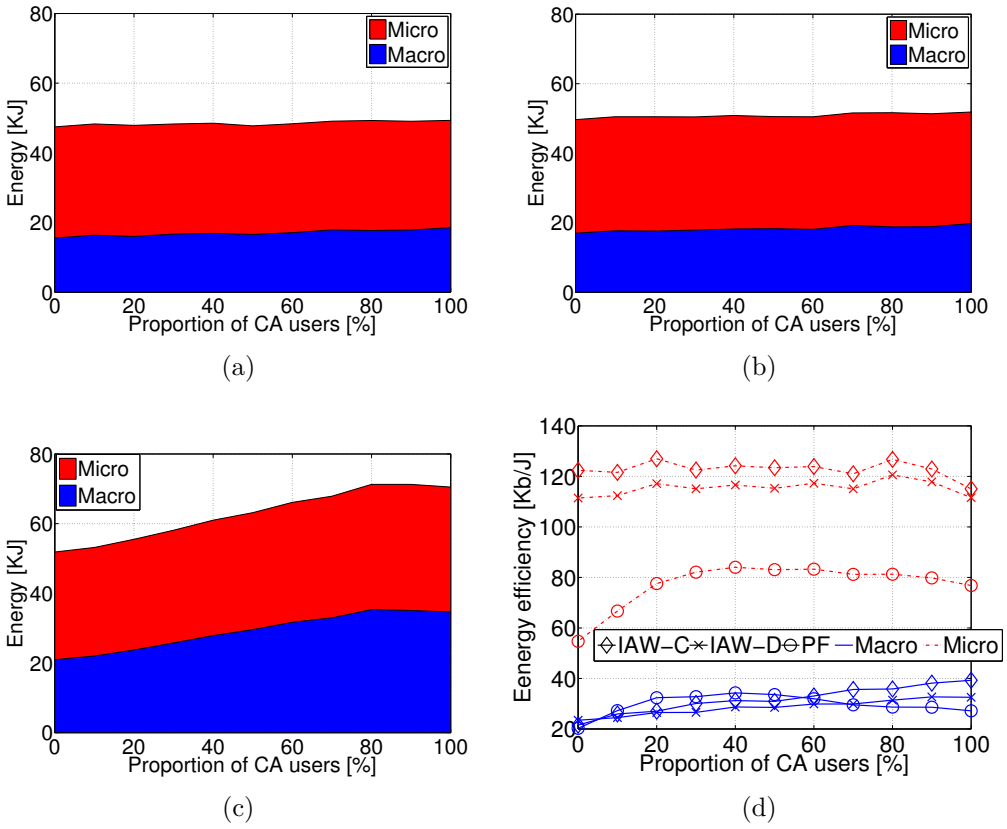


Figure 5.6. Overall amount of energy consumed during the simulation period by Macro (blue) and Micro (red) PoAs: (a) IAW-C; (b) IAW-D; (c) PF; and (d) energy efficiency measured in Kb/J.

are not as time sensitive as video content, the performance of all solutions is quite good, however for video content PF trails significantly behind. Overall, it is evident that IAW centralised is the superior implementation compared to the distributed one, but not by a significant margin.

Average aggregated user throughput The average aggregated user throughput is shown in Fig. 5.5. Specifically in Fig. 5.5(a) the overall average aggregated user throughput is shown, which as expected grows proportionally with the number of CA users in the network for all solutions. For small proportions of CA UEs in the network (less than 40%) PF performs seemingly as good as IAW distributed. However, when we look at the differentiated average aggregated user throughput for CA and non-CA UEs in Fig. 5.5(b), we observe that, again, IAW is able to ensure a constant average user throughput for both types of UEs. PF in contrast is able to ensure very high rates for CA UEs when they are a minority, and, again, it experiences a sharp drop when the number grows. The gain in performance for

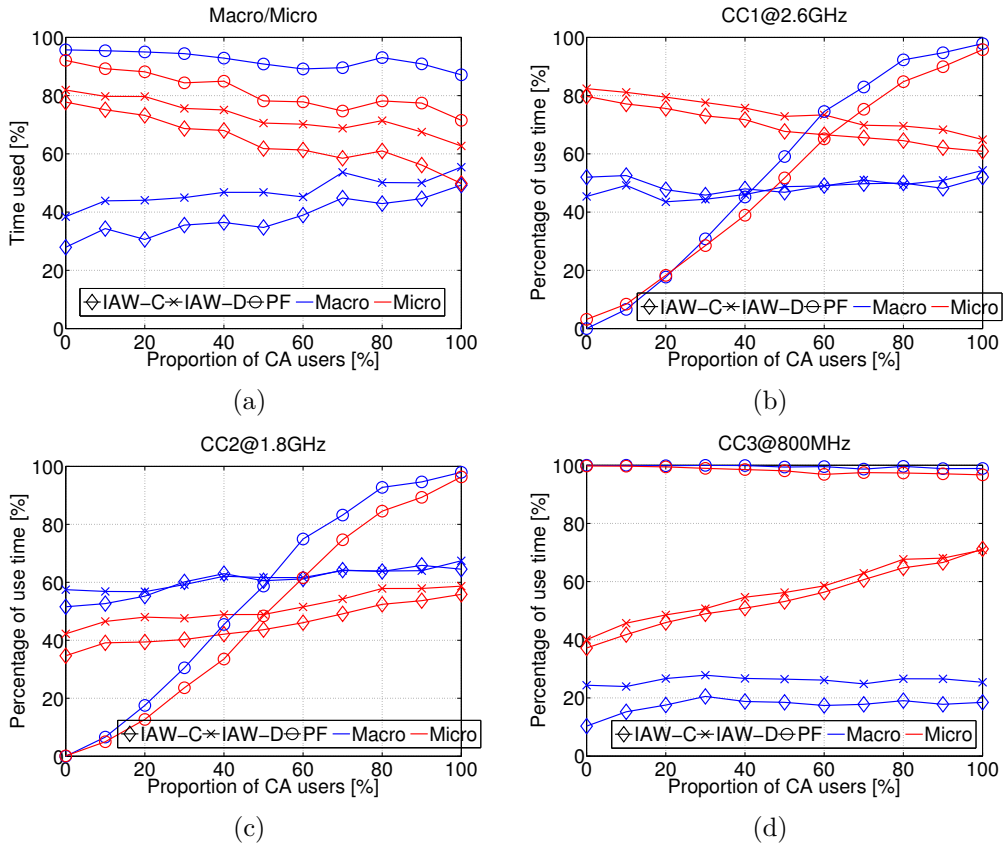


Figure 5.7. (a) Average percentage of use time for Macro (blue) and Micro (red) PoAs; Average percentage of CC use time by Macro (blue) and Micro (red) PoAs: (b) CC1@2.6 GHz; (c) CC2@1.8 GHz; (d) CC3@800 MHz.

IAW in all these metrics is mostly due to the algorithm's ability to efficiently match the serving cells and RBs with the receiving UEs, in an interference-aware manner which ensures higher SINR values at the destination, and hence higher data rates. In Fig. 5.5(c) and (d) we show the average throughput for inner and edge UEs respectively. While the above analysis holds both for inner and edge UEs, we note that the difference in performance between IAW-C and IAW-D is more pronounced for edge UEs. Indeed, if we look at the top 5% and bottom 5% UEs in Fig. 5.5(e) and (f), we note that PF performs better or as well as IAW-C for the top 5%; while for the bottom 5%, it is clear that IAW-C performs significantly better than both IAW-D and PF. While IAW-D's performance is very close to the performance of IAW-C in most metrics, it is clear that most of the loss in performance due to the distributed implementation is borne by UEs already facing bad conditions.

Energy consumption. The IAW approach manages to meet the demand consuming less energy than the baseline solution. Fig. 5.6(a), (b) and (c) depicts the

overall energy consumption incurred by the infrastructure nodes in the network, for the three different solutions. It is clear that for PF the energy consumption grows proportionally with the number of CA UEs in the network, while both IAW implementations have relatively consistent consumption curves. To gain more insight, the energy efficiency metric measured in Kb transmitted per Joule consumed, is plotted in Fig. 5.6(d). IAW considerably outperforms PF especially for micro PoAs (red), which is quite significant since we showed that micro PoAs in our network are responsible for delivering the larger bulk of the data. The energy efficiency for macro PoAs tends to be on the same level for both algorithms. However, for the same reason as discussed above, we note that IAW and PF exhibit opposite trends as the number of CA users increases.

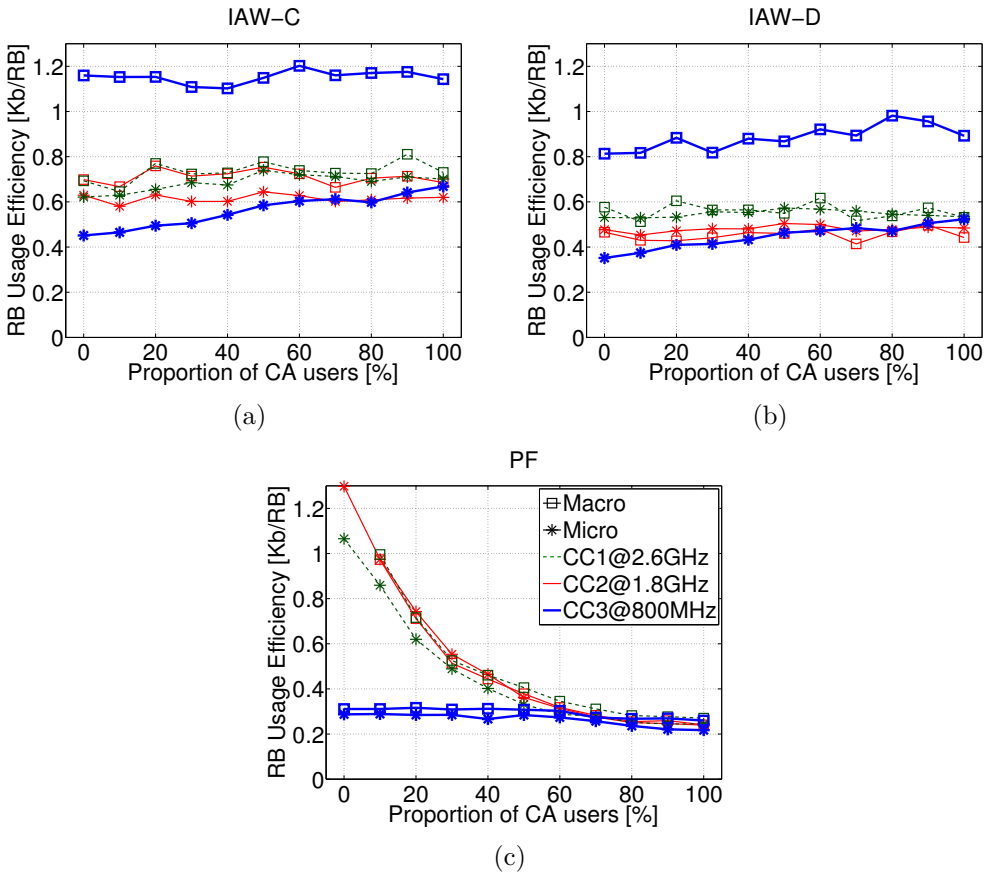


Figure 5.8. RB usage measured in Kb/RB for each CC: (a) IAW Centralized; (b) IAW Distributed; and (c) Proportional-Fair. The legend for all there subfigures is shown in the top right-hand corner of subfigure (c). We differentiate between different types of BSs using different markers, and between different CCs using different line styles and colours.

Some more insight on how the IAW algorithm works and consumes energy can be

gained by looking at Fig. 5.7 where we have plotted the average use time of Macro and Micro PoAs, as well as the average use times of the individual CCs. We see in Fig. 5.7 (a) that in both IAW implementations, Macro PoAs are used very little, in fact, on average a Macro PoA is used only 40%-55% of the time in the distributed implementation, and only 30-45% of the time in the centralized one, as the use time percentage increases with proportion of CA UEs. This behaviour partially explains why the energy consumed by Macro PoAs accounts for such a small portion of the total energy consumed with the IAW algorithm, and also why the energy efficiency of Macro PoA is so low. It is clear that most of the energy consumed by Macro PoAs is when they are in idle mode, powered on, but not transmitting any data. Micro PoAs, by contrast, are employed much more often when proportion of CA UEs in the network is small, while the use time is similar when all the UEs in the network are CA-enabled. Micro PoAs are used around 82%-60% of the time in the distributed mode and 77%-50% in the centralised mode, as the percentage tends to slightly decrease with the number of CA users. PF, on the other hand, uses both types of PoA almost all of the time. On average, when there are no CA users, a Macro PoA is used 95% of the time, while a Micro PoA is used roughly 90% of the time. These percentages drop to 90%, respectively to 70% when all of the users in the network are CA-enabled, however they are still significantly higher than in the IAW cases. The inefficiency of PF becomes more clear now: despite the higher usage values, it is not able to deliver nearly as much content as IAW is.

Similarly, IAW does not utilise all the CCs all the time either. As we can see in Fig. 5.7 (b), (c) and (d) both IAW implementations tend to use CC1 and CC2 significantly more than CC3 for Macro PoA. Indeed on average with IAW, a Macro PoA will use CC1 around 50% of the time (in both implementations), and CC2 around 60%-65% in the distributed mode, and 55%-65% in the centralised mode. CC3 on the other hand is used only 10%-20% in the centralised implementation and around 25% in the distributed one. For Micro PoAs on the other hand utilise CC1 significantly more, as they are in a better position to use its spatial reusability potential combined with their low transmit power. Indeed CC1 is used roughly 80% to 60% of the time, as the number of CA UEs increases, in both implementations. CC2 and CC3 on the other hand are used roughly 40% to 65% of the time as the number of CA UEs increases. To summarise, we can conclude that IAW in general prefers CC1 the most for both types of PoA when the number of CA UEs in the network is low. The choice is to be expected since CC1 has the smallest interference footprint as well. However, as the number of CA UEs in the network increases, we see that all CCs tend to be used roughly the same amount of time which is about 50% to 60%. With PF on the other hand, CC3 is used almost 100% of the time by both PoAs, while the use time of the other two CCs increases linearly with the proportion of the CA users in the network.

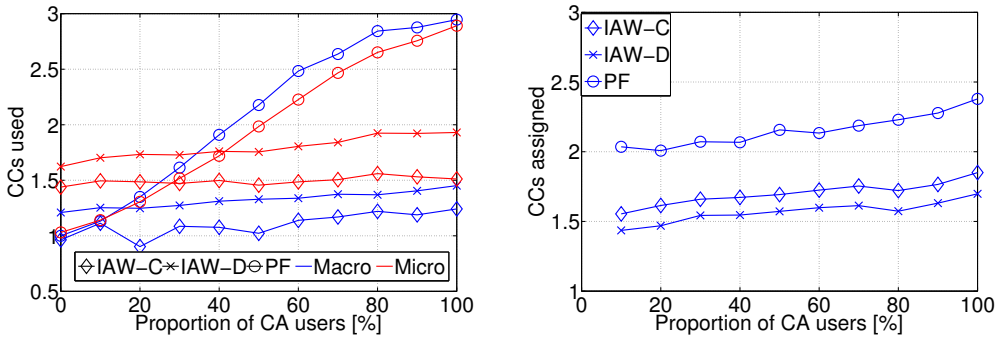
From these observations we reach the conclusion that to improve network performance, it is necessary to allow for PoAs and CCs alike not to be used at all times. Indeed, it is clear that the efficiency of IAW stems from its ability to choose, in an intelligent and interference-aware manner, which PoAs should transmit on which CCs and which not, at a given time step, to ensure best use of resources. That said, this clearly indicates that there is room for improving energy efficiency by devising appropriate sleeping patterns for the different PoAs and CCs, so that they consume even less energy when they are not used. Another possibility, which provides even more flexibility, is to allow PoAs to reduce the transmit powers, and dynamically adjust the transmit power levels of the different CCs so that optimal network configurations are achieved. Indeed, the latter option is exhaustively researched in Chapter 6.

RB usage efficiency Finally, in Fig. 5.8 we look at the RB usage efficiency, which is measured in Kb transmitted per allocated RB during one subframe. We note that IAW-C, Fig. 5.8(a), is able to transmit the largest amount of data per allocated RB in CC3, both for macro and micro PoAs. Note that this ratio is significantly higher for macro PoAs than for micro PoAs, in the case of IAW-C and IAW-D. That is to be expected, since macro PoAs have higher transmit power and are used less (hence suffer less interference) than micro PoAs, especially in combination with the low frequency CC3.

Under both IAW-C and IAW-D, the RB usage efficiency improves for CC3 as the percentage of CA users increases. This is due to the fact that more CA-enabled UEs imply a higher flexibility in assigning the different CCs to user traffic flows. This is consistent with the fact that the RB usage efficiency of CC1 and CC2 is practically independent of the number of CA users. On the contrary, Fig. 5.8(c), shows that PF again exhibits a sharp drop in RB usage efficiency for CC1 and CC2, with the increase of CA users. Indeed, as the number of CA UEs grows, PF is unable to smartly allocate traffic flows over the various CCs, hence CC1 and CC2 tend to become over-crowded.

In general, the gains in RB usage efficiency can also be attributed to IAW's ability to use the CCs more efficiently. In Fig. 5.9 we see that with IAW a Macro PoA will on average use less than 1.5 CCs simultaneously and this average grows only slightly with the number of CA users in the network. Micro PoAs on the other hand use on average 1.5 to 2 CCs, the average slightly growing with the number of CA users in the network. Average number of CCs used with PF is quite similar for Macro PoAs and Micro PoAs and depends heavily on the percentage of CA users in the network. When there are no CA users in the network the average number of CCs used is 1, meaning that PoAs almost never take advantage of the additional carriers at hand. The average grows almost in a linear fashion with the number of CA users in the network. In Fig. 5.9 we have plotted the average number of CCs assigned to CA users. We note that with IAW this average tends to be somewhere

between 1.5 and 2. With PF it is always higher than 2. This indicates that with IAW there is room for introducing power saving policies at the UE terminal, by letting the UE know that it does not need to listen to all available CCs. This could be a topic for further investigation.



(a) (b)

Figure 5.9. (a) Average number of CCs used simultaneously by Macro (blue) and Micro (red) PoAs; (b) average number of CCs assigned to CA UEs.

Load distribution. Fig. 5.10 shows the average number of UEs associated to Macro and Micro PoAs and the individual CCs. We note that both implementations of IAW tend to distribute the traffic load similarly over the different PoAs and CCs. Namely, while CC1 and CC2 are clearly favoured, IAW tends to balance the load between all CCs, and the number of associated UEs on all carriers increases, as expected, with the number of CA UEs in the network. The baseline solution, on the other hand, despite applying CRE, fails at taking advantage of the availability of multiple CCs, especially when there are more legacy UEs in the network, as non-CA UEs are consistently associated to CC3, and only UEs with CA capabilities are assigned on the other two CCs. In addition, with PF, there is also load imbalance between Macro and Micro PoAs in the network. It is evident that IAW's ability to fairly balance the load, is one of the key reasons it performs significantly better than PF in most performance metrics.

5.4 Conclusions

In this chapter we investigated the resource allocation problem in a two-tier, dense network, with carrier aggregation support. We proposed an interference-aware heuristic solution to efficiently accommodate the traffic requested by UEs in the network. For each traffic request, our algorithm matches the requesting UE with the most appropriate serving cell, carrier and set of RBs. Besides interference, our algorithm also takes into account traffic characteristics to ensure that users receive

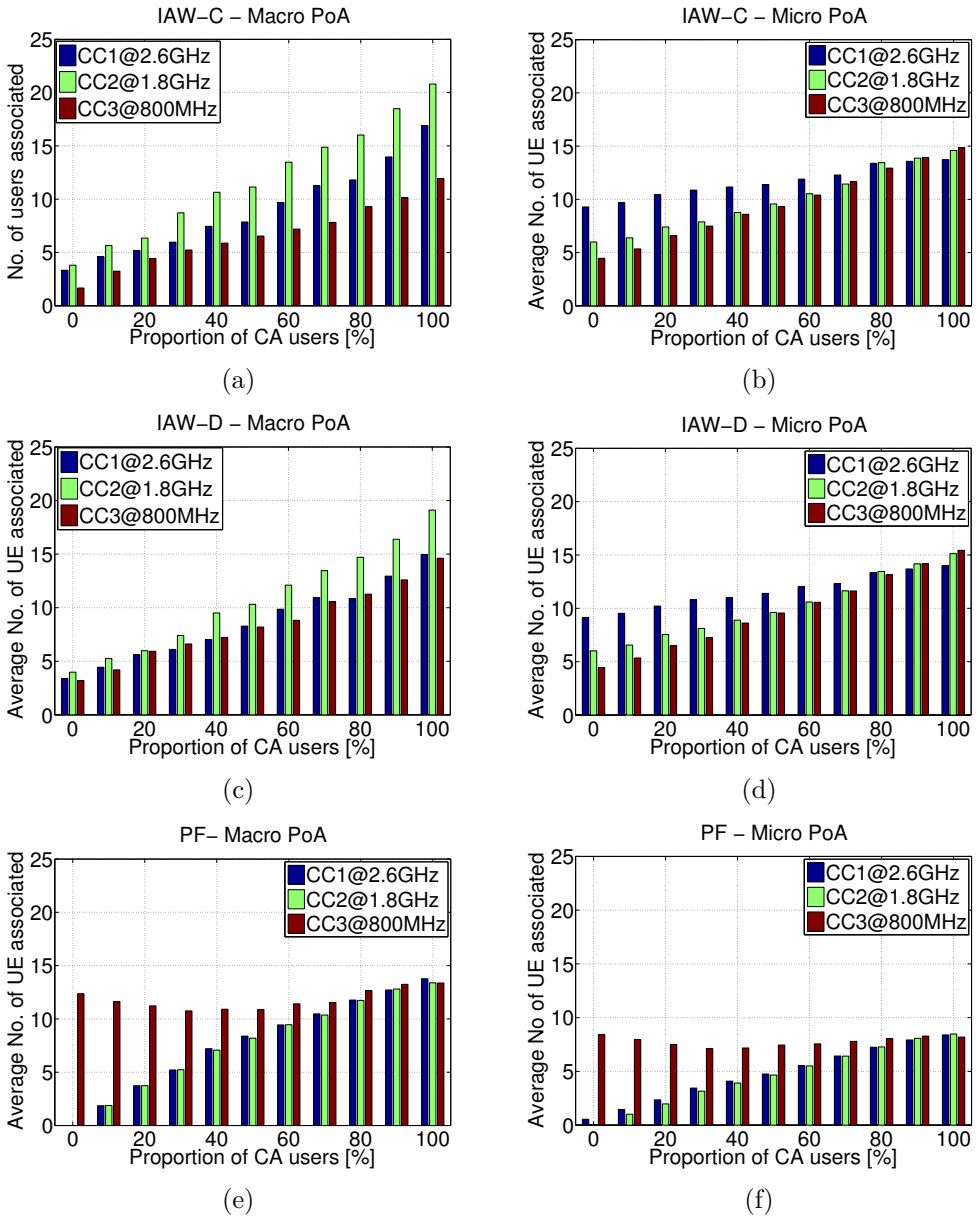


Figure 5.10. Average number of UEs associated per PoA on the different carriers: (a) IAW Centralized - Macro PoA; (b) IAW Centralized - Macro PoA; (c) IAW Distributed -Macro PoA; (d) AW Distributed - Micro PoA; and (e) Proportional-Fair Macro PoA; (f) Proportional-Fair Micro PoA.

their content within the prescribed deadlines. We evaluated two possible implementations of the proposed algorithm, centralised and distributed, and compared its performance with a baseline solution which couples eICIC and Proportional-Fair scheduling, a combination widely used in today's networks. Simulation results show

that our scheme outperforms the baseline solution, in terms of overall number of flows served, average aggregated user throughput, energy and spectrum efficiency, and is able to better meet the traffic demands of UEs with different capabilities. The results show that although, as expected, the centralised implementation performs better in most metrics, the distributed implementations performs almost as well, and significantly better than the baseline.

Part III

Downlink transmit power management

Chapter 6

Game-theoretic approach to carrier downlink power setting in dense networks

As we already mentioned, CCs belonging to different frequency bands may have very different coverage areas and impact in terms of interference, due to both their different transmit power level and their propagation characteristics. In this chapter we lay down a proposal that aims at leveraging this diversity to mitigate the interference in the network, by adjusting the carrier downlink transmit power. To this end we formulate a downlink power setting problem for the different available carriers and use game theory to design a distributed algorithm that lets cells dynamically adjust different transmit powers for the different carriers.

As envisioned in LTE-A systems and unlike most of previous work, we consider that each CC at each Point of Access (PoA) has an independent power budget, and that PoAs can choose the transmit power on each carrier from a discrete set of values. Therefore, our goal is to adequately choose a power level from a range of choices to ensure optimal network performance. It is easy to see that the complexity of the problem increases exponentially with the number of cells, CCs and the granularity of the power levels available to the PoAs. In addition, if one of the objectives is to maximise the network throughput, the problem becomes non linear since transmission data rates depend on the signal-to-interference-plus-noise ratio (SINR) experienced by the users. It follows that an optimal solution requiring a centralised approach would be both unfeasible and unrealistic, given the large number of cells in the network.

We therefore study the above problem through the lens of game theory, which is an excellent mathematical tool to obtain a multi-objective, distributed solution in a scenario with entities (PoAs) sharing the same pool of resources (available CCs). We model each group of PoAs in the coverage area of a macrocell as a

team so that we can capture both (i) cooperation between the macrocell and the small cells with overlapping coverage areas, and (ii) the competitive interests of different macrocells. The framework we provide however allows for straightforward extension to teams that include several macrocells. We prove that the game we model belongs to the class of *pseudo-potential* games, which are known to admit pure Nash Equilibria (NE) [50]. This allows us to propose a distributed algorithm based on best-reply dynamics that enables the network to dynamically reach an NE representing the preferred solution in terms of throughput, user coverage and power consumption. As shown by simulation results, our scheme outperforms fixed transmit power strategies, even when advanced interference mitigation techniques such as eICIC are employed.

The proposed solution greatly improves network performance by reducing interference and power consumption, while ensuring coverage for as many users as possible. We compare our scheme to other interference mitigation techniques, in a realistic large-scale scenario. Numerical results show that our solution outperforms the existing schemes in terms of user throughput, energy and spectral efficiency.

6.1 System model and assumptions

We consider a CA-enabled two-tier dense network composed of macro and microcells, each controlled by different types of PoAs. The network serves a large number of CA-enabled user equipments (UEs), which may move at low-speed (pedestrian) or high-speed (vehicles).

To make the problem tractable, we partition the entire network area into a set of tiles, or zones, denoted by $\mathcal{Z}$. From the perspective of downlink power setting, the propagation conditions within a tile from a specific PoA represent averages of the conditions experienced by the UEs within the tile. Note that the tile size can be arbitrarily set, and represents a trade-off between complexity and realism. We will assume for ease of presentation that tiles (i.e., the UEs therein) are associated with the strongest received reference power, although the extension to other, dynamic association schemes as well as to the case where a tile is served by multiple PoAs can be easily obtained. For simplicity, the user equipments (UEs) in the network area are all assumed to be CA enabled. Note, however, that the extension to a higher number of tiers as well as to the case where there is a mix of CA-enabled and non CA-enabled UEs is straightforward. All cells share the same radio resources. In particular, a comprehensive set of component carriers (CC), indicated by $\mathcal{C}$, is available simultaneously at all PoAs (PoAs having at their disposal a subset of CCs is a sub-case of this scenario). Each CC is defined by a central frequency and a certain bandwidth. The central frequency affects the carrier's coverage area, as the propagation conditions deteriorate greatly with increasing frequency. The level

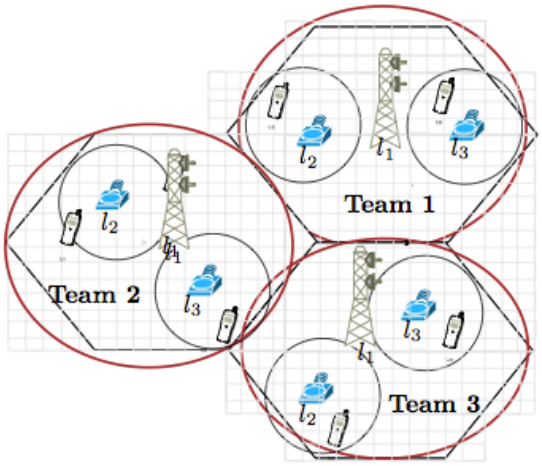


Figure 6.1. Network model and teams. Team locations are denoted by l_1, l_2, l_3. Solid red lines represent team boundaries, while black solid lines represent coverage areas. Tiles are represented by grey squares.

of transmit power irradiated by each PoA on the available CCs can be updated periodically depending on the traffic and propagation conditions in the served tiles, or it can be triggered by changes in UE distribution or traffic demand. The update time interval, however, is expected to be substantially longer than an RB allocation period, e.g., order of hundreds of subframes. The PoAs can choose from a discrete set of available power levels, including 0 that corresponds to switching off the CC. The possible power values are expressed as fractions of the maximum transmit power, i.e., $\boldsymbol{P} = \{0.1, 0.2, ...,1\}$. Note that the maximum transmit power in general varies depending on the type of PoA. As noted before, each CC at each PoA has an independent power budget.

In order to determine the downlink power setting, PoAs can leverage the feedback they receive from their users on the channel quality that UEs experience. Also, we assume that PoAs within the same macrocell are interconnected, or at least connected to the macro PoA, via, e.g., optical fiber connections, which allows for swift communication between them. As a result, we assume that it is possible for the macro PoA and the corresponding micro PoAs to cooperate and exchange information in order to reach common decisions. This is a reasonable assumption since it is expected that the architecture foreseen for future networks will allow PoAs that are geographically close to share a common baseband [3]. Furthermore, we assume that neighbouring macro PoAs can also communicate with each other.

6.2 Game theory framework

As mentioned before, game theory is an excellent tool in addressing complex problems, for which an optimal centralised solution might not be feasible. Considering that the complexity of carrier power setting increases exponentially with the number of PoAs, CCs and the granularity of the transmit power levels, we adopt a game theoretic approach to the problem in order to derive low-complexity, distributed solutions that are applicable in practice.

Table 6.1: List of symbols

'Symbol	Description	Symbol	Description
$\mathcal{Z}$	Set of tiles z covering the entire network coverage area	$\mathcal{C}$	Comprehensive set of carriers $\{c_1, \cdots, c_C\}$
$\boldsymbol{P}$	Set of strategy levels available to players, expressed in fractions of the maximum transmit power	$\mathcal{T}$	Set of teams $\{t_1, \cdots, t_T\}$
$\mathcal{L}_t$	Set of locations (players) $\{l_1, \cdots, l_L\}$ forming team t	$\mathcal{Z}_t$, $\mathcal{Z}_l$	Set of tiles under the coverage area of a particular team t and location l, respectively
E_t, E_l, E_c	The total number of UEs under the coverage area of team t, location l and tile z, respectively	α, β	Sigmoid utility function tuneable parameters
$\boldsymbol{s}^t$	An $L \times C$ matrix indicating the strategy selected by team t	$\boldsymbol{s}^{-t}$	The strategies selected by all teams other than t
$\boldsymbol{I}^t$	The interference matrix of team t	$a_{l,z,c}$	The factor of signal attenuation during transmission from location l on carrier c to the UEs in tile z
$\gamma^t_{z,c}$	The SINR at tile z, when served by location l in team t	$u^t(\boldsymbol{s}^t, \boldsymbol{s}^{-t})$	Utility function of team t
$\xi^t_{l,c}$	rice per received power unit for location l and carrier c	δ	Unit price paid for each unserved user

Table 6.1: List of symbols

‘Symbol	Description	Symbol	Description
e_t	The fraction of UEs within the team area that experience SINR levels below a certain threshold	$\pi^t(\boldsymbol{s}^t, \boldsymbol{s}^{-t})$	The cost function of team t
$w^t(\boldsymbol{s}^t, \boldsymbol{s}^{-t})$	Payoff function of team t	$\mathcal{S}$	The comprehensive set of strategies available to the teams
$\mathcal{W}$	Set of payoff functions	γ_{min}	The minimum SINR threshold indicating coverage

Specifically, we formulate the problem of power setting in dense CA-enabled networks as a competitive game between *teams* of PoAs (see Fig. 6.1), where each team wants to maximize its own payoff. Indeed, given the network architecture at hand, PoAs within an overlapping geographical area have the common objective to provide the UEs under their coverage high data throughput. Thus, they may choose to cooperate with each other in order to improve their individual payoffs as well as contribute to the “social welfare" of the team. Cooperation among such PoAs is beneficial especially since the inter-tier interference is most significant within the cell. Although increasing the transmit power of one PoA may increase the SINR that its UEs experience, such increase hurts the UEs being served by other PoAs since all PoAs share the same frequency spectrum. It follows that teams will compete between each other for the same resources, each aiming at maximising their own benefits. The game we model and its analysis are detailed below. We note that the formulation can be easily extended to accommodate various team configurations and clusters of teams, each controlled by a central controller.

6.2.1 Game model

Let $\mathcal{T} = \{t_1, ..., t_T\}$ be the set of teams in our network, where T is the number of teams. Each team consists of a macro PoA and the micro PoAs whose coverage areas geographically overlap with that of the macrocell. Note that not only can team players exchange information between each other, but we can also assume that the macro PoA plays the role of team leader, i.e., it makes the decisions for all team members in a way that maximizes the overall team benefits.

To generalise the formulation further, we will refer to the PoAs forming a team t as the *locations* of the team, $\mathcal{L}_t = \{l_1, l_2, ..., l_L\}$ where, for simplicity of notation, the number of locations within a team is assumed to be constant and equal to L. Such

a generalisation is particularly useful since the interference caused within the team depends also on the relative position between the different players. We indicate the set of tiles under the coverage area of a particular location l by $\mathcal{Z}_l$, and their union, denoting the comprehensive set of tiles of the team, by $\mathcal{Z}_t$. Also, let us denote by E_l the number of UEs under the coverage of location l, by E_z the number of UEs in tile z, and by $E_t = \sum_{l \in \mathcal{L}_t} E_l$ the total number of UEs served by the team.

Each team, comprising a set of locations (PoAs located at different positions within the macrocell), has to decide which transmit power level to use (out of the possible values in $\boldsymbol{P}$), at each one of those locations and for each of the available carriers $\mathcal{C} = \{c_1, c_2, ..., c_C\}$. It follows that the strategy selected by a team t, $\boldsymbol{s^t}$, is an $L \times C$ matrix, where each (l, c) entry indicates the power level set at location l on carrier c.

We now provide the definitions for the team utility and payoff, which are used in game theory to model the objectives of the players when choosing their strategy. Since network throughput is an important performance metric, it is natural that the utility of each team is defined as a function of the data rates it can serve to its UEs. The data rate a UE obtains is closely linked to the SINR it experiences, which depends on the transmit power chosen by the serving location (PoA), the CC that is used and the transmit power levels chosen by neighbouring locations. Assuming that all UEs within the same tile experience the same amount of interference, for each team we can first define an interference matrix of size $|\mathcal{Z}_t| \times C$, denoted by $\boldsymbol{I^t}$. Each entry in the matrix indicates the interference experienced by UEs in tile z on carrier c, which is caused by other teams:

$$I^t_{z,c}(\boldsymbol{s^{-t}}) = \sum_{t' \in \mathcal{T} \wedge t' \neq t} \sum_{l' \in \mathcal{L}_{t'}} s^{t'}_{l',c} a_{l',z,c} \tag{6.1}$$

where $\boldsymbol{s^{-t}}$ represents the strategies adopted by all teams other than t, $s^{t'}_{l',c}$ is the power level (the strategy) of team t' for location l' on carrier c and $a_{l',z,c}$ is the factor of the attenuation ($0 \leq a_{l',z,c} \leq 1$) experienced by the signal transmitted from location l' on c and received by the UEs in tile z. The attenuation values are pre-calculated using the urban propagation models specified in [69].

The SINR at tile z, when served by location l in team t, is:

$$\gamma^t_{z,c} = \frac{s^t_{l,c} a_{l,z,c}}{N + \sum_{l' \in \mathcal{L}_t \wedge l' \neq l} a_{l',z,c} s^t_{l',c} + I^t_{z,c}} \tag{6.2}$$

where N represents the average noise power level. Note that, besides N and $I^t_{z,c}$, we have an additional term at the denominator, which stands for the intra-team interference and indicates the sum of all power received from the locations within the same team, other than location l.

Then the utility of each team can be defined as a function of the individual tiles' SINR values. In particular, the sigmoid-like function has been often used for this

purpose in uplink power control [51]. We note that this function is suited to capture also the utility in downlink power setting, as it has features that closely resemble the realistic relationship between the SINR and the data rate. We therefore adopt the sigmoid function proposed in [51], as the utility function of each (tile, carrier) duplet in the team, and write the team utility as:

$$u^t(\boldsymbol{s^t}, \boldsymbol{s^{-t}}) = \sum_{l \in \mathcal{L}_t} \sum_{z \in \mathcal{Z}_l} \sum_{c \in \mathcal{C}} \frac{E_z}{E_t \left(1 + e^{-\alpha(\gamma^t_{z,c} - \beta)}\right)}. \tag{6.3}$$

The sigmoid function in Eq. (6.3) has two tuneable parameters, α, which controls the steepness of the function, and β, which controls its centre. They can be tweaked to best meet the scenario of interest. In particular, the higher the α, the closer the function resembles a step function, i.e., the utility becomes more discontinuous with the increase of the SINR. The higher the β, the larger the SINR for which a tile obtains a positive utility. Also, the individual utility of each tile z in team t is weighted by the fraction of UEs covered by the team in the tile (E_z/E_t) so as to give more weight to more populated tiles. This enables us to account for the user spatial distribution whenever this is not uniform over the network area.

Next, we introduce a cost function to account for the interference and its detrimental effect, as well as for fairness in the service level to users. We define a first cost component that aims at penalising players who choose high power strategies, as: $\sum_{l \in \mathcal{L}_t} \sum_{c \in \mathcal{C}} \xi^t_{l,c} \bar{a}_{l,c} s^t_{l,c}$ where $\bar{a}_{l,c}$ is the link quality on carrier c averaged over all tiles served by location l, and $\xi^t_{l,c}$ is the price per received power unit for location l and carrier c. This cost component increases with the increase in the chosen level of transmit power, however it also accounts for the propagation conditions of the users served by the location. In other words, locations that have to serve UEs experiencing poor channel quality will incur a lower cost, which ensures some level of fairness. The way the unit price, ξ, should be set is investigated in Sec.6.2.2.

The second term of the cost function further provides fairness in the network by penalising those strategies that leave UEs without coverage. It is defined as δe_t, where δ is a unit price paid for each unserved user and e_t is the fraction of UEs within the team area that experience SINR levels below a certain threshold. We remark that since a macro PoA can communicate with the micro PoAs in the macrocell, the team leader has knowledge of the UE density under the coverage of its team players. Thus, it can easily estimate the fraction of users, e_t, depending on the strategy chosen for each of its players ($\boldsymbol{s^t}$) as well as on all other teams' strategies ($\boldsymbol{s^{-t}}$). The total cost function is then given by:

$$\pi^t(\boldsymbol{s^t}, \boldsymbol{s^{-t}}) = \sum_{l \in \mathcal{L}_t} \sum_{c \in \mathcal{C}} \xi^t_{l,c} \bar{a}_{l,c} s^t_{l,c} + \delta e_t \tag{6.4}$$

where ξ and δ indicate the weight that is assigned to each part of the cost function.

Finally, we define the payoff of each team t as the utility minus the cost paid:

$$w^t(\boldsymbol{s^t}, \boldsymbol{s^{-t}}) = u^t(\boldsymbol{s^t}, \boldsymbol{s^{-t}}) - \pi^t(\boldsymbol{s^t}, \boldsymbol{s^{-t}})\,. \tag{6.5}$$

In summary, we can formulate the problem as a competitive game $G = \{\mathcal{T}, \mathcal{S}, \mathcal{W}\}$, where $\mathcal{T}$ is the set of teams, $\mathcal{S}$ is the comprehensive set of strategies available to the teams, and $\mathcal{W}$ is the set of payoff functions. The objective of each team is to choose a strategy that maximises its payoff. Because its payoff depends also on the strategies of the other teams, a team must make decisions accounting for the strategies, it estimates or knows, the other teams have selected. Thus, using game-theory terminology, we will refer to the strategy chosen by a team as best reply. Moreover, to reduce both power consumption and the interference towards other teams, a team will select its best reply among strategies that maximise its payoff, as follows:

(i) Between strategies that are equivalent in terms of payoff, it will choose the one with the lowest total power, to reduce the overall power consumption.

(ii) When indifferent between strategies with equal total power but assigned to different locations, it will select the strategy that assigns higher power levels to micro PoAs that are closer to the centre of the cell, to minimise interference.

(iii) When indifferent with respect to the two above criteria, it will choose the strategy that assigns higher power levels to higher frequency carriers, again, to minimise interference.

6.2.2 Price setting

The price parameter $\xi^t_{l,c}$ introduced in Eq. (6.4) is an important parameter which affects the nature of the game. To gain some insight into the possible values of ξ we can start by considering a single carrier and reducing the number of players in the team to one. We further simplify the scenario to consider one tile per location, and dropping the superfluous notation, the team payoff becomes:

$$w^t = \frac{1}{\left(1 + e^{-\alpha(\frac{as^t}{\mathcal{I}^t} - \beta)}\right)} - \xi^t \bar{a} s^t \tag{6.6}$$

where $\mathcal{I}^t(\boldsymbol{s^{-t}})$ indicates the interference determined by other teams' strategies. We set $\delta = 0$, since the two cost components are independent of each other, therefore

the second component bears no effect on the analysis of the first component. Differentiating with respect to the team's chosen strategy, s^t, which now is scalar, and solving for 0, we obtain the following result:

$$e^{-2\alpha(\gamma^t-\beta)} - \left(\frac{\alpha}{\xi^t(\mathcal{I}^t)} - 2\right) e^{-\alpha(\gamma^t-\beta)} + 1 = 0 \tag{6.7}$$

It turns out that, in order to obtain a real and positive value for s^t, ξ^t must satisfy the following condition:

$$\xi^t \leq \frac{\alpha}{4\,(\mathcal{I}^t)} \tag{6.8}$$

The last expression indicates that the price parameter ξ^t is inversely proportional to the interference experienced by the UEs served by the player. If the interference experienced by the UEs in tiles served by the player increases, it is clear that the value of ξ^t needs to be lowered in order to ensure that the chosen power is a positive value. Note that, if we fix the value of ξ^t and the interference increases beyond a certain value, Eq. (6.8) will no longer be satisfied (i.e., the value obtained becomes complex), and the only possible strategy left for that player is to turn off its transmitter. This suggests that, in order to achieve high performing operational points for our network, a dynamic price setting is required, so that the teams can adapt to the changing interference, as other teams change their strategies.

We further remark that aside from being dynamically updated depending on the value of the interference, the price must also be tailored individually for each team player. Indeed, the interference experienced by UEs served by a specific location l depends not only on the strategies selected by other teams, but also on the topology of the network, i.e., the relative position and distance between the interfering players and said UEs. A team leader can leverage the knowledge it has about its team topology to adjust the price parameter, according to each player's expected external interference coming from other teams, and the expected intra-team interference.

Algorithm 6.1 Dynamic team price setting

Require: c, $\boldsymbol{s_c}$, t

1: **for all** $l \in \mathcal{L}_t$ **do**
2: $\quad \bar{I}_l^t = 0$
3: $\quad$ **for all** $z \in \mathcal{Z}_l$ **do**
4: $\quad\quad$ Compute $I_{z,c}^t$ by using Eq. (6.1)
5: $\quad\quad$ $I_{z,c}^{int} = \sum_{l' \in \mathcal{L}_t \wedge l' \neq l} s_{l',c}^t a_{l',z,c}$
6: $\quad\quad$ $\bar{I}_{l,c}^t = \bar{I}_{l,c}^t + \frac{E_z}{E_l}\left(I_{z,c}^t + I_{z,c}^{int}\right)$
7: $\quad$ $\xi_{l,c}^t = \frac{k\alpha}{\bar{I}_{l,c}^t}$

How to dynamically update the price for each team player under general settings is shown in Alg. 6.1. The procedure takes into account both the external interference coming from the other competing teams, calculated in line 4, as well as the internal interference coming from the other locations of the team, calculated in line 5. Once these values are obtained, the price parameter $\xi_{l,c}^t$ is updated in line 7 using $\xi_{l,c}^t = \frac{k\alpha}{\bar{I}_{l,c}^t}$, where k is a weight factor used to indicate the importance we place on the cost function; higher k values indicate that consuming less power will be given more consideration when selecting the best response. As a result, for higher k we obtain overall lower best response values, and vice versa. Note that $k \leq 1/4$ must hold in order to satisfy Eq. (6.8).

The algorithm can be applied at different update frequencies. In general, it can be executed either prior to a game start, or at the beginning of every iteration during the game. Note that the initial price for each team player is determined given an initial strategy, which can be any of the fixed strategies, and then updated every iteration/game. In practice, an update of the price parameter at the start of the game is sufficient as numerical results show that the performance is as good as when considering higher update frequencies.

6.2.3 Game analysis

To analyse the behaviour of the above-defined game, and discuss the existence of NEs, we rely on the definition of games of *strategic complements/substitutes with aggregation* as provided in [50, 53].

A game $\Gamma = \{\mathcal{P}, \mathcal{S}, \mathcal{W}\}$, where $\mathcal{P}$ is the set of players, and $\mathcal{S}$ and $\mathcal{W}$ are defined as above, is a game of **strategic substitutes** with aggregation if for each player $p \in \mathcal{P}$ there exists a best-reply function $\theta_p : \boldsymbol{S}^{-\boldsymbol{p}} \rightarrow \boldsymbol{S}^{\boldsymbol{p}}$ such that:

$$1) \quad \theta_p(I^p) \in \Theta(I^p) \tag{6.9}$$

$$2) \quad \theta_p \text{ is continuous in } \boldsymbol{S}^{-\boldsymbol{p}} \tag{6.10}$$

$$3) \quad \theta_p(\hat{I}^p) \leq \theta_p(I^p), \quad \forall \hat{I}^p > I^p \,. \tag{6.11}$$

$\Theta(I^p)$ is the set of best replies for player p and $\boldsymbol{S}^{-\boldsymbol{p}}$ is the Cartesian product of the strategy sets of all participating players other than p. I^p is an additive function of all other players' strategies, also referred to as the *aggregator* [53]:

$$I^p(\boldsymbol{s}^{-\boldsymbol{p}}) = \sum_{p' \in \mathcal{P}, p' \neq p} b_{p'} s_{p'} \tag{6.12}$$

where $b_{p'}$ are scalar values. Condition 1) is fulfilled whenever the dependence of the payoff function on the other players' strategies can be completely encompassed by the aggregator. Condition 2), also known as the *continuity* condition, implies that

for each possible value of I^p, the best reply function θ_p provides unique best replies. Condition 3) implies that the best reply of the team decreases with the value of the aggregator.

A game of **strategic complements** with aggregation is identical, except for condition 3), which changes into:

$$\theta_p(\hat{I}^p) \leq \theta_p(I^p), \quad \forall \hat{I}^p < I^p\,, \tag{6.13}$$

i.e., in the case of games of strategic complements, the best reply of the team increases with the value of the aggregator.

Next, we show the following important result.

Theorem 1. Our competitive team-based game G is a game of **strategic complements/substitutes with aggregation**.

Proof. Let us first consider a single-carrier, single-player team game and further simplify the scenario by assuming one tile per location. After removing the unnecessary notation, the interference expression given in Eq. (6.1) becomes $I^t(\boldsymbol{s^{-t}}) = \sum_{t' \in \mathcal{T} \wedge t' \neq t} s^{t'} a_{t'}$. It is clear that this expression fits the aggregator definition provided in Eq. (6.12), and consequently game G meets the conditions set out in Eqs. (6.9)-(6.10) and in either Eq. (6.11) or Eq. (6.13), as shown in [50]. The extension to a multi-carrier game with multi-player teams, implies that the strategy chosen by the team is not a scalar value but a matrix. Likewise, the interference experienced by each team (i.e., the aggregator) is a matrix. Without loss of generality, we can assume that the number of locations is the same in each team, and the set of the available carriers is the same for all teams. A decision has to be made for each location and each carrier. As already defined, the strategy of a team t, $\boldsymbol{s^t}$, is now an $L \times C$ matrix, while the team interference caused to team t, aggregated from other teams' strategies, can now be modelled as a $|Z_t| \times C$ matrix, each element of which is given by Eq. (6.1). The $\boldsymbol{I^t}$ matrix can be therefore expressed as:

$$\boldsymbol{I^t} = \sum_{t' \neq t} \sum_{l' \in \mathcal{L}^{t'}} \boldsymbol{a^t_{l'}} \boldsymbol{\sigma^{t'}_{l'}} \tag{6.14}$$

where $\boldsymbol{a^t_{l'}}$ is a $|Z_t| \times C$ matrix, populated by the attenuation values $a_{l',z,c}$, with each entry (z, c) indicating the attenuation factor from location l' in team t' to tile z in team t, on carrier c. $\boldsymbol{\bar{\sigma}^{t'}_{l'}}$ is a diagonal $C \times C$ matrix, where $diag(\boldsymbol{\bar{\sigma}^{t'}_{l'}}) = [s^{t'}_{l',c_1} s^{t'}_{l',c_2} \cdots s^{t'}_{l',c_C}]$. It is clear that the final interference matrix can be represented as an aggregation of interference matrices caused by each individual team, therefore the aggregator definition still fits. Condition 1) set out in Eq. (6.9), is fulfilled due to the very definition of our payoff function, since the dependence on the other teams' strategies is completely captured by the aggregator.

Regarding conditions 2) and 3), some further explanations have to be made to account for the fact that strategies are no longer single power level (scalar), but instead sets of power levels for the different locations and carriers within the team. Since both interference and strategy are formulated as matrices in our scenario, we have to define what signifies an increase/decrease in interference, and be able to distinguish between higher level strategies/lower level strategies. A natural way to quantify the value of a matrix would be to use the Frobenius norm, in which case, condition 3) becomes:

$$3) \quad ||\theta_t(\boldsymbol{I'^t})||_F \leq ||\theta_i(\boldsymbol{I^t})||_F, \quad \forall ||\boldsymbol{I'^t}||_F > ||\boldsymbol{I^t}||_F \tag{6.15}$$

$$3) \quad ||\theta_t(\boldsymbol{I'^t})||_F \leq ||\theta_i(\boldsymbol{I^t})||_F, \quad \forall ||\boldsymbol{I'^t}||_F < ||\boldsymbol{I^t}||_F \tag{6.16}$$

for games of strategic substitutes and games of strategic complements, respectively.

Note that the output of $\theta_t(\boldsymbol{I'^t})$ is a strategy, $\boldsymbol{s^t}$, for team t which is, as we said, a matrix. Similarly, we may use the Frobenius norm to differentiate between higher/lower strategies. To fulfil condition 2), the best-reply function must be continuous, i.e., the output of θ_t given a specific value of $\boldsymbol{I'^t}$ must be unique. In general, there may be cases in which a team may be indifferent between several strategies, in terms of payoff. In such a scenario, we consider that the team can apply the list of preferences (i)-(iii) in Sec. 6.2.1 to fulfil this condition. Concerning condition 3), we closely analyse the payoff function of the team given in Eq. (6.5). We note that the team payoff is a sum of individual payoffs obtained at each tile for each carrier. The payoff in each tile is directly linked to the interference value corresponding to that tile. Since we know that at the individual tile condition 3) holds (it is identical to the single-carrier single-player case), then it will hold also at the team level. Namely, when the level of interference experienced by a specific tile increases, increasing thus the value of the Frobenius norm of the interference matrix, then we know that the best reply of the individual location (which serves the specific tile) will be lower/higher depending on whether the game is of strategic substitutes or complements. Lowering/increasing the transmit power at one of the locations, indicates that the Frobenius norm of the strategy matrix will also decrease/increase. ■

As a further remark to the above result, it is worth stressing that the cost introduced in Eq. (6.4) is an important function that determines whether the game is of strategic complements or substitutes. Indeed, if we consider the payoff to coincide with the utility function (i.e., $\xi = \delta = 0$), a team's best reply will consist in increasing its transmit power as the interference grows, implying that the game is of strategic complements. This would lead to an NE in which all teams transmit at maximum power level, without consideration for the interference caused. Instead, imposing some $\xi > 0$, the game will turn into a game of strategic substitutes. This is because the first term of the cost function is linear with the received power, and hence increasing with the chosen strategies. Therefore, the payoff function will

Algorithm 6.2 BPS Algorithm run by team t at iteration $i+1$

Require: c, $\boldsymbol{s}_{\boldsymbol{c}}^{-\boldsymbol{t}}(i)$, $\xi, \delta, \alpha, \beta, \gamma_{min}$

for all $\boldsymbol{s} \in \boldsymbol{S}_{\boldsymbol{c}}^{\boldsymbol{t}}$ **do**
  Set $u^t(\boldsymbol{s}, \boldsymbol{s}_{\boldsymbol{c}}^{-\boldsymbol{t}}(i))$, $w^t(\boldsymbol{s}, \boldsymbol{s}_{\boldsymbol{c}}^{-\boldsymbol{t}}(i))$, $\pi^t(\boldsymbol{s}, \boldsymbol{s}_{\boldsymbol{c}}^{-\boldsymbol{t}}(i))$, e_t to 0
  for all $l \in \mathcal{L}_t$ **and** $z \in \mathcal{Z}_l$ **do**
    Compute $I_{z,c}^t$ by using Eq. (6.1)
    Compute $\gamma_{z,c}^t$ by using Eq. (6.2)
    $u^t(\boldsymbol{s}, \boldsymbol{s}_{\boldsymbol{c}}^{-\boldsymbol{t}}(i)) \leftarrow u^t(\boldsymbol{s}, \boldsymbol{s}_{\boldsymbol{c}}^{-\boldsymbol{t}}(i)) + \frac{E_z}{E_t\left(1+e^{-\alpha(\gamma_{z,c}^t-\beta)}\right)}$
    $\pi^t(\boldsymbol{s}, \boldsymbol{s}_{\boldsymbol{c}}^{-\boldsymbol{t}}(i)) \leftarrow \pi^t(\boldsymbol{s}, \boldsymbol{s}_{\boldsymbol{c}}^{-\boldsymbol{t}}(i)) + \xi \bar{a}_{l,c} s_{l,c}$
    if $\gamma_{z,c}^t \leq \gamma_{min}$ **then**
      $e_t \leftarrow e_t + \frac{E_z}{E_t}$
    $\pi^t(\boldsymbol{s}, \boldsymbol{s}_{\boldsymbol{c}}^{-\boldsymbol{t}}(i)) \leftarrow \pi^t(\boldsymbol{s}, \boldsymbol{s}_{\boldsymbol{c}}^{-\boldsymbol{t}}(i)) + \delta e_t)$
  $w^t(\boldsymbol{s}, \boldsymbol{s}_{\boldsymbol{c}}^{-\boldsymbol{t}}(i)) \leftarrow u^t(\boldsymbol{s}, \boldsymbol{s}_{\boldsymbol{c}}^{-\boldsymbol{t}}(i)) - \pi^t(\boldsymbol{s}, \boldsymbol{s}_{\boldsymbol{c}}^{-\boldsymbol{t}}(i))$
$\boldsymbol{s}_{\boldsymbol{c}}^{\boldsymbol{t}}(i+1) \leftarrow \arg\max_{\boldsymbol{s}}^{\star} w^t(\boldsymbol{s}, \boldsymbol{s}_{\boldsymbol{c}}^{-\boldsymbol{t}}(i))$

start decreasing once the increase in the chosen transmit powers does not justify the price the team has to pay. Note that, throughout the paper, we will consider $\xi > 0$, therefore our game is of strategic substitutes. Imposing some $\delta > 0$ (i.e., activating the second cost component), the relationship between transmit power and cost becomes more complicated but it does not change the nature of the game: the fraction of unserved UEs within the team will be high for very low power strategies, then it will decrease as the transmit power is increased, and increase again as the strategies chosen cause high intra-team interference. In other words, the second cost component strengthens the trend in the payoff function imposed by the utility for increasing interference in presence of low power strategies. For those mid-level strategies that ensure good coverage, it does not affect the cost function. Instead, it resembles the behaviour of the first cost component for high power strategies, as it is still able to discriminate against high power strategies that may harm the system performance.

Main results from [50, 53] and references therein show that games of strategic complements/substitutes with aggregation belong to the class of potential games, specifically to the subclass of *pseudo-potential games*. These games admit pure Nash Equilibria (NE), i.e., action profiles that are a consistent and stable prediction of the outcome of the game, in the sense that no player has incentive to unilaterally deviate from such strategies. Another important result that holds for such games with a discrete set of strategies is that, thanks to the continuity condition in Eq. (6.10), convergence to an NE is ensured by best reply dynamics [50, 53].

6.3 The power setting algorithm

We now use the above model and results to build a distributed, low-complexity scheme that enables efficient downlink power setting on each CC. We first consider a single carrier and show that it converges to the best NE among the possible ones, in terms of payoff. We aim for an NE because it is the only solution of the game which the participating teams can reach independently, although it may not be the most optimal one in terms of utility. We then extend the algorithm to the multiple-carrier case and discuss its complexity.

6.3.1 Single-carrier scenario

Let us first focus on a single carrier and consider two possible borderline strategies that a team may adopt: the *max-power* strategy in which all locations transmit at the highest power level, and the *min-power* strategy in which all locations transmit at the lowest available power level greater than 0. Evaluating the utility values obtained for the two extreme strategies, both at the global and individual team level, it transpires

that the *min-power* always outperforms the *max-power* in a multi-tier dense scenario. Indeed, the inter-tier and inter-team interference seriously undermines the overall network performance in terms of social welfare, expressed as the sum of all individual team utilities (see Eq. (6.2)–(6.3)), as shown in Sec. 6.4.3. With regard to the cost, as discussed in Sec. 6.2.3, the first component increases with the increase in the selected transmit power. The second component strengthens the trend imposed by the first cost component for the *max-power* strategy, and by the utility for the *min-power* strategy.

We therefore devise the following procedure that should be executed by each team leader (macro PoA), in order to update the PoAs downlink power setting, either periodically or upon changes in the user traffic or propagation conditions. At a given update period, all teams initialise their transmit power to zero. Then, they sequentially run the Best-reply Power Setting (BPS) algorithm reported in Alg. 6.2. We refer to the single execution of the BPS algorithm by any of the teams as an iteration. Note that the order in which teams play does not affect the convergence or the outcome of the game, since all teams start from the zero-power strategy. At each iteration, the leader of the team that is playing determines the strategy (i.e., the power level to be used at each location in the team) that represents the best reply to the strategies selected so far by the other teams. The team leader will then notify it to the neighbouring team leaders that can be affected by this choice. BPS will be run by the teams till convergence is reached, which, as shown in Sec. 6.4.3, occurs very swiftly. Also, we remark that the strategies identified over the different iterations are not actually implemented by the PoAs. Only the strategies representing the game outcome will be implemented by the PoAs, which will set their downlink power accordingly for the current time period.

In order to detail how the BPS algorithm (Alg. 6.2) works, let us consider the generic $i+1$-th iteration and denote the team that is currently playing by t. The algorithm requires as input the carrier c at disposal of the PoAs and the strategies selected so far by the other teams, $\boldsymbol{s_c^{-t}}(i)$. Additionally, it requires the cost components weights ξ and δ, the SINR threshold γ_{min}, used to qualify unserved users, and the utility function parameters α and β. This latter set of parameters are calculated offline and provided to the teams by the network operator. The algorithm loops over all possible strategies in the strategy set of team t, $\boldsymbol{S_c^t}$. For each possible strategy, $\boldsymbol{s}$, and each location l within the team, it evaluates the interference experienced by the tiles within the location area (line 4). This value is used to calculate the SINR and the utility (lines 5-6), then the first cost component is updated (line 7). In line 8, it is verified whether UEs in tile z achieve the minimum SINR value. If not, the cost component e_t is amended to include the affected UEs. The overall team utility for each potential strategy $\boldsymbol{s}$ is obtained by summing over the individual tile utilities weighted by the fraction of UEs present in each tile. We recall that such weight factor ensures that the UE distribution affects the outcome of the game accordingly.

Once the utility and cost are obtained, the team payoff corresponding to strategy $\boldsymbol{s}$ is calculated (line 11). After this is done for all possible strategies, the leader chooses the strategy $\boldsymbol{s}^t(i+1)$ that maximises the team payoff. Note that, according to our game model, $\arg\max^\star$ in line 12 denotes the following operation: it applies the $\arg\max$ function and, if more than one strategy is returned, the best strategy is selected by applying the list of preferences in Sec. 6.2.1.

Theorem 2. When the NE is not unique, then the BPS algorithm reaches the NE that maximises the social welfare, i.e., the sum of individual payoffs.

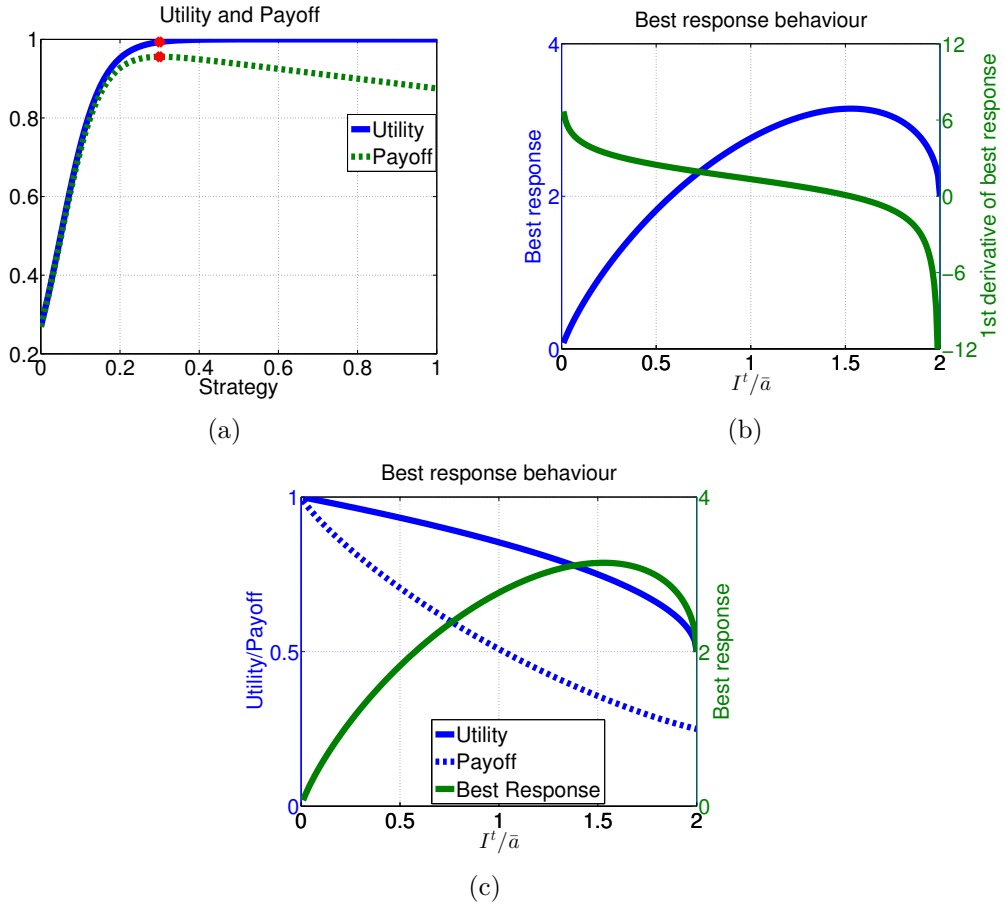


Figure 6.2. Utility and payoff as functions of the player's strategy, assuming fixed interference (a); the behaviour of the best response and its first derivative as interference increases (b); utility and payoff obtained by applying best response strategy as interference increases (c). The price is assumed to be fixed at $\xi^t = \alpha/8\bar{a}$, where $\bar{a}$ is the average attenuation.

Proof. Given that ours is a game of strategic substitutes with aggregation and that teams are provided with discrete strategy sets, the best reply convergence is defined

as the convergence of sequential best replies to other teams' strategies [50]. Taking this into account, let us analyse how the game proceeds when teams sequentially play their best reply strategies, according to the BPS algorithm.

For ease of presentation, we consider the simplified scenario we referred to before, with two single-location teams and one tile per location. Allowing for a continuous, rather than discrete strategy set, the best reply response has the following analytical expression:

$$\begin{aligned} \boldsymbol{s}^{\boldsymbol{t}}_{\boldsymbol{br}}(I^t) &= \arg\max w^t(\boldsymbol{s}^{\boldsymbol{t}}, \boldsymbol{s}^{-\boldsymbol{t}}) \\ &= -\frac{I^t+N}{\alpha\bar{a}}\left[\ln\left(\frac{\alpha}{2\xi^t(I^t+N)} - 1 - \sqrt{\left(1-\frac{\alpha}{2\xi^t(I^t+N)}\right)^2 - 1}\right) - \alpha\beta\right]. \end{aligned} \tag{6.17}$$

where the expression in the second line of Eq. (6.17) has been obtained by evaluating the first derivative of w^t with respect to $\boldsymbol{s}^{\boldsymbol{t}}$ and solving for zero. Then recall that the user's payoff is given by its utility, represented by a sigmoid function, discounted by the power cost. The value of the sigmoid function does not exceed 1 and increases quickly as it reaches its saturation point (see Eq. (6.3) and Fig. 6.2 (a)). It follows that, since the cost is linear with respect to increasing power, a team's best reply coincides with the lowest strategy that reaches the saturation region. Also, from the expression in Eq. (6.17) and its derivative with respect to interference, we can see that, initially when interference is low, a team's best response increases quickly with interference, then it reaches a maximum point and finally it decreases as interference becomes too high (see also Fig. 6.2 (b)).

As a consequence, when both teams start from zero strategy as foreseen in BPS, the game will evolve as follows. The first team to play will increase its power, in order to receive a payoff that is higher than zero. However, since the interference it receives from the other team is zero, the first team will be satisfied with choosing the leftmost strategy that reaches the saturation point, which coincides with the peak of the payoff function (see also Fig. 6.2 (a)). The first team's increased transmit power will prompt an increase in the best response of the second player, who plays next. The second move of the first team will be to increase its power again (i.e., looking at Fig. 6.2 (b), its best response moves to the right due to the increased interference), and the pattern is continued by the second player. When the interference reaches the point at which the best response has a maximum for at least one of the teams, they will react by not increasing their transmit power further. It is sufficient for one of the teams at some point in the game to stop increasing its strategy, for the other team to follow suit, causing a domino effect. This will happen because the other team will experience the same interference as the previous iteration, therefore it will have no interest to move from its previous strategy. At this point in the game, we therefore have converged to a Nash equilibrium (none of the teams has

any incentives to move from its chosen strategy). Because each team started from zero, incrementally increasing its strategy (hence the interference), we are sure that the NE that is reached is at the lowest possible strategy level for each team. Hence, this is the NE corresponding to the lowest overall transmit power.

We are now left to show that such an NE is the best in terms of social welfare, which is defined as the sum of individual payoffs of the teams. To this end, we substitute the expression obtained for $\boldsymbol{s_{br}^t}$ (given in Eq. (6.17)) in the utility and payoff functions, so that we obtain the expressions that describe the trends of the utility and payoff when applying the best response, as the interference increases:

$$u_{br}^t(I^t) = \frac{2\xi^t(I^t+N)}{\alpha - \sqrt{\alpha^2 - 4\xi^t(I^t+N)}} \tag{6.18}$$

$$w_{br}^t(I^t) = u_{br}^t(I^t) - \boldsymbol{s_{br}^t}(I^t)\,. \tag{6.19}$$

For clarity, the expressions for $u_{br}^t(I^t)$ and $w_{br}^t(I^t)$ are plotted in Fig. 6.2 (c). From Eqs. (6.18)-(6.19), it is clear that, as the interference increases, u_{br}^t and w_{br}^t decrease, even though the teams are playing their best responses. Let us assume, by contradiction, that there is a second NE which provides a better social welfare. Because this NE was not reached according to BPS, it must be a point that has not been explored by the algorithm. Therefore we can conclude that at least for one of the teams, the strategy in this second NE is higher than in the previous NE. That is, the overall transmitted power in this second NE must be higher, implying that at least some teams are facing higher interference than in the previous NE. As shown by the above equations, the utility and payoff values are always decreasing with interference, therefore the utility/payoff values of teams facing higher interference must definitely be lower than those obtained in the previous NE, implying that the social welfare must also be lower. If, on the other hand, we assume that an NE is further to the right, in the region of interference values where we have a decreasing best response behaviour at least for one or more teams, then that would clearly trigger a cycle where all teams start decreasing their best responses due to the low interference, therefore proving that such a point cannot be an NE.

This is exemplified in Fig. 6.2 (c), which shows that the payoff value decreases even more rapidly with the increase in interference, due to the increase in the cost component. Thus, an NE associated with the lowest overall power, i.e., with the least amount of interference, is the most optimal NE in terms of social welfare. At last we remark that the NE reached through our BPS algorithm may not coincide with the global optimal point in terms of social welfare, but it is the optimum among the game's pure NEs. ■

6.3.2 Multi-carrier scenario

We now extend the BPS algorithm to the multi-carrier case. As mentioned before, the team leader has to decide on the power level to be used at each available carrier, at each location within the team. Thus the team strategy is no longer a vector, but an $L \times C$ matrix, each entry (l, c) indicating the power level to be used for carrier c at location l. A straightforward extension of Alg. 6.2 would imply that lines 1–11 are executed for each element in the new extended strategy set. However, the new strategy set, depending on the number of carriers, may become too large and therefore make the algorithm impractical to use in realistic scenarios.

Analysing the utility expression obtained in Eq. (6.3), we can note that since the carriers are in different frequency bands and have separate power budgets (as foreseen in LTE-A), the utilities secured at each carrier are independent of each other. In other words, the utility a team will get at one of the carriers, is not affected by the strategy chosen at another carrier. The same holds for the first cost component in Eq. (6.4). However, the overall payoff value is dependent on the interaction between carriers due to the second cost component. Indeed, in networks with CA support, a UE can be considered unserved only if the SINR it experiences is below the threshold in all carriers. In order to obtain a practical and effective solution in the multi-carrier scenario, we take advantage of the partial independence between the carriers, and run Alg. 6.2 independently for each carrier, keeping the size of the strategy set the same as in the single-carrier scenario. Then, to account for the dependence exhibited by the second cost component, we set the order in which the per-carrier games are played, using the order of preferences listed in Sec. 6.2.1. Since the teams prefer to use high-frequency carriers over low-frequency ones, due to their smaller interference impact, it is logical that the game is played starting from the highest-frequency carrier. It follows that low-frequency carriers will likely be used to ensure coverage to UEs not served otherwise.

Importantly, our algorithm is still able to converge to an NE, since surely none of the teams will deviate from the strategies they chose at each carrier. Also, since the game for the lowest frequency carrier is played last, the number of served UEs cannot be further improved without increasing the power level on the other carriers, which we already know is not a preferable move as it has not been selected earlier. Thus, although it does not search throughout the entire solution space as for the single-carrier scenario, the procedure is still able to converge to an NE that provides a close-to-optimum tradeoff among throughput, user coverage and power consumption. The results presented in Sec. 6.4.3 obtained for toy scenarios, confirm that our scheme provides performance as good as that achieved by an exhaustive search in the strategy space.

6.3.3 Complexity

The complexity of the algorithm depends largely on the size of the strategy sets that are available to the teams, $\boldsymbol{S^t}$, since each team has to find the strategy which maximises its payoff value by searching throughout the entire set. The set size depends on the number of discrete power levels available to the PoAs ($|\boldsymbol{P}|$), the number of locations in the team (L) and the number of CCs available at each location (C). In the single-carrier scenario, we have $|\boldsymbol{S^t}| = |\boldsymbol{P}|^L$, while in the multi-carrier scenario the size exponentially grows to $|\boldsymbol{S^t}| = |\boldsymbol{P}|^{LC}$, which is reduced to $|\boldsymbol{S^t}| = C|\boldsymbol{P}|^L$ by our approach.

6.4 Performance evaluation

In this section we look at the performance of the proposed approach when compared to several baseline solutions. We first look at the game behaviour and evaluate the performance of the algorithm in terms of utility, payoff and convergence, using the static scenario we used in the previous chapter. In the second section, instead we use a dynamic and more extended scenario, in order to evaluate the performance of the algorithm, in almost realistic conditions. Considering that the algorithm's ability to adapt to evolving user distributions and traffic loads in one of the key features of our approach, the scenario used during simulation is dynamic both in space and time. The second scenario is used to evaluate the performance of the algorithm in key performance metrics relevant to cellular networks, such as energy efficiency (with respect to actual consumed power), spectrum usage efficiency, average user throughput and other quality of service related metrics.

6.4.1 Static simulation scenario

We consider the same two-tier network scenario as in the previous chapter, that is typically used within 3GPP for evaluating LTE networks [78]. The network is composed of 57 macrocells and 228 microcells. Macrocells are controlled by 19 three-sector macro PoAs, while micro PoAs are deployed over the coverage area so that there are 4 non-overlapping microcells per macrocell. The inter-site distance is set to 500 m. The overall network area is divided into 2,478 square tiles of equal size. The PoAs are grouped into 57 five-player teams, each consisting of 1 macro BS and 4 micro PoAs within its macrocell. There are about 34,400 UEs in the area, distributed non-uniformly with a user density around micro PoAs that is three times higher than over the macro BS coverage area. All UEs are assumed to be CA enabled and associated to the closest PoA. PoAs can use three CCs, each 10 MHz wide, with the central frequencies: 2.6 GHz (CC1), 1.8 GHz (CC2) and 800 MHz (CC3). We assume a slow-fading environment, where signal attenuation and losses

follow the ITU specification for urban environments [69], while the SINR values are mapped to throughput using the look-up table in [82]. The maximum transmit powers for macro and micro PoAs are set at 20 W and 1 W, respectively. The set of discrete power levels is given by $\boldsymbol{P} = \{0,0.1,0.2,...,1\}$, each representing a fraction of the maximum power. The game is played by all teams using Alg. 6.2 applied in a multi-carrier setup. The sigmoid function parameters are $\alpha = 1$ and $\beta = 1$, which were selected as the most appropriate to model the relationship between the selected strategy and final user rate. The SINR threshold is set at $\gamma_{min} = -10$ dB, based on [82]. The value of the cost parameter is fixed and the same for all teams, set as $\xi = \frac{k\alpha}{\bar{\mathbf{I}}}$, where k is the weight factor used to indicate the importance we place on the first cost component and $\bar{\mathbf{I}}$ is an average value for interference calculated by the network operator, obtained by fixing the transmit power of all teams at half the maximum power. Unless otherwise specified, the weight factor k is set to 0.25 while $\delta = 0.6$. These values were selected based on their effect on the performance metrics, as shown below in Sec. 6.4.3.

6.4.2 Baseline solutions

The performance of the BPS algorithm is compared to four baseline power setting strategies:

Max-power strategy

The max-power strategy implies that all PoAs transmit at maximum allowable transmit power for their respective type. No interference mitigation technique is applied. Distance-based fixed user association is applied and single-flow implementation of CA.

Min-power strategy

Similarly, the min-power strategy implies that all PoAs transmit at the lowest non-zero transmit power from their respective strategy sets. No interference mitigation technique is applied. Distance-based fixed user association is applied and single-flow implementation of CA.

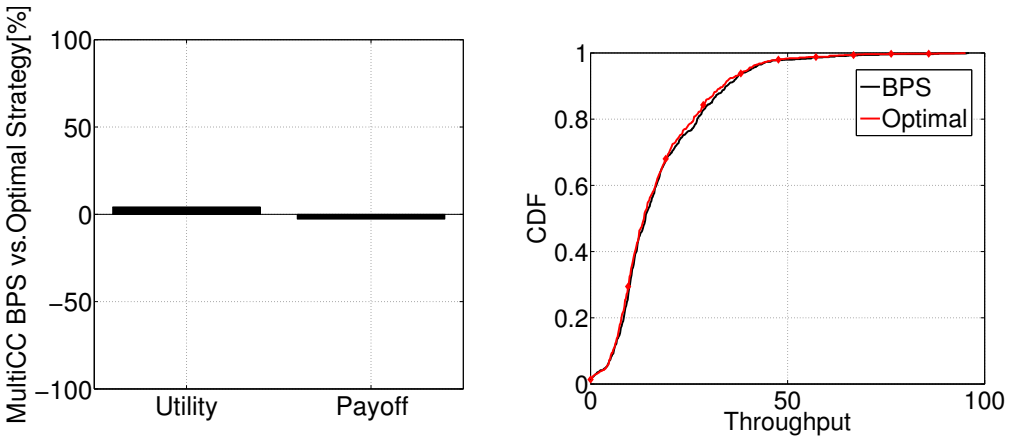
eICIC

The *max-power* strategy coupled with eICIC technique is also considered. CRE is considered also for microcells, with biasing factor set at 8 dB, as and macro PoA downlink transmissions muted in 25% of subframes (ABS). These values were chosen to represent the mid-range of those applied in the surveyed literature [22]. User association is performed as described in Sec. 5.3.2 for single-flow implementation.

Low Power ABS (LP-ABS)

LP-ABS is a modification of eICIC, which adopts reduced transmit power at the macro PoAs, instead of muting, during ABS. During the simulation we apply 6 dB microcell biasing, ABS subframe ratio of 50% and macro BS power reduction of 6 dB during ABS, which were shown to perform best in [83].

6.4.3 Game behaviour



(a) (b)

Figure 6.3. Deviation from optimal strategy: utility, payoff and overall transmitted power (a) and CDF of the per-user throughput (b).

In Fig. 6.3 we compare BPS in a multi-carrier setting with the optimal solution obtained via exhaustive search. Due to the problem complexity, the comparison is performed only for a toy scenario in which two teams compete, each consisting of one macro and one micro PoA. The results, obtained by averaging the behaviour of ten different sets of teams, show that there is negative deviation in terms of payoff as expected, but BPS yields higher utility. Looking at the per-user throughput CDF curves, however, we note that the two strategies perform almost identically.

In Fig. 6.4, we look at a snapshot of the NE strategy reached via the BPS algorithm, in a game with 57 teams. The strategies chosen by the teams for each CC are differentiated using different shades, from white (*zero* power) to black (*maximum* power). Hexagons represent the macro PoA, while circles represent micro PoAs. The figure shows that CC1, i.e., the high frequency carrier, allows for higher transmit power to be used by both macro and micro PoAs, due to its low interference impact. CC1 can be also used simultaneously by macro and micro PoAs in the same team, which is not always the case for the other two CCs. CC2 and CC3 are used to complement each other to ensure overall coverage. Histograms of chosen strategies for macro and micro PoAs, shown in Fig. 6.5, confirm these observations. Here note that CC1 is activated for most macro and micro PoAs, however macro PoAs often set low power levels for CC1, while most micro PoAs set CC1 at maximum power level. On the contrary, CC2 is rarely activated for macro PoAs, while CC3 (the low frequency CC) is the least utilised, and tends to be especially unfavored by micro PoAs, due to its high interfering impact. These results validate the intuition that far reaching low-frequency carriers are not appropriate to be used by micro PoAs, rather they should be used only to ensure broader coverage for edge UEs.

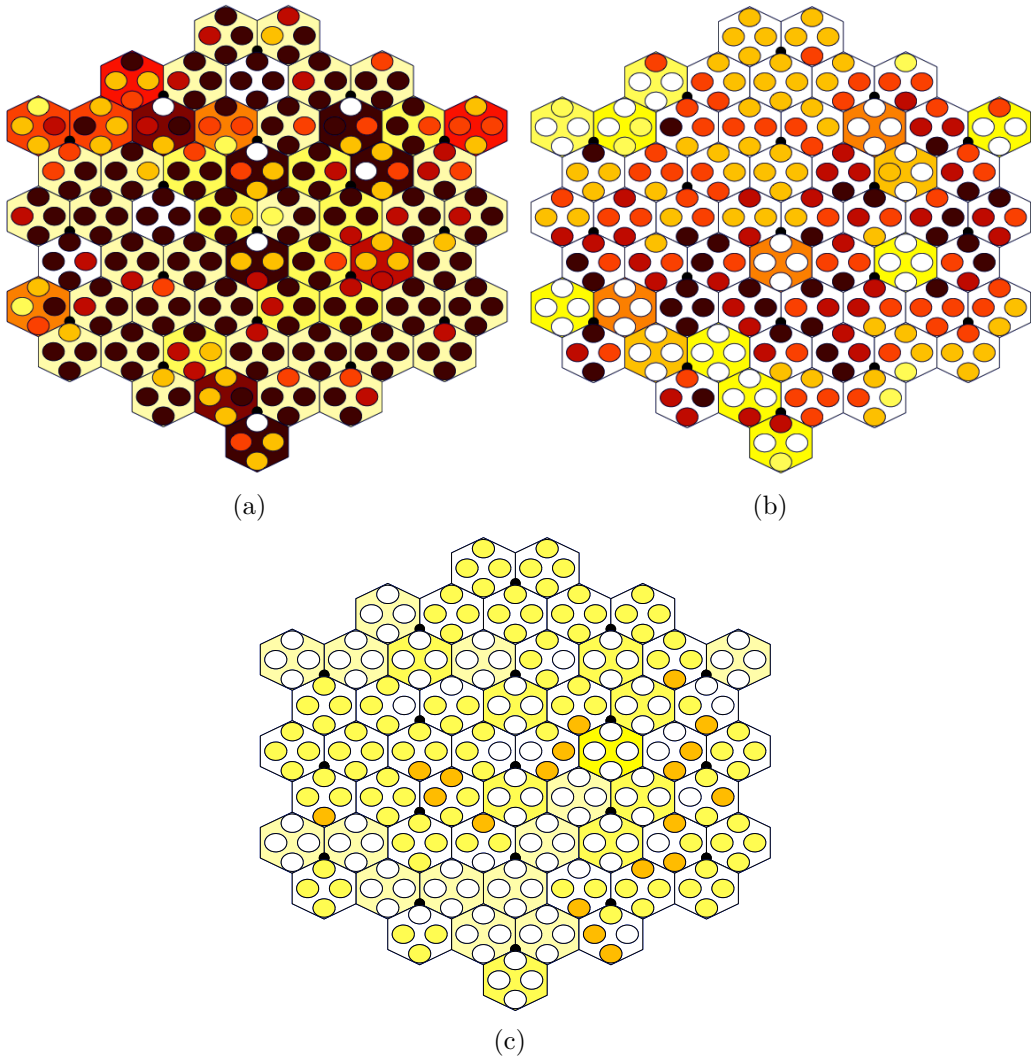
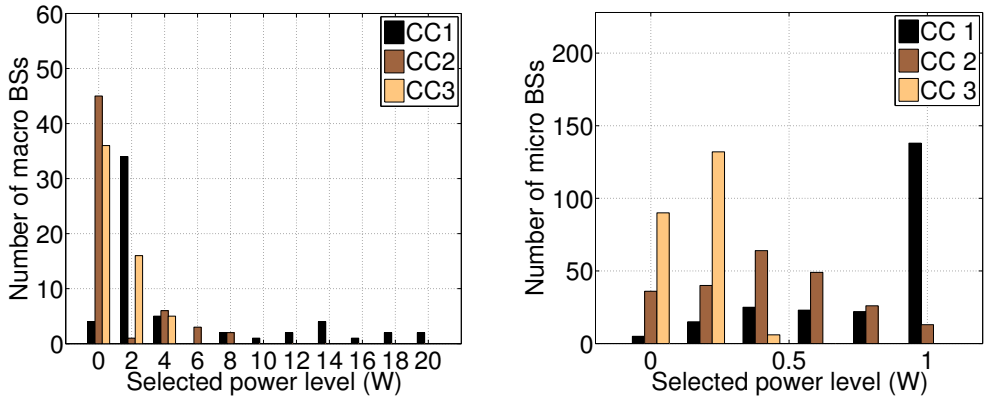


Figure 6.4. BPS strategies for a 57-team game for CC1 (a), CC2 (b) and CC3 (c). Darker shades represent higher power level, while the white color corresponds to the *off* state. Hexagons are macro PoAs while circles are micro PoAs.

Next, in plots (a) and (b) of Fig. 6.6 we compare the performance of the strategy reached via our scheme (labelled by "BPS") to the fixed baseline strategies, in terms of global utility and overall transmitted power, and for a varying number of teams. The strategy reached via the BPS mechanism outperforms all other solutions in terms of global utility, calculated as the sum of the individual team utilities. Also, the gap in performance grows with the number of teams. This gain in performance is achieved at much lower transmit power, which implies that the BPS strategy is very efficient. The overall transmit power of the BPS strategy, calculated as the sum of the selected transmit powers over all BSs and CCs in the network, closely approaches



(a) (b)

Figure 6.5. BPS strategies for a 57-team game: chosen strategies by macro (a) and micro (b) PoAs.

that of the *min-power* strategy and is much lower than the power consumption of all other schemes. Also, as anticipated in Sec. 6.3.1, the *min-power* strategy always outperforms the *max-power* strategy in terms of utility, regardless of the number of teams, while keeping the overall transmit power at the minimum level.

The final comparison is performed in Fig. 6.6(c), which depicts the cumulative distributive function (CDF) of the per-user throughput for the strategies under consideration. Overall, our solution outperforms all other schemes. This holds especially for the top 70% of UEs. eICIC and LP-ABS give slightly better results in ensuring a positive throughput to the worst UEs. However, BPS provides a very low fraction of UEs that are left unserved (about 2%), while transmitting at much lower overall power. Note also that the strategies with eICIC and LP-ABS are at a slight advantage since user association is performed based on the best downlink pilot signal, which, at least for downlink communication, is always better than the fixed distance-based user association scheme that we assumed for simplicity. In summary, it is clear that BPS is a very well-balanced strategy in terms of level of service: it provides slightly lower per-user throughput than eICIC and LP-ABS for the worst UEs, but much better throughput than all other strategies for the rest of the UEs, and it consumes very little power (almost the same as the *min-power* strategy).

In Fig. 6.7, we look at the behaviour of our algorithm. First, we evaluate the effect of k, i.e., the weight we assign to the cost of received power, on the global utility and the fraction of low SINR users, by varying its value from 0 to 1 and fixing $\delta = 0.6$. We see that increasing k is beneficial in terms of global utility (solid, blue line), but only up to some value (around 0.4). Beyond that, the global utility experiences a sharp drop, which signifies that, due to the high power price, BPS is more inclined to provide strategies that optimise power consumption rather than

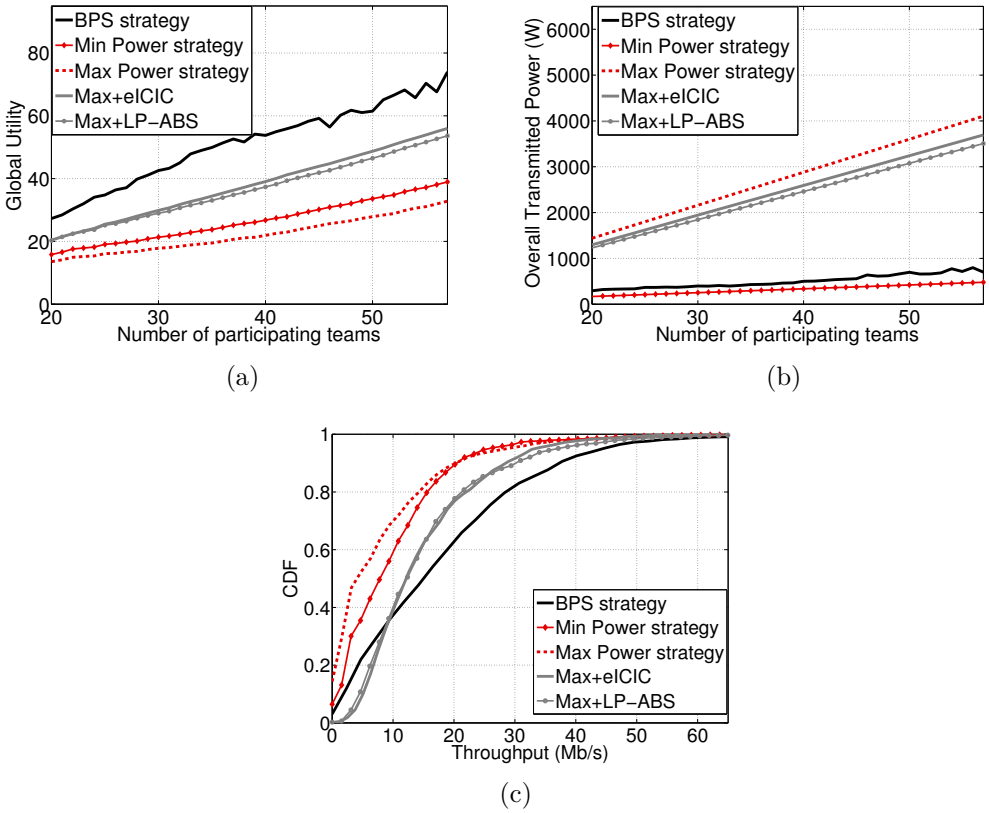


Figure 6.6. BPS strategy for a 57-team game: comparison with baseline strategies for varying number of teams. Global utility (a), overall transmitted power (b) and CDF of the per-user throughput (c).

the utility. Also, k has little effect on the fraction of unserved users (dashed, green line): just a small improvement can be noticed around $k = 0.25$. Conversely, the cost parameter δ plays an instrumental role in ensuring that the number of UEs experiencing an SINR below the acceptable threshold is kept low, as can be seen by the dashed green line in the second plot of Fig. 6.7 (here $k = 0.25$). The third plot depicts the effect of k (solid, blue line) on the overall transmitted power when $\delta = 0.6$, and the effect of δ (dashed, green line) when $k = 0.25$. Note that increasing k leads BPS to converge to strategies with overall lower power, however, as observed before, this comes at the expense of the utility. As expected, the increase in δ does not lead to strategies with higher overall transmit power, which confirms our earlier statement that introducing the second cost component does not change the nature of the game.

Fig. 6.7(d) presents the average number of iterations it takes to each team to converge to the final best strategy. Depending on the intra-team dynamics, teams

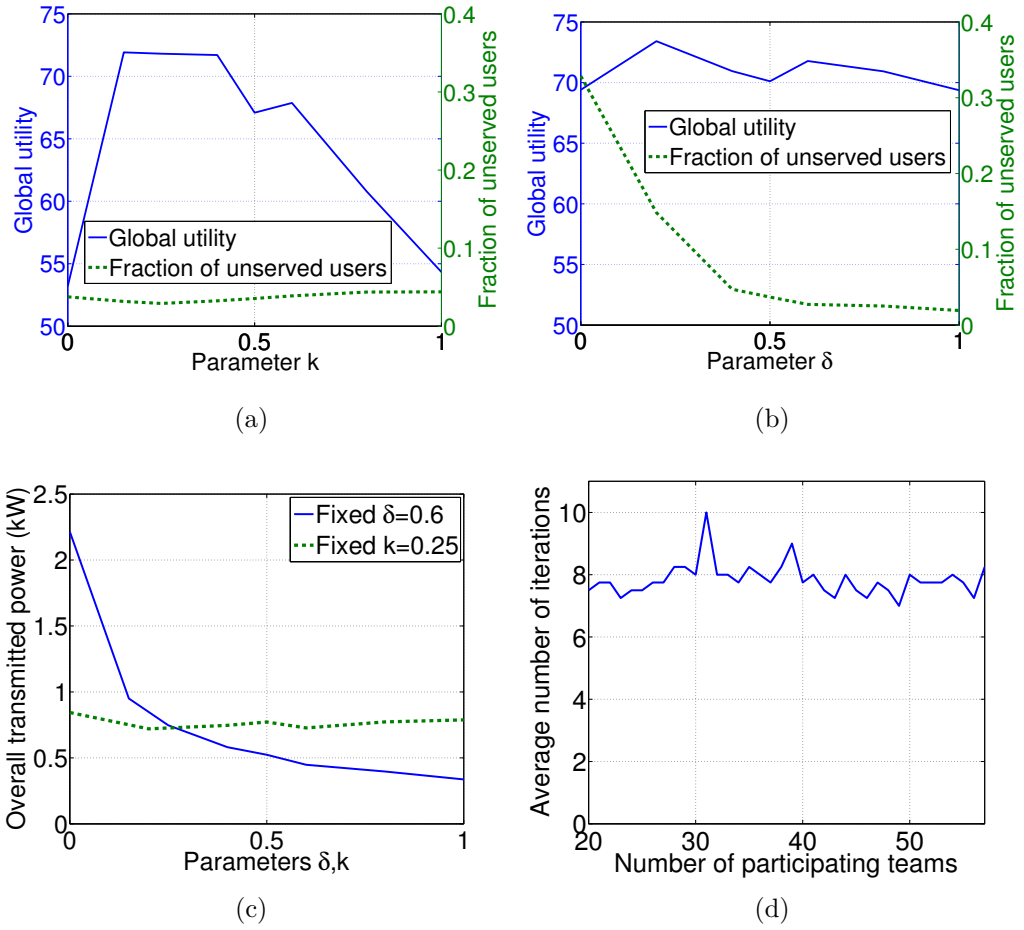


Figure 6.7. (a) Effect of the price parameter k on global utility (solid, blue) and fraction of unserved users (dashed, green) with for $\delta = 0.6$; (b) Effect of the coverage cost parameter δ on global utility (solid, blue) and fraction of unserved users (dashed, green), with $k = 0.25$; (c) Effect of k (solid, blue) and δ (dashed, green) on the overall transmitted power; (d) Per-team number of iterations for game convergence vs. number of teams.

may take a different time, however the game always converges quite fast (in about 8 iterations). Importantly, the average number of iterations required by each team does not grow with the number of teams. To demonstrate how fast the teams reach convergence, an example is shown in Fig. 6.8 depicting BPS outputs for different team members at each iteration. For the particular team in question, the convergence is reached quite fast in all three carriers.

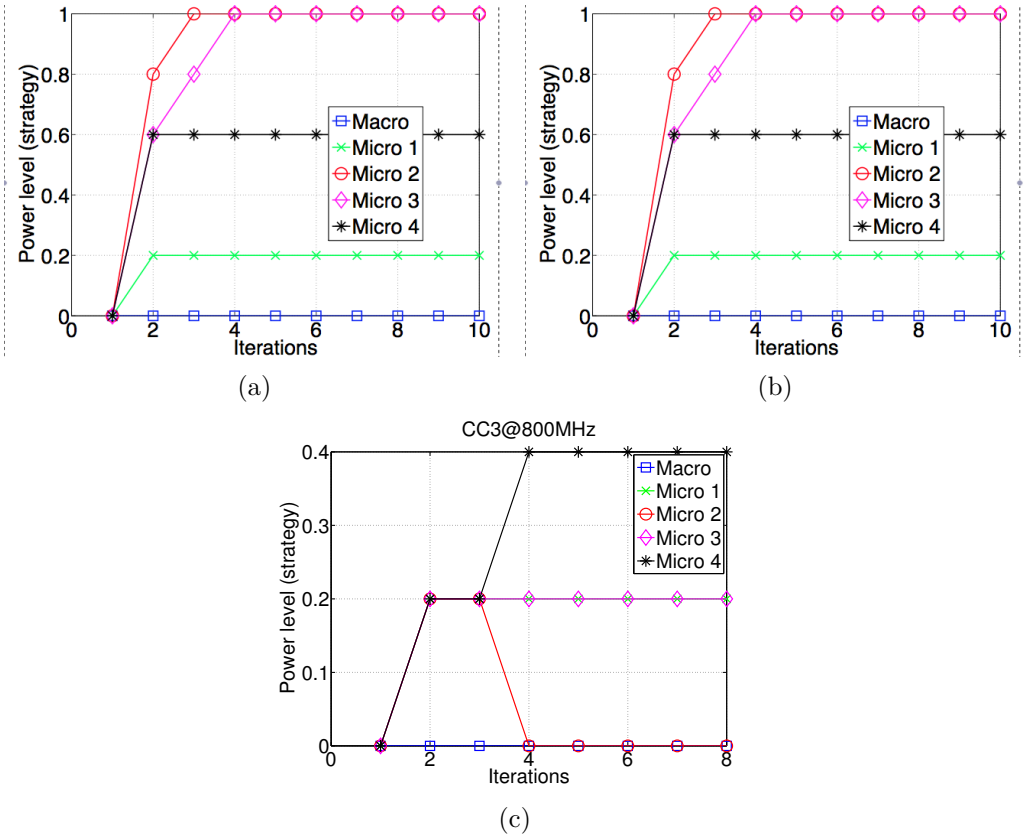


Figure 6.8. BPS algorithm convergence for an arbitrarily selected team. (a) Convergence in CC1; (b) Convergence in CC2 and (c) Convergence in CC3.

6.4.4 Dynamic simulation scenario

While the number of macrocells and microcells remain the same, in the dynamic scenario we consider that micro PoAs are deployed randomly over the coverage area so that there are 4 non-overlapping microcells per macrocell. The number of tiles is increased to 4,560 for finer resolution. Specifically, to make the scenario more realistic and comparable to an actual urban scenario, we divide the network coverage area into five types of urban areas: city centre, residential area, commercial area, parks and school area, as shown in Fig. 6.9. The UEs are also randomly dropped with varying density depending on the population density of the area type as well as time of the day (morning, afternoon or evening). Reference values for UE density were obtained using official population statistics of the city of Rome (Italy) [84], and then scaled to represent realistic values for cellular users of a single network provider. The UE densities were further scaled for the different urban areas and times of the day, using weights extracted from the data provided in the MIT Senseable City Lab

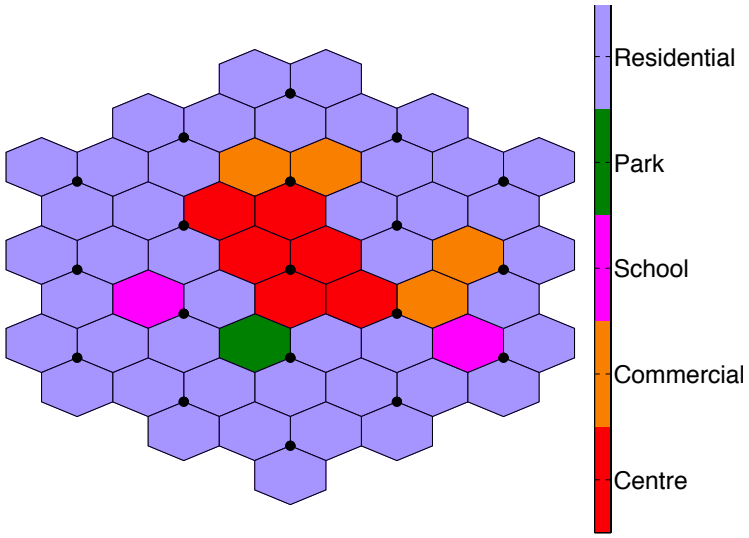


Figure 6.9. The network scenario and the different types of urban areas.

project [85]. The final values obtained and used for the numerical simulations are shown in Table. 6.2. Note that, in addition, user density around micro PoAs is four times higher than over the rest of macrocell. The mobility of pedestrian UEs was modelled using the random walk model, while the mobility of vehicular UEs was modelled using real mobility traces collected from taxi cabs in Rome [86], assuming an average velocity of 30 km/h. Snapshots of user distribution at different times of the day are shown in Fig. 6.10. The data traffic is simulated by generating download requests, whereby a random user requests to download a file which can be either video (file size: 1 Mb) or a generic file (file size: 500 kb), with equal probability. The number of requests per cell follows a Poisson distribution with a certain arrival rate λ per cell. The UEs making the requests are chosen randomly from the set of cell UEs, which are not currently downloading. The arrival rate λ varies depending on the urban area and time of the day. We assume a maximum λ value of 1.5 arrivals per subframe during peak hours in the city centre, which is then proportionally scaled using the density weights. The values obtained for the different λ's are shown in Table. 6.2.

The value of the cost parameter ξ, is calculated before running the BPS, using the dynamic pricing algorithm Alg. 6.1 in Sec. 6.2.2. The power setting update period is set at 100 ms. Unless otherwise specified, the weight factor for the second cost component is kept at $\delta = 0.6$.

The performance of the approach is compared to the fixed strategy in which all PoAs transmit at highest power coupled with the eICIC technique described in Sec. 6.4.2 and denoted as *eICIC* in the results. As this combination was shown to perform almost identically to the LP-ABS solution, and much better than the two

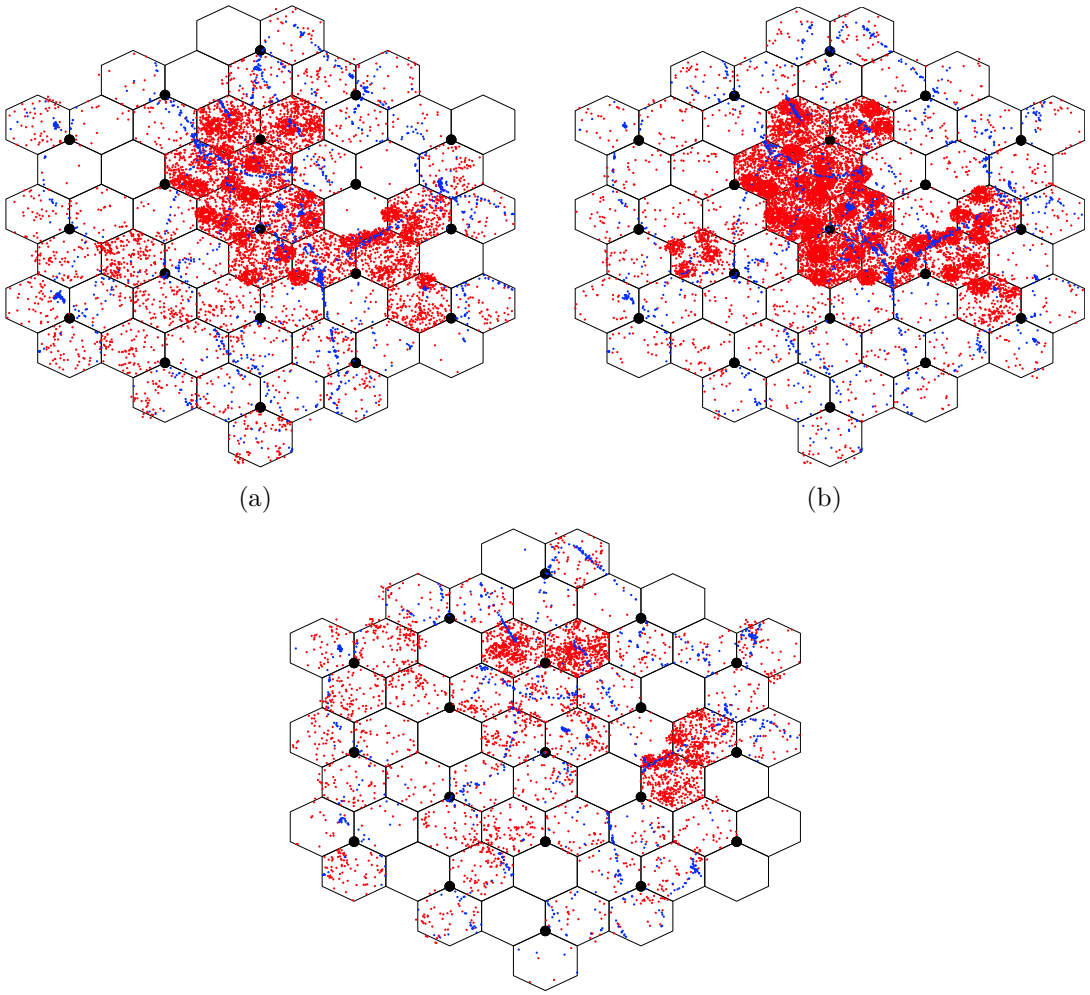


Figure 6.10. Snapshots of user distribution. The red dots represent pedestrian UEs, while the blue dots represent vehicle UEs. (a) Morning; (b): Afternoon; (c): Evening.

fixed strategies, it was chosen as the only baseline solution in this scenario. Note that this combination is most widely used in literature and applied in practice as well.

6.4.5 Numerical results

BPS behaviour. First, we take a look at the power setting strategies that the BPS algorithm produces. In Fig. 6.11, we depict the averaged strategies reached through the BPS algorithm during the simulation period for the morning scenario. The

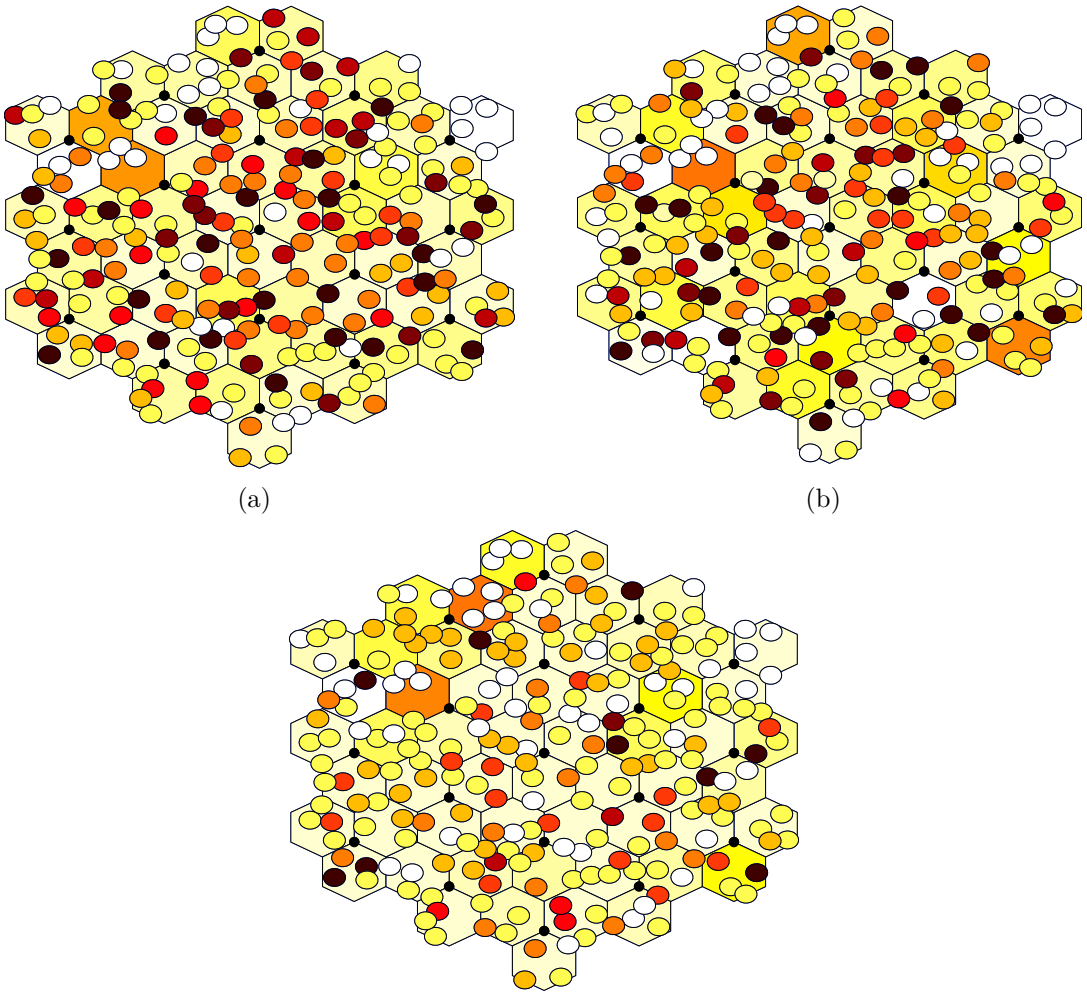


Figure 6.11. BPS achieved power strategy for the morning scenario. (a) CC1; (b) CC2; (c) CC3.

strategies chosen by the teams for each CC are differentiated using different shades, from white (*zero* power) to black (*maximum* power). Recall that the maximum power varies depending on the type of PoA. Hexagons represent the macro PoAs, while circles represent the micro PoAs. The figure shows that CC1, i.e., the high frequency carrier, allows for higher transmit power levels to be used by both macro and micro PoAs, due to its low interference impact. CC2 and especially CC3 are used to complement each other to ensure overall coverage. Average transmit power levels selected during the simulation period, for the the three different times of the day, are shown in Fig. 6.12 for Macro and Fig. 6.13 for Micro PoAs. In general we see that low transmit power levels are preferred for macro PoA across all CCs. In highly concentrated areas such as the city centre and commercial areas, the

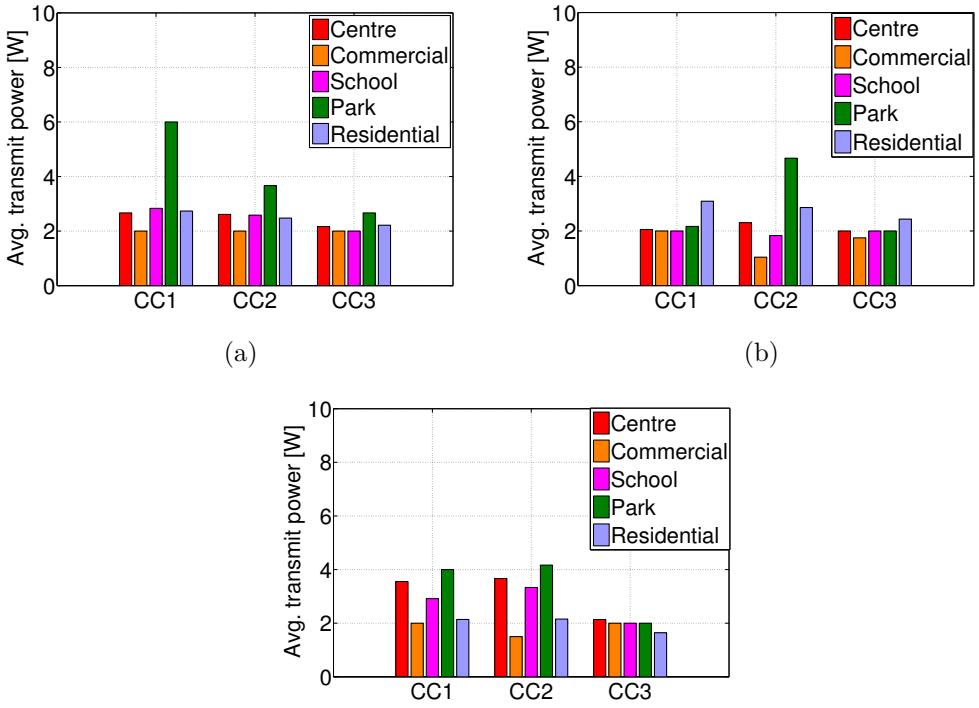


Figure 6.12. Average downlink transmit power selected by the BPS algorithm for Macro PoAs, in the different urban areas. (a) Morning; (b) Afternoon; (c) Evening.

Table 6.2. UE densities and cell request arrival rates

	City centre	Comm. area	School	Park	Res. area
Baseline UE density [UE/msq]	0.0245	0.0147	0.0074	0.0009	0.0009
Percentage of vehicles	30%	5%	5%	5%	50%
Density weights					
Morning (7-9 AM)	0.5	0.6	0.6	0.8	0.8
Afternoon (3-5 PM)	1	0.95	0.95	0.7	1
Evening (10 PM-12 AM)	0.08	0.5	0.01	0.5	0.6
Cell request arrival rates λ					
Morning (7-9 AM)	0.75	0.54	0.27	0.04	0.04
Afternoon (3-5 PM)	1.5	0.9	0.4	0.03	0.05
Evening (10 PM-12 AM)	0.12	0.45	0.005	0.02	0.03

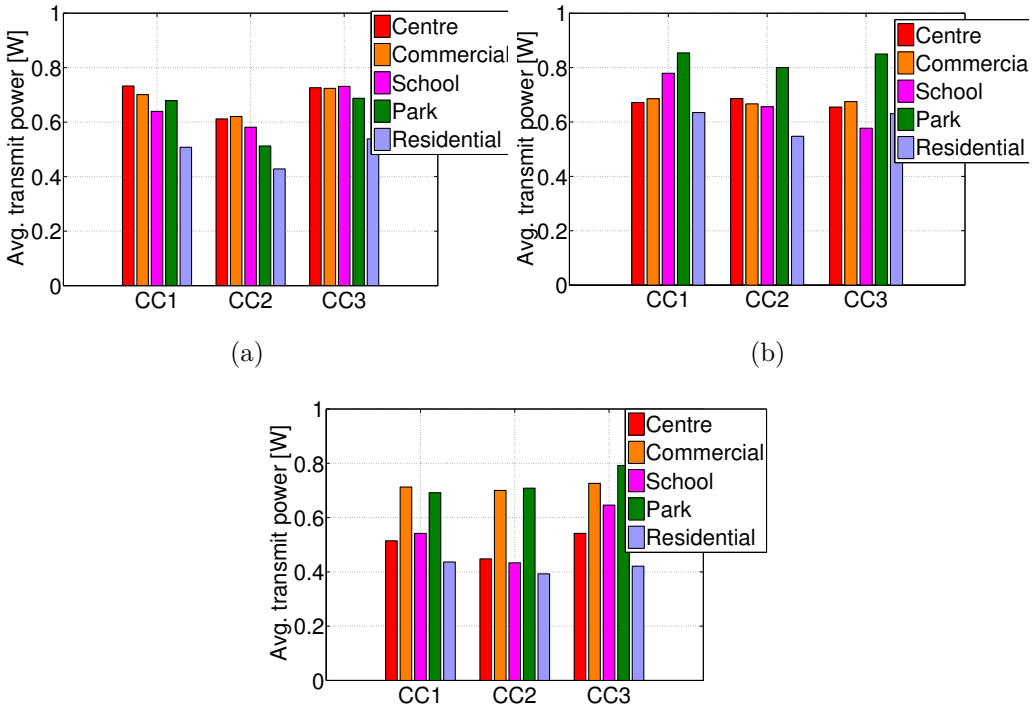


Figure 6.13. Average downlink transmit power selected by the BPS algorithm for Micro PoAs, in the different urban areas. (a) Morning; (b) Afternoon; (c) Evening.

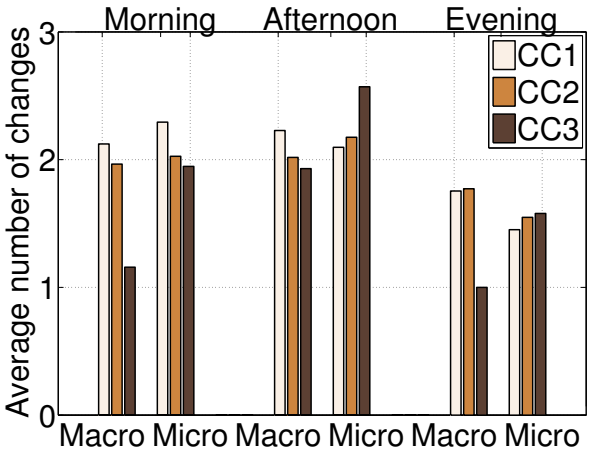


Figure 6.14. Average number of strategy changes during the simulation period.

micro PoAs tend to transmit at higher power levels, while macro PoAs at lower power levels, especially during morning and afternoon. Such strategies enable micro

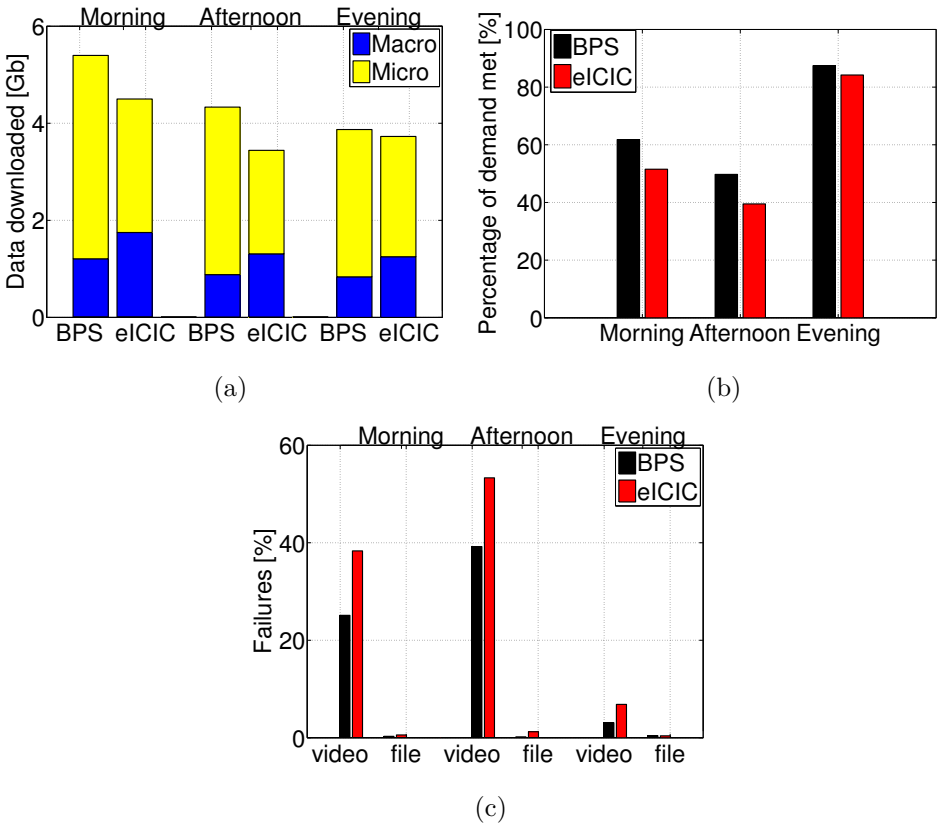


Figure 6.15. (a) Total amount of content downloaded over the network; (b) Percentage of demand met; (c) Percentage of failed downloads depending on content type.

PoAs in these areas, who support most of the traffic demand, to transmit using a higher modulation coding scheme (MCS), which in turn implies higher bit rate and therefore throughput. In residential areas instead, traffic demand is lower and more spatially spread; thus, it is the macro PoAs that serve most of the traffic demand and therefore need to use higher power.

In Fig. 6.14 we have plotted the average number of strategy changes, i.e., switching from one power level to another, enacted by BPS during the simulation period for Macro and Micro PoAs. Considering that the simulation period was 1200 ms long, and the BPS update frequency 100 ms, this implies that BPS algorithm was performed 11 times. Out of these 11 times, the average number of strategy changes per PoA is around 2 for both Macro and Micro PoAs, during morning and afternoon, which are the busier times of the day. This value is smaller, around 1.7, during the evening when the traffic load is much lighter and more spatially spread. This indicates that BPS is responsive to changes caused by short-term user mobility, but it is not overly sensitive to such variations in user distribution. For Macro PoAs,

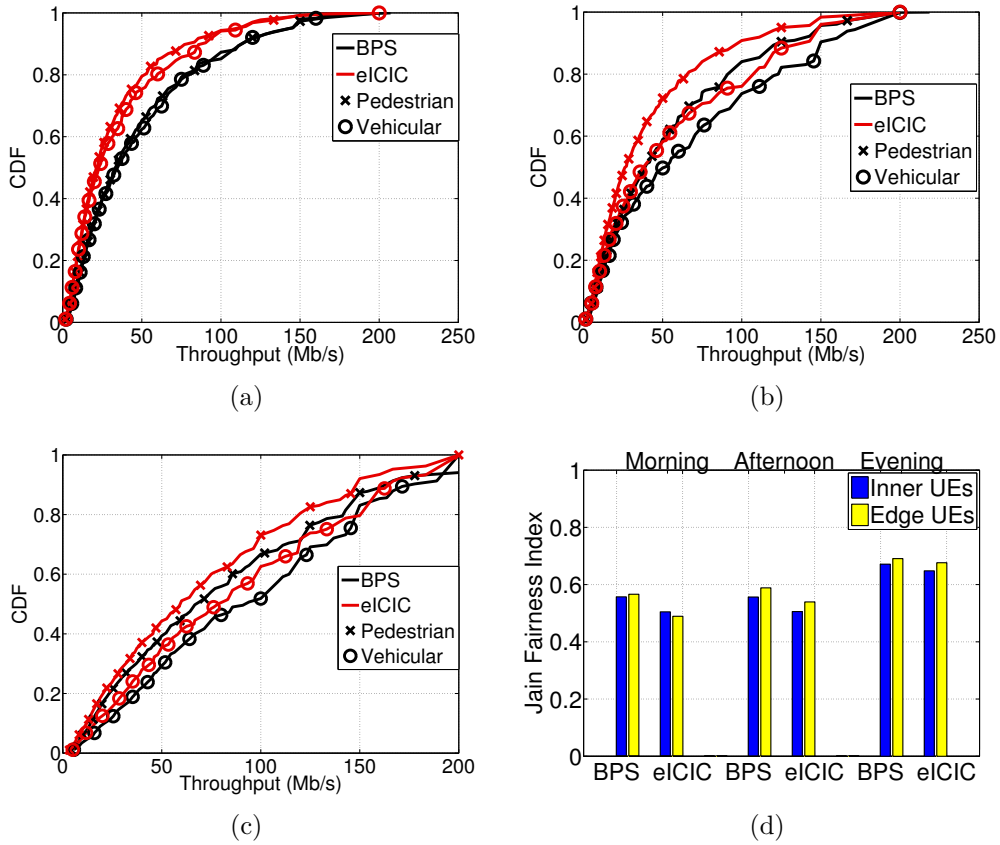


Figure 6.16. CDF of the average user throughput achieved by pedestrian and vehicular UEs. (a) Morning; (b) Afternoon; (c) Evening. (d) Fairness between inner and edge UEs in terms of average user throughput (Jain index).

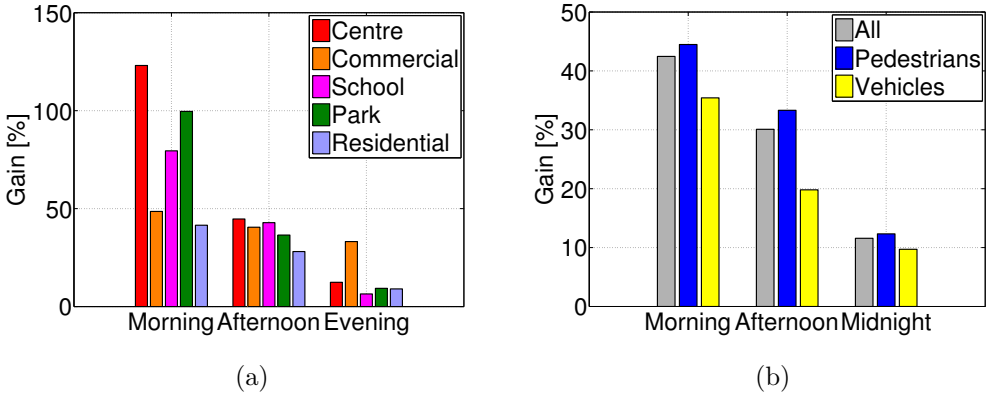


Figure 6.17. Gains over eICIC in terms of average throughput: (a) For the different urban areas; (b) For vehicular and pedestrian UEs

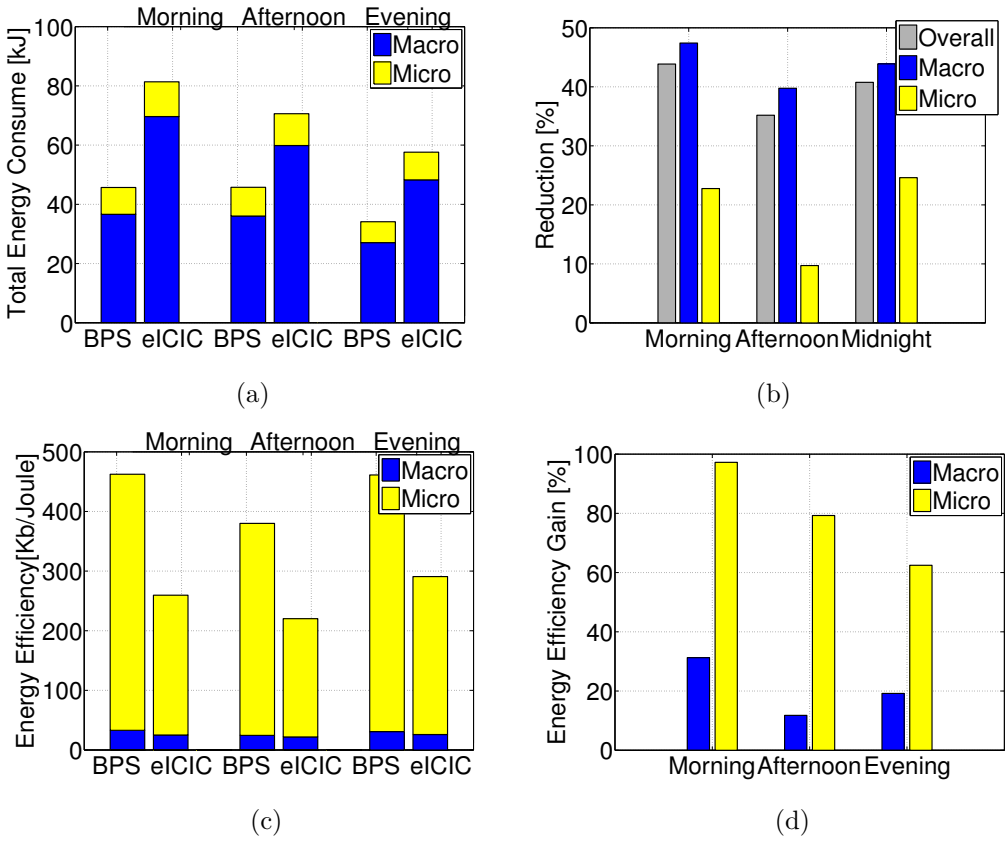


Figure 6.18. (a)Total energy consumed; (b) Reduction in energy consumed; (c) Energy efficiency (expressed in Kb/Joule) achieved by the PoAs at different times of the day (d) Gain in energy efficiency.

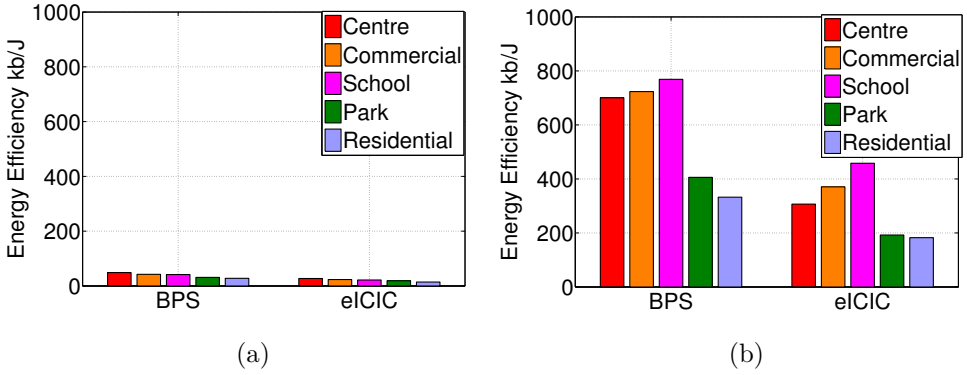


Figure 6.19. Energy efficiency for the different areas, morning scenario: (a) macro PoAs; (b) micro PoAs.

we notice that the average value is slightly lower for CC3; this combination is less affected by user mobility as it is usually applied to ensure overall coverage.

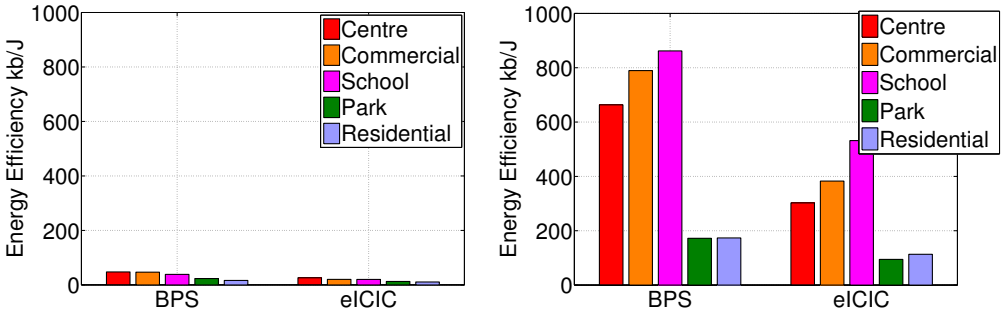


Figure 6.20. Energy efficiency for the different areas, afternoon scenario: (a) macro PoAs; (b) micro PoAs.

In the following plots we show how the dynamically obtained power strategies outperform eICIC in some of the main performance metrics.

Level of service and fairness. Fig. 6.15 (a) shows that when BPS is employed the amount of data downloaded over the network is always higher, especially during high intensity periods like morning and afternoon. BPS also improves the service experienced by the UEs in terms of demand met and percentage of failed downloads, as shown in Fig. 6.15 (b) and (c). Note that for each file we have set a specific deadline, within which we expect the download to complete, otherwise it is considered a failed download. Specifically, we consider a deadline of 0.5 seconds for videos and 1 second for generic files. In Fig. 6.15 (b), we show that during intensive periods, BPS improves the percentage of demand met across the entire network by around 10%, and reduces the rate of failed downloads by approximately 15%. It is clear that as the traffic load intensifies in certain areas, which is the case in the afternoon scenario, both approaches have difficulties in managing the demand, however BPS does ensure an improvement, especially for video content, without applying any intelligent content-aware resource allocation techniques.

In Fig. 6.16 we look at the cumulative distribution function (CDF) of the achieved average user throughput at the different times of the day, differentiated for vehicular (circle) and pedestrian (cross) UEs. Note that in general BPS offers higher average user throughput for both types of traffic but the improvement compared to eICIC, is more significant during peak hours. This is true especially for pedestrian UEs, who are concentrated in the high density areas with heavy traffic load. While it may look counterintuitive, vehicular UEs tend to have better average throughput. That reason is that most of the vehicle UEs tend to be spread in the residential area where the traffic demand is lower and, they tend to be situated in well covered areas. These two factors influence the performance more than the fast-fading effects.

In Fig. 6.16 (d) we show the level of fairness between inner and edge UEs in terms of average user throughput, by calculating the Jain fairness index. While the level of

fairness for inner and edge UEs tends to be the same, there is a modest improvement for both categories when BPS is applied. Note that in multi-tier networks with high density of small cells the line between inner and edge UEs tends to blur, as edge UEs under the coverage of a micro PoA may experience even better conditions than inner UEs. As a result, the average throughput may vary greatly between UEs of the same category. BPS, however is able to improve the fairness by limiting the overall interference.

Finally, Fig. 6.17 depicts the gains obtained by using BPS, compared to eICIC, in terms of average user throughput. We can observe that in the different urban areas significant gains are shown, especially during morning and afternoon. In particular, in highly dense areas such as the city centre the gains can be well over 100%. The gains are lower during evening when the traffic load lightens and disperses over the entire area. Overall, gains in the average user throughput are slightly above 40% during morning, which decreases to around 10% during evening. Pedestrian UEs in general experience higher gains, which is significant as they account for the majority of UEs.

Energy efficiency. While in terms of amount of data and average throughput, the difference is smaller in the evening when the traffic load decreases significantly, BPS still retains a considerable edge in energy efficiency (see Fig. 6.19). This is because BPS is able to serve higher amounts of data, while consuming significantly less power. From Fig. 6.19 (a) it is clear that BPS, put simply, consumes considerably less energy. The overall reductions in energy consumption range from 35% to 45%, irrespective of the time of the day. The reduction is even more pronounced for Macro PoAs; as we mentioned Macro PoAs tend to choose lower transmit power across all CCs, which accounts for the significant reduction in consumption. It is clear therefore that since BPS is able to consume less energy while delivering larger amounts of data, that it outperforms eICIC in terms of energy efficiency for both macro PoAs and micro PoAs. As we see in Fig. 6.19(c) the effect is more significant for the latter: the gain in energy efficiency for macro PoAs varies between 15 and 20%, while for micro PoAs it can be as high as 100% during the morning and it drops to around 60% in the evening. BPS tends to choose lower transmit powers for macro PoAs, especially for dense areas with heavy traffic load, which significantly reduces the interference experienced by micro PoAs who are responsible for serving the bulk of the data. Indeed, if we look at the energy efficiency values for the different areas, shown in Fig. 6.18-6.21 for the different times of the day, it is clear that energy efficiency is highest in the city centre, commercial and school areas where the traffic load is more intense. Again, this is true for both macro and micro PoAs, but it is more significant for the latter.

RB usage efficiency. Fig. 6.22 shows the RB usage efficiency for macro (a) and micro PoAs (b) calculated in terms of kilobits transmitted per number of RBs used. Note that this metric takes into account only those RBs allocated to UEs, not

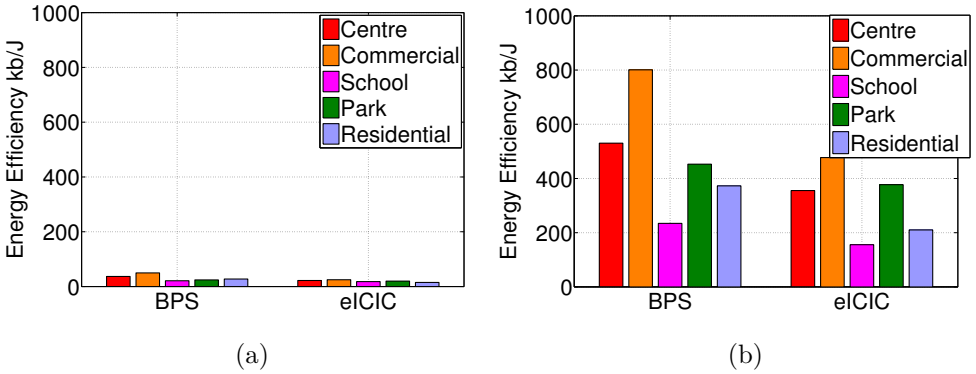


Figure 6.21. Energy efficiency for the different areas, evening scenario: (a) macro PoAs; (b) micro PoAs.

the overall number of RBs available. Again, BPS improves the performance of the network for all types of PoAs, but more significantly for micro PoAs, especially in the morning and afternoon. The gains are shown in Fig. 6.22(c) and (d). We see that for Micro PoA during these times of the day can range between 70% and 80%. During evening, the gains drop significantly; during this time of the day, users are more evenly distributed in residential areas, hence, most of the service is offered by Macro rather than Micro PoA. It should be noted that eICIC itself introduces important improvement in this metric especially for macro PoAs by offloading some of their UEs to the micro PoAs, however BPS provides an additional edge while lowering the overall power consumption, as seen in the previous figures. For micro PoAs, BPS improves this metric significantly by strategically varying the transmit power of the different macro PoAs to reduce the overall interference. The performance of micro PoAs is further improved by the fact that the power setting of the micro PoAs within the same cell is decided at the team level, ensuring optimal coordination in terms of interference. It is worth noting that BPS could also be applied jointly with eICIC, especially to take advantage of the CRE feature.

Varying the number of microcells. In Fig. 6.23 (a), we look at the improvement obtained by applying BPS when compared to eICIC alone, for different network configurations with a varying number of microcells within each cell. The improvement in the three core metrics: energy efficiency, average user throughput and RB usage efficiency, tends to be significant and consistent as the number of microcells is increased.

Increasing the maximum transmit power. n Fig. 6.23 (b), we also show the gains achieved in the same core metrics, when we consider a higher maximum power for macro PoAs, i.e., 46 dBm (40 W), which is foreseen for 5G systems [13], instead of 43 dBm (20 W), which we typically assume. As expected, the effect of BPS is increased when the maximum power of the macro PoAs is elevated, since the

effects of interference, which BPS effectively mitigates, are even more pronounced.

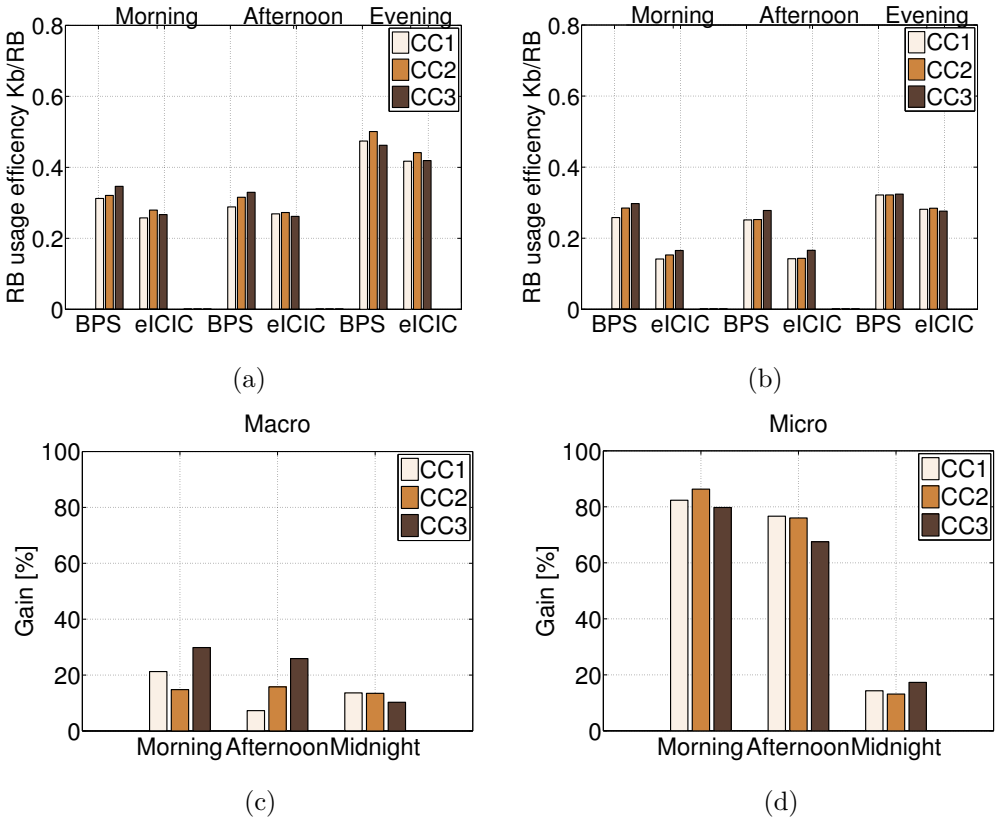


Figure 6.22. RB usage efficiency expressed in Kb transmitted per RB. (a) Macro PoAs; (b) Micro PoAs; (c) Gain in RB usage for Macro PoAs; (d) Gain in RB usage for Micro PoAs.

6.5 Conclusions

Given the devastating effects interference will have in future networks as they become more dense and heterogeneous, effective means to contain and mitigate it will be key to enabling the optimal use of resources. In this chapter, we proposed a novel solution for downlink power setting in dense networks with CA, which aims to reduce interference and power consumption, and to provide high quality of service to users. Our approach leverages the different propagation conditions of the carriers and the different transmit powers that the various types of PoAs in the network can use for each carrier.

Applying game theory, we framed the problem as a competitive game among teams of macro and micro PoAs, and identified it as a game of strategic substitutes/complements with aggregation. We then introduced a distributed algorithm

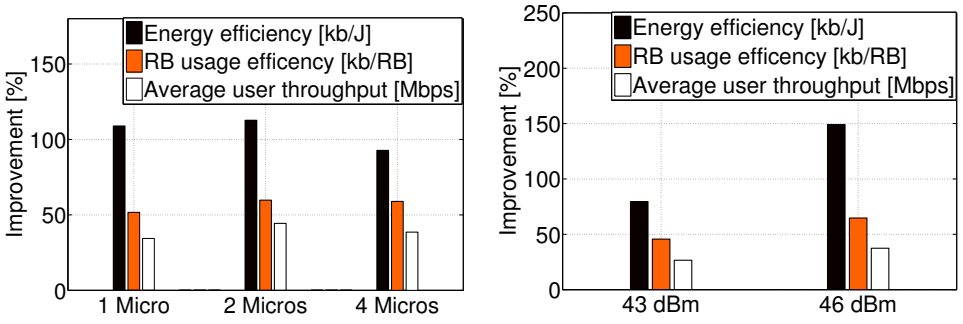


Figure 6.23. Improvement obtained using BPS over eICIC in three core metrics: energy efficiency, RB usage efficiency, and average user throughput, for a varying number of micro PoAs within a cell (a) and for different maximum transmit power for macro PoAs (b).

that enables the teams to reach a desirable NE in very few iterations. Simulation results, obtained in a realistic large-scale scenario, show that our solution greatly outperforms the existing strategies in the main performance metrics, such as energy efficiency, user throughput and spectral efficiency, while consuming little power.

Chapter 7

Conclusions and future work

In this thesis we addressed resource and power management aspects of next generation cellular networks. In particular we looked at dense heterogeneous networks with multi-tiered architecture and support for advanced technologies such as D2D communications and carrier aggregation. As we noted, such networks suffer from severe interference, especially in scenarios where radio resources are fully shared and base stations transmit at full power indiscriminately. Therefore we proposed resource allocation techniques and transmit power algorithms which alleviate the interference, and in turn significantly improve overall network performance.

We started by addressing the resource allocation problem in a two-tier heterogeneous networks with support for D2D links. We used a dynamic programming model to address the problem, and proposed a centralised scheme to allow the nodes in the network to make interference and traffic aware resource allocation decisions. To reduce the complexity of the centralised problem, we applied approximate dynamic programming principles to control and limit the size of the solution space. Thus we were able to swiftly reach resource allocation decision which are enacted every subframe, i.e., 1 ms. It should be noted that the algorithm we proposed takes into consideration both interference and traffic characteristics when making scheduling decisions. Through extensive simulations we evaluated the performance of our approach in large-scale realistic scenario comprising a two-tier heterogeneous network. The approach we proposed shows significant promise in enhancing network performance when compared to state-of-the-art solutions which couple interference mitigation techniques such as eICIC with standard scheduling algorithms such as *proportional-fair* (PF). Our numerical results showed that our approach significantly improves network throughput, quality of service and energy and spectrum efficiency. Another interesting finding was the advantage of accommodating D2D communications in the uplink or in the downlink bandwidth. Under current traffic load conditions, the two options are equally effective. However, as upload and download traffic tend to even out, using the downlink bandwidth will be preferable. Finally,

D2D was found to be an effective, low-energy and low-cost replacement for microcell deployment.

Next, we addressed the resource allocation problem in a carrier-aggregation enabled heterogeneous dense network. Due to the considerable increase in complexity, instead of a centralised approach, we proposed a heuristic distributed algorithm to solve the resource allocation problem. This algorithm too takes into account traffic and interference to produce interference-aware scheduling decisions. In general, the algorithm assigns *urgency* values to the different traffic flows, and *pollution* values to the available links, to rank and sort potential combinations of traffic flows, serving cells and radio resources. It then aims at scheduling the most urgent flows on the least toxic links for improved network performance. We evaluated the performance of the algorithm, on the same large-scale two-tier network, and we showed that the heuristic algorithm we proposed significantly outperforms the eICIC and PF combination. The algorithm is especially apt at fully taking advantage of the availability of the multiple carriers in the network to better distribute the traffic load, and improve throughput and spectral efficiency by reducing overall interference.

Finally, we proposed a downlink transmit power management scheme to alleviate the problem of interference in dense multi-tiered network, by leveraging the diversity offered by the multiple carriers which can be used with carrier aggregation. Using game theory principle we modelled the problem at hand as a competitive game between teams of PoAs within the coverage area of a macrocell so we could capture both (i) cooperation between the macrocell and the small cells with overlapping coverage areas, and (ii) the competitive interests of different macrocells. The framework we provided however allowed for straightforward extension to teams that include several macrocells. We proved that the game we model belongs to the class of pseudo-potential games, which are known to admit pure Nash Equilibria (NE). This allowed us to propose a distributed algorithm based on best-reply dynamics that enables the network to dynamically reach an NE representing the preferred solution in terms of throughput, user coverage and power consumption. As shown by simulation results, our scheme outperforms fixed transmit power strategies, even when advanced interference mitigation techniques such as eICIC are employed. Simulation results, obtained in a realistic large-scale and dynamic scenario, showed that our solution greatly outperforms the existing strategies in the main performance metrics, such as energy efficiency, user throughput and spectral efficiency, while consuming little power.

In terms of future work, with respect to the last part of the thesis, we plan to further investigate aspects of team composition and size for optimal game behaviour. While the algorithm we propose is applied by the individual teams in an independent fashion, which PoAs to group in a team and what is the optimal team size are still open questions. Further analysis of how the game is affected by incomplete or erroneous sharing of information between teams would also be of interest, as it

pertains to challenges often encountered in real cellular networks.

As we mentioned in the beginning, expanding to mmWave bands seems to be an inevitable in the evolution of cellular networks, as they continue to cater to a growing number of spectrum hungry applications. Such an expansion represents a real paradigm shift, as integration of mmWave communications will have far-reaching implications, particularly in aspects of radio resource and power management and higher layer tasks as well. Due to their specific signal propagation characteristics which require constant line-of-sight conditions during transmissions, mmWave integration will represent a range of challenges which we plan to tackle in future work.

Appendix A

PhD Activity Report

==========

SCHEDA INFORMATIVA SULL'ATTIVITA' SVOLTA NEL DOTTORATO

IN INGEGNERIA ELETTRONICA E DELLE COMUNICAZIONI

DA Zana Limani Fazliu

- PhD Ciclo/ course 28 Anni accademici di riferimento/Years 4
- Department of Electronics and Telecommunication
- Coordinatore: Prof. Ivo Montrosset
- Tutore/ Supervisor : Prof. Carla-Fabiana Chiasserini
- Titolo della Tesi di Ricerca / PhD thesis title (nella lingua in cui viene scritta)

Resource allocation and power management in next generation cellular networks

A. Descrizione dell'argomento della tesi/ Topic of the thesis (massimo 20 righe/ max 20 rows)

Densification of wireless cellular networks, by overlaying smaller cells over the traditional macrocell, is seen as an inevitable step in enabling future networks to support the expected increase in data rate demand. Moving towards 5G, networks will become more heterogeneous as services will be offered via various types of points of access (PoAs). Indeed, besides the traditional macro base station, it is expected that users will also be able to access the network through WiFi access points, small cell (i.e., micro, pico and femto) base stations, or even other users when device-to-device communications are supported. This approach will improve both the capacity and the coverage of current cellular networks. However, since the different PoAs are expected to fully share the available radio resources, inter-cell interference as well as the interference between the different tiers will pose a significant challenge. Future networks are also expected to support carrier aggregation (CA), which allows the simultaneous use of several component carriers (CCs), in order to guarantee higher data rates for end users. Device-to-device communications are also expected to be widely supported in future networks. We propose to leverage the diversity of the advanced features to mitigate inter-tier interference while enabling a wide range of network configurations which reduce power consumption, provide high throughput and ensure a high level of coverage to network users. An interference-aware heuristic algorithm that jointly performs carrier selection and resource allocation is also proposed, to improve spectrum utilization, while also ensuring better fairness between CA-enabled and legacy user terminals.

B. Attività di ricerca svolta nel triennio/ Research activity during the whole PhD.

B.1 **descrizione** complessiva e sintetica dell'attività di ricerca/ synthetic description of the research activity

The exponential increase in mobile data traffic in recent years has become a serious challenge for today's cellular communication networks. To tackle this challenge, one of the strategies foreseen in the LTE-Advanced (LTE-A) and 5G specifications, among others, is the densification of the network infrastructure both by diversifying the types of points of access (PoAs) and by increasing their density. Dense networks are seen as a potential cost-efficient approach to effectively meet the challenge, by introducing smaller cells, i.e., micro, pico and femtocells, nested within the traditional macrocell and allowing direct device-to-device (D2D) communications. This approach promises to improve both the capacity and the coverage of current cellular networks. However, it also introduces several technical challenges, the most prominent being the interference between different architectural layers sharing the same spectrum resources.

Carrier aggregation is another expected feature of future networks, which aims at guaranteeing higher data rates for end users so as to meet the IMT-Advanced requirements. It enables the concurrent use of several LTE component carriers with, potentially, different bandwidth and belonging to different frequency bands. Downlink transmissions over each carrier will occur at maximum output power and each carrier will have an independent power budget. Thus, different component carriers may have very different coverage areas and impact in terms of interference, due to both their different transmit power level and propagation characteristics.

To tackle the challenges that future networks will face as well to ensure optimal use of advanced features such availability of heterogeneous cells, CA and D2D, during our research we have focused closely on two topics: radio resource allocation in such networks and downlink power management for the component carriers. The goal is to improve the performance network in several key metrics, such as overall network throughput, energy-efficiency, spectrum utilization efficiency and fairness. The approach we take to address these issues and reach the above-listed goals, is described in more detail next.

B.2 **argomenti di ricerca specifici affrontati/ specific research topics covered**

- **Interference-aware resource scheduling in LTE heterogeneous networks with carrier aggregation support**

  Optimal resource allocation in LTE networks is known to be a hard problem, and is further exacerbated when support for advanced features such as heterogeneity and carrier aggregation are also considered. In our research, we address the resource allocation problem in dense heterogeneous networks that support carrier aggregation with backward compatibility for legacy user terminals which do not support carrier aggregation. Unlike previous work, we look at the resource allocation problem by tackling the two main problems affecting these kind of networks: inter-cell interference, and the complexity imposed by the availability of multiple carriers with potentially very different propagation and coverage characteristics. To this end we propose a solution that jointly addresses carrier selection and resource allocation, while taking into account interference, in order to fairly serve CA-enabled and legacy user terminals.

  The aim of the proposed algorithm is to enable the controller to decide at each LTE subframe (i) which point of access (PoA) should serve each traffic request, i.e., flow, (ii) on which available carriers and (iii) which resources to employ for such communication. To achieve this, the algorithm assigns an urgency value to every active traffic flow, which depends on the characteristics of the requested traffic, such as data size and delivery deadline. Additionally, for each potential resource allocation decision, it calculates a pollution value that accounts for the interfering impact a specific decision may have on the system. Using these parameters, the algorithm produces a resource allocation strategy, indicating which PoA and carrier are chosen to serve which traffic flow on which resources.

We evaluate the performance of our approach in a large-scale scenario and compare it with other widely used heuristic algorithms such as Proportional-Fair scheduling and Enhanced Inter Cell Interference Coordination (eICIC) techniques. Simulation results show that the solution we propose increases system throughput, minimises energy consumption and improves spectrum utilisation, while also ensuring better fairness between CA-enabled and legacy user terminals.

- **Resource allocation in D2D enabled LTE heterogeneous networks**

  During the PhD activity, a part of the research was devoted to addressing the challenges of resource allocation in D2D enabled heterogeneous networks. We proposed an interference-aware resource scheduling algorithm for an LTE-based, two-tier heterogeneous networks with D2D support. We consider that D2D will take place within the LTE bands, in what is often called "in-band underlay" mode, where D2D opportunistically accesses the same spectrum resources used by the other nodes in the cellular network. In principle, D2D communications can take place in either the uplink or the downlink resources. Currently, it is widely accepted that uplink resources should be used, since, at present, traffic is significantly heavier in downlink than in uplink. However, it is expected that in the future traffic will be much less asymmetric, then the use of downlink resources will represent a viable option. In both scenarios, D2D can cause significant interference to normal infrastructure-to-device communications, either to nearby receiving devices when implemented in the downlink bands, or to nearby receiving PoAs when deployed in the uplink bands. Without proper management of this interference, D2D communication may easily end up doing more harm than good.
  We therefore address and compare both D2D scenarios, and propose a resource allocation procedure based on *approximate dynamic-programming*. The procedure itself is adaptable to both downlink and uplink D2D scenarios, it is updated every subframe and is efficient enough to be applied to large-scale scenarios. The performance of our approach is numerically evaluated and compared to standard resource scheduling algorithms adopted in today's cellular networks, employing interference mitigation techniques. Results highlight that the proposed approach is apt at fully exploiting the potential of both the heterogeneity of the network and D2D support, while consuming far less energy. Results further reveal that D2D interactions act inherently as an additional layer in the heterogeneous network, thereby potentially reducing the need for deploying more microcells. Finally, while the uplink and downlink scenarios provide similar performance in current traffic load conditions, the downlink will become preferable as the upload and the download traffic tend to even out.

- **Downlink Power Setting in Carrier-Aggregation Enabled Dense Networks**
  Given the proven benefits cell densification brings in terms of capacity and coverage, it is certain that 5G networks will be even more heterogeneous and dense. However, as smaller cells are introduced in the network, interference will inevitably become a serious problem as they are expected to share the same radio resources. We argue that by exploiting the diversity of the different carriers, CA can be used to effectively mitigate the interference in the network. In our research we leverage the key features of next-generation cellular networks and formulate a downlink power setting problem for the different available carriers.

  Currently, three main approaches have been proposed to address the interference problem in dense networks: per-tier assignment of carriers, Enhanced Inter Cell Interference Coordination (eICIC), which has been adopted in LTE-A systems, and downlink power control. Per-tier assignment of carriers simply implies that in CA-enabled networks, each tier should be assigned a different CC so as to nullify inter-tier interference. eICIC includes techniques such as Cell Range Expansion (CRE) to incentivise users to associate with microcells, and Almost Blank Subframes (ABS), i.e., subframes during which macrocells mute their transmissions to alleviate the interference caused to microcells. Algorithms to optimise biasing coefficients and ABS patterns in LTE heterogeneous networks have been studied, however they do not address CA. Also, modifications to the eICIC techniques that

allow macro base stations to transmit at reduced power during ABS subframes have also been proposed. In our research we do not consider a solution within the framework of eICIC or its modifications, rather we use them as comparison benchmarks for the solutions we propose.
We adopt instead the third approach, which consists in properly setting the downlink transmit power of the different CA-enabled PoAs so as to avoid interference between different tiers. We propose to leverage the diversity in the component carrier coverage areas to mitigate inter-tier interference by varying their downlink transmit power. Thus, we enable a wide range of network configurations which reduce power consumption, provide high throughput and ensure a high level of coverage to network users.

As envisioned in LTE-A systems we consider that each CC at each PoA has an independent power budget, and that PoAs can choose the transmit power on each carrier from a discrete set of values. Therefore, our goal is to adequately choose a power level from a range of choices to ensure optimal network performance. It is easy to see that the complexity of the problem increases exponentially with the number of cells, CCs and the granularity of the power levels available to the PoAs. In addition, if one of the objectives is to maximise the network throughput, the problem becomes non linear since transmission data rates depend on the signal-to-interference-plus-noise ratio (SINR) experienced by the users. It follows that an optimal solution requiring a centralised approach would be both unfeasible and unrealistic, given the large number of cells in the network.
We therefore study the above problem through the lens of game theory, which is an excellent mathematical tool to obtain a multi-objective, distributed solution in a scenario with entities (PoAs) sharing the same pool of resources (available CCs). We model each group of PoAs in the coverage area of a macrocell as a team so that we can capture both (i) cooperation between the macrocell and the small cells with overlapping coverage areas, and (ii) the competitive interests of different macrocells. The framework we provide however allows for straightforward extension to teams that include several macrocells. We prove that the game we model belongs to the class of pseudo-potential games, which are known to admit pure Nash Equilibria (NE) . This allows us to propose a distributed algorithm based on best-reply dynamics that enables the network to dynamically reach an NE representing the preferred solution in terms of throughput, user coverage and power consumption. As shown by simulation results, our scheme outperform fixed transmit power strategies, even when advanced interference mitigation techniques such as eICIC are employed.

B.3 **risultati più rilevanti ottenuti nel triennio/most relevant results**

Samples of the most relevant results obtained during the PhD research activity are listed below:

1. An interference-aware algorithm based on approximate dynamic programming (ADP) principles was proposed, to perform the complex problem of resource allocation task in LTE heterogeneous networks with D2D support. While we cannot prove that ADP can provide the optimal resource allocation strategy, our numerical results show that the performance of this algorithm is at least *near-optimal.* As we already mentioned, D2D communications can take place in either the uplink or the downlink resources. Therefore, in our research we addressed both scenarios: the Downlink (DL) scenario and the Uplink (UL) scenario. The algorithm proposed was applicable without modifications in both scenarios. It should also be noted that the resource allocation applied to both downlink and uplink resources in both scenarios. A flow-chart of the algorithm is shown below:

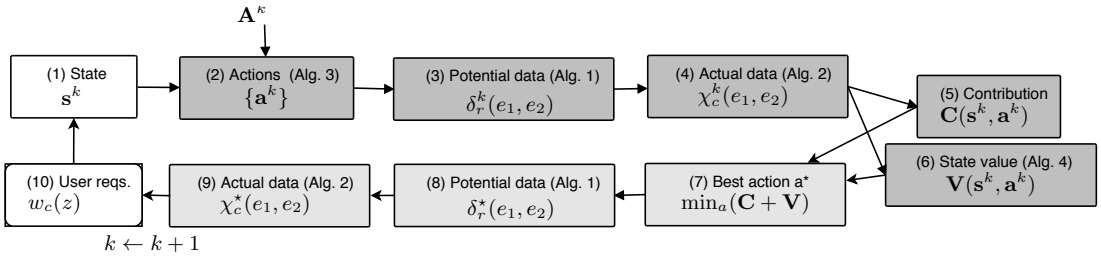


Fig. 1 Detailed view of the algorithm. Given the current state (1), the set of possible actions can be determined (2). For each action, the potential (3) and actual (4) amount of content transferred between the pairs of endpoints can be computed. These values are further used to compute the cost (5) of an action, and to estimate the value of the state it leads to (6). The latter two figures are used (7) to select the best action. The resulting data transfers (8-9), along with the users that just became interested in a content, define the next state.

The algorithm is explained in detail in three of our published papers:

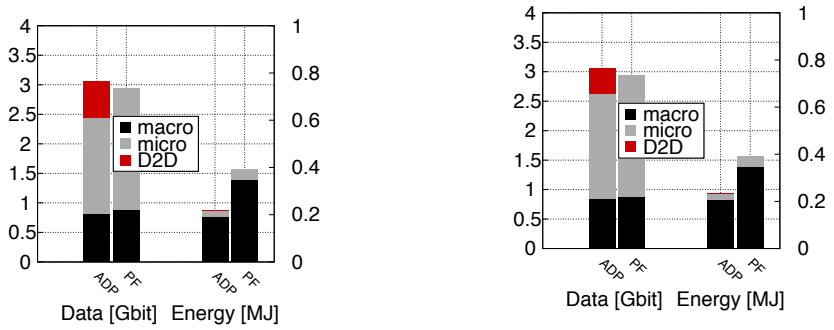
i. F. Malandrino, **Z. Limani,** C. Casetti, C.-F. Chiasserini, "Interference-Aware Downlink and Uplink Resource Allocation in HetNets with D2D Support," IEEE Transactions on Wireless Communications, vol. 14 n. 5, pp. 2729-2741, 2015.

ii. F. Malandrino, C. Casetti, C.-F. Chiasserini, **Z. Limani** "Fast Resource Scheduling in HetNets with D2D Support," IEEE INFOCOM, April 2014.

iii. F. Malandrino, C. Casetti, C.-F. Chiasserini, **Z. Limani** "Uplink and downlink resource allocation in D2D-enabled heterogeneous networks," IEEE WCNC, April 2014.

The performance of our approach was evaluated in the two-tier scenario that is typically used within 3GPP for LTE network evaluation. The scenario covered a service network area of 12.34 km2, covered by 57 macrocells and 228 microcells. Macrocells are controlled by 19 three-sector base stations with inter-site distance set to 500 m. Microcells are deployed over the network area, so that there are 4 non-overlapping microcells per macrocell. A total of 3420 users are present in the area. The approach we propose showed significant improvement in some key network performance metrics when compared to the application of the *Proportional Fair* algorithm (the standard algorithm used in LTE). Interference is managed by applying eICIC.

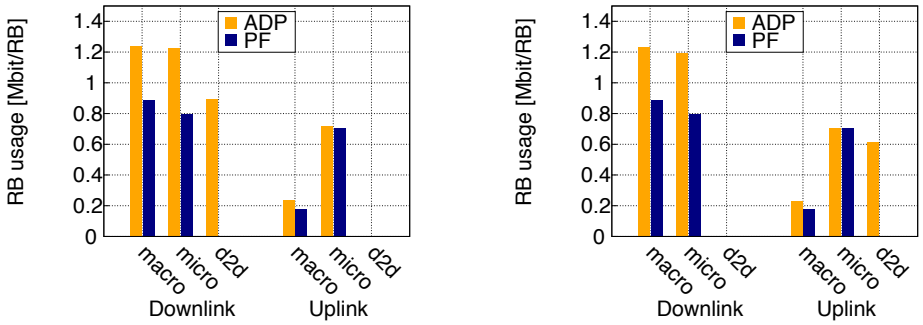
Specifically, in terms of overall data transferred over the network and energy consumed we obtained the results presented in Fig. 2 show that ADP is able to carry slightly more data over network than PF while consuming significantly less energy.



a) DL scenario b) UL scenario

Fig. 2 Total amount of transferred data and consumed energy

ADP also performs better in the spectrum utilization metric. In Fig. 3 we present the results for both scenarios. As can be observed ADP utilizes the spectrum much more efficiently at all network tiers, especially in the downlink resources.



a) DL scenario b) UL scenario

Fig. 3 RB usage expressed in Mbit transferred per RB.

Note: Results shown here were taken from paper iii).

2. An interference-aware heuristic algorithm was proposed for resource allocation in LTE heterogeneous networks with carrier aggregation support. The goal of the heuristic algorithm is to construct an interference and traffic-aware resource allocation strategy, in a way that efficiently schedules download traffic flows while ensuring backward compatibility and fairness for legacy users. To achieve this, the algorithm calculates and assigns several customized parameters such as *urgency* and *pollution* to each active traffic flow and possible strategy and then uses these values to produce an overall resource allocation strategy defined as a set of triplets indicating which serving cell (PoA-CC combination) is selected to serve which traffic flow on which resources. It is assumed that the algorithm is repeated every subframe and can performed in a centralised manner (wherein one area controller making scheduling decision for all the cells under its control), or in a distributed manner by grouping cell into clusters controlled by a cluster head. A flow chart of the algorithm is shown in Fig. 4. The detailed description, as well as discussions and results, can be found in the following paper:

i. Z. Limani, C.-F. Chiasserini, G. M. Dell'Aera "Interference-Aware Resource Scheduling in LTE HetNets with Carrier Aggregation Support", IEEE ICC June 2015.

The performance of the proposed approach was evaluated in the same realistic scenario, with a two-layer LTE network comprising of 57 macrocells and 228 microcells. We assume that there is a mix of CA-enabled and non-CA (legacy) users in the network and assume they move according to the cave-man model, with average speed of 1 m/s. We assume a transmitting power of 43 dBm for macrocells, and 30 dBm for microcells, equally distributed over all RBs, and antenna height values of 25 m and 10 m, respectively.
We considered a set of three carriers at the following frequency bands: 2.6 GHz (CC1), 1.8 GHz (CC2) and 800 MHz (CC3). All carriers have a bandwidth of 10 MHz band, hence each CC has 50 RBs. All CCs are simultaneously available at all BSs, while it is assumed that carrier-aggregation enabled are able to receive simultaneously from different cells on different CCs. Both implementations of the algorithm, the centralized and the distributed one are evaluated against the *Proportional Fair* algorithm coupled with eICIC.

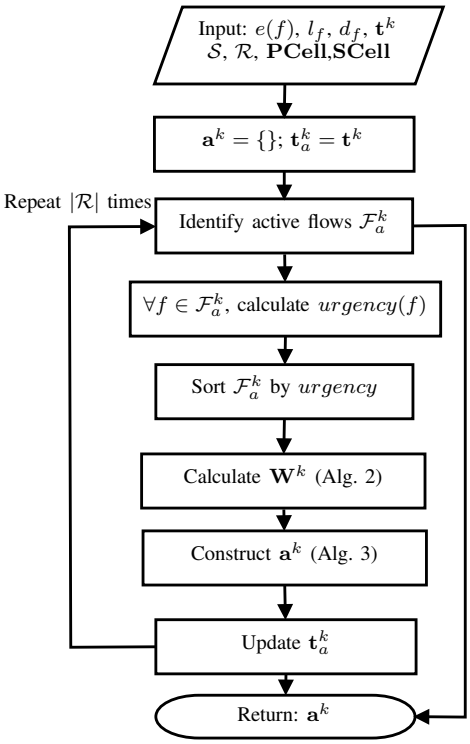
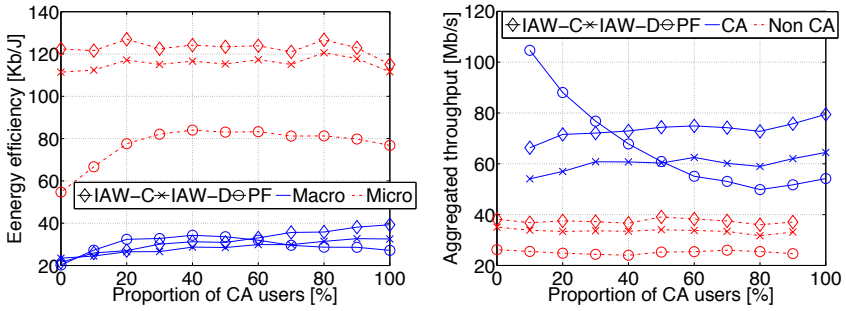


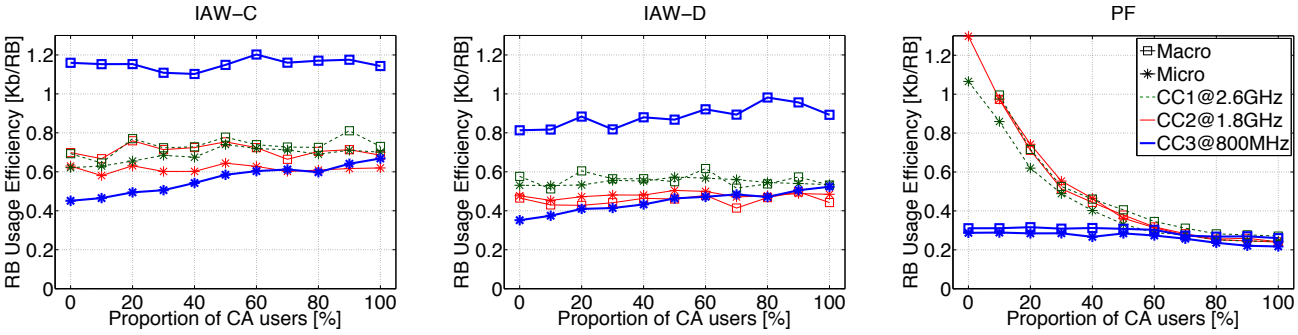
Fig. 4 Main algorithm: Building the allocation strategy

As shown in Fig. 5, the algorithm we propose denoted by IAW-C (centralized implementation) and IAW-D (distributed implementation) perform significantly better in terms of the energy efficiency metric measured in Kb transmitted per Joule consumed (Fig.5 a)). IAW considerably outperforms PF especially for microcells (red), which is quite significant since our experiments showed that microcells in our network are responsible for delivering the larger bulk of the data. The results show the value obtained for this metric as the percentage of the CA users in the network increases. Fig. 5(b) shows the average aggregated user throughput for CA-enabled and non-CA-enabled users. Again, IAW is able to ensure a constant average user throughput for both types of UEs. PF in contrast is able to ensure very high rates for CA UEs when they are a minority, and, again, it experiences a sharp drop when the number grows.



a) b)

Fig. 5. a) Energy efficiency measured in Kb/J for macrocells (blue/solid line) and microcells (red/dashed line); b) average aggregated user throughput in Mb/s for CA users (blue/solid line) and legacy users (red/dashed line). We differentiate between the different algorithms using different markers.



a) b) c)

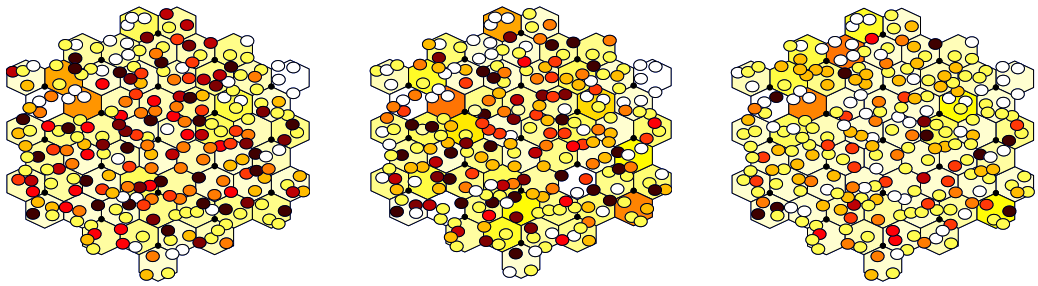
Fig. 6 RB usage measured in Kb/RB for each CC: (a) IAW Centralized; (b) IAW Distributed; and (c) Proportional-Fair. The legend for all there subfigures is shown in the top right-hand corner of subfigure (c). We differentiate between different types of BSs using different markers, and between different CCs using different line styles and colours.

Finally, in Fig. 6 we show the performance for the RB usage efficiency metric, which is measured in Kb transmitted per allocated RB during one subframe. We note that IAW-C, Fig. 6(a), is able to transmit the largest amount of data per allocated RB in CC3, both for macro and microcells. Note that this ratio is significantly higher for macrocells than for microcells, in the case of IAW- C and IAW-D. That is to be expected, since macrocells have higher transmit power and are used less (hence suffer less interference) than microcells, especially in combination with the low frequency CC3. The RB usage efficiency of CC1 and CC2 is practically independent of the number of CA users. On the contrary, Fig. 6(c), shows that PF exhibits a sharp drop in RB usage efficiency for CC1 and CC2, with the increase of CA users. Indeed, as the number of CA UEs grows, PF is unable to smartly allocate traffic flows over the various CCs, hence CC1 and CC2 tend to become over-crowded.

3. We developed a detailed a game theoretical framework to address the downlink power setting problem in dense CA-enabled networks. The goal is to frame the problem in such a way that would enable us to reach distributed and efficient solutions for such a complex problem. To summarize, we frame the problem as a competitive game between teams of PoAs who have to compete for the same set (or overlapping sets) of radio resources. As each team's goal is to maximize its payoff function, we model such function in a way that captures both the benefits gained by each team when selecting a certain strategy (using a utility function) as well as the harm caused to other teams (using a cost function) by penalizing high power strategies. The cost functions also contains a component which ensures fairness for users experiencing bad channel conditions. We provide proof that the game we model belongs to the class of *pseudo-potential* games, which are know to possess pure Nash equillibria that can be reached via best-reply dynamics. Using this important result, we develop a distributed algorithm (*Best reply Power Setting – BPS)* to be performed at the teams level in order to reach a Nash equilibria (NE). Furthermore, we prove that at least in the single carrier case our algorithm leads to the preferred NE in terms of social welfare. A reduced-complexity algorithm is provided for the multi-carrier case, which, our numerical results performs almost as well as the optimal solution. We also provide a price-setting algorithm which addresses an important practical aspect when determining the value of the cost function.
The results of this part of the research were presented in two of our papers:

i. Z. Limani, C.-F. Chiasserini, G. M. Dell'Aera "Downlink Transmit Power Setting in LTE HetNets with Carrier Aggregation", IEEE WoWMoM, June 2016.
ii. Z. Limani, C.-F. Chiasserini, G. M. Dell'Aera "Downlink Power Setting in Carrier Aggreagation Enabled Dense Networks", *submitted to IEEE Transactions on Wireless Networks.*

We use the same two-tier network scenario that is used within 3GPP for evaluating LTE networks, composed of 57 macrocells and 228 microcells , to evaluate the performance of our approach Micro PoAs are deployed randomly over the coverage area so that there are 4 non-overlapping microcells per macrocell. The inter-site distance is set to 500 m. The PoAs are grouped into 57 five-player teams, each consisting of 1 macro PoA and all the micro PoAs within its macrocell, unless stated otherwise. Specifically, to make the scenario more realistic and comparable to an actual urban scenario, we divide the network coverage area into five types of urban areas: city centre, residential area, commercial area, parks and school area. The users are randomly dropped with varying density depending on the population density of the area type as well as time of the day (morning, afternoon or evening). Reference values for user density were obtained using official population statistics of the city of Rome (Italy) and then scaled to represent realistic values for cellular users of a single network provider. The user densities were further scaled for the different urban areas and times of the day, using weights extracted from the data provided in the MIT Senseable City Lab project. The mobility of pedestrian users was modelled using the random walk model, while the mobility of vehicular users was modelled using real mobility traces collected from taxi cabs in Rome, assuming an average velocity of 30 km/h. All user are assumed to be CA enabled. PoAs can use three CCs, each 10 MHz wide, with the following central frequencies: 2.6 GHz (CC1), 1.8 GHz (CC2) and 800 MHz (CC3). The maximum transmit powers for macro and micro PoAs are set at 20 W and 1 W, respectively. The performance of our approach is compared to the fixed strategy in which all PoAs transmit at highest power coupled with the eICIC technique, denoted as *eICIC* in the figures. Below are some of the main results obtained through numerical evaluation.

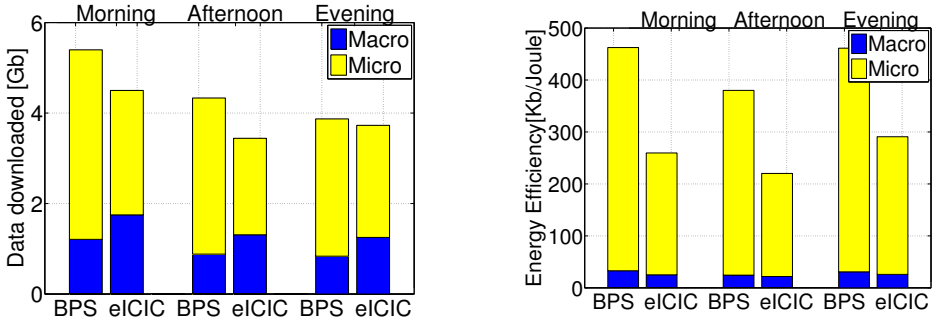


a) CC1 b) CC2 c) CC3

Fig. 7. BPS achieved power strategy.

In Fig. 7, we depict the averaged strategies reached through the BPS algorithm during the simulation period for the morning scenario. The strategies chosen by the teams for each CC are differentiated using different shades, from white (zero power) to black (maximum power). The maximum power varies depending on the type of PoA. Hexagons represent the macro PoAs, while circles represent the micro PoAs. The figure shows that CC1, i.e., the high frequency carrier, allows for higher transmit power levels to be used by both macro and micro PoAs, due to its low interference impact. CC2 and especially CC3 are used to complement each other to ensure overall coverage. In general we see that low transmit power levels are preferred for macro PoA across all CCs, while for micro PoAs the chosen transmit power levels tend to be higher for higher frequency CCs such as CC1. It can also be noted that in highly concentrated areas such as the city centre and commercial areas, the micro PoAs tend to transmit at higher power levels, while macro PoAs at lower power levels. Such a strategy enables micro PoAs in these areas,

who support most of the traffic demand, to transmit using a higher modulation coding scheme (MCS), which in turn implies higher bit rate and therefore throughput. In residential areas instead, traffic demand is lower and more spatially spread; thus, it is the macro PoAs that serve most of the traffic demand and therefore need to use higher power.



a) b)

Fig. 8. a) Total amount of content downloaded over the network; b) Energy efficiency achieved by the PoAs at different times of the day.

Fig. 8 (a) shows that when BPS is employed the amount of data downloaded over the network is always higher, especially during high intensity periods like morning and afternoon. While the difference is smaller in the evening when the traffic load decreases significantly, BPS still retains a considerable edge in energy efficiency (see Fig. 8 (b)). This is because BPS is able to serve higher amounts of data, while consuming significantly less power. From Fig. 8 (b) it is clear that BPS improves the energy efficiency for both macro PoAs and micro PoAs, however the effect is more significant for the latter: the gain in energy efficiency for macro PoAs varies between 15 and 20%, while for micro PoAs it can be as high as 100% during the morning and it drops to around 60% in the evening. BPS tends to choose lower transmit powers for macro PoAs, especially for dense areas with heavy traffic load, which significantly reduces the interference experienced by micro PoAs who are responsible for serving the bulk of the data.

B.4 **collaborazioni di ricerca avute in TUTTO IL TRIENNIO con Università , Centri di ricerca ed Industrie nazionali ed internazionali /Research collaborations during the whole period with national and international Universities, research centers and industries** (specificare la durata, il quadro entro cui sono avvenute: contratto di ricerca, periodi di formazione, collaborazione non formalizzata, e la eventuale durata della attivita' **presso** l'ente etc./ specify the duration, the reasons of these activities: research contract, research stage, informal collaboration,... and the the duration of your activity in the research center...)

Periodo Durata collaborazione e/o permanenza Ente (nominativo , luogo, nazionalita')

Period: March 2013-Jan 2017 Duration of the collaboration activity: 4 yrs Name of the research center and location: Telecom Italia, Turin, Italy

As a recipient of a Telecom Italia scholarship, I have collaborated with engineers from Telecom Italia throughout my PhD activity. I was assigned a tutor by Telecom Italia, Mr. Marco Caretti, with whom we have worked closely. We have also closely collaborated with other Telecom Italia engineers, namely Mr. Gian Michele Dell'Aera (also co-author of three of our works) and Mr. Bruno Melis. The collaboration was mostly performed through periodic meetings and sharing and exchanging of ideas about the direction of the research activity but also about technical and practical details relevant to our research.

B.5 **attività di ricerca di tutto il triennio** in progetti e contratti di ricerca nazionali ed internazionali **finanziati al dipartimento o a POLITO**./ **research activity** developed in the framework of projects or research contract between the Department DET or the POLITO and other institutions
Indicare/ Specify:

Durata complessiva Vs attivita', acronimo o identificativo del progetto
Full duration of your activity, project acronym or title

B.6 **brevetti** conseguenti l'attività di ricerca? Patent applications from the research activity

1. **Scheduling method and system for fourth generation radio mobile networks,** PCT International Application No. PCT/EP2014/069982

Authors: Zana Limani (PdT), Carla-Fabiana Chiasserini (PdT), Marco Caretti (Telecom Italia), Gian Michele Dell'Aera (Telecom Italia), Bruno Melis (Telecom Italia)

B.7 **altre attività** che si ritengono degne di menzione/ Other relevant activities

1. I was a recipient of the Politecnico Premio di Qualità 2014, for the 28th cycle
2. Our paper "Downlink Transmit Power Setting in LTE HetNets with Carrier Aggregation", presented at the IEEE WoWMoM 2016, received the Best Paper Award.
3. I was a recipient of a travel grant from the Ministry of Education, Science and Technology of Republic Kosovo, to attend IEEE WoWMoM 2016.

C. Attività di formazione / Learning activity

C.1 **partecipazione ad attività interne di supporto alla didattica/support activity to department teaching** (specificare il titolo corsi indicando se di diploma o laurea o altro, il tipo di attivita' (esercitaz, laboratorio, seminari e/o esami) indicandone il n.ro ore approssimativo, il docente responsabile corso e se l'attivita' e' stata remunerata con contratto o altra forma. / (specify the title of the course, nature of the course (I level, II level, other) and the nature of your activity(exercises, laboratory, seminar..) the teacher responsible of the course ..)
None

C.2 **corsi e seminari più significativi seguiti/list courses and seminars you attended** (interni, esterni, etc. – indicare il titolo, n.ro ore, docente, tipo e l'ente organizzatore)/ (internal or external, title, number of hour duration, name of the teacher and the sponsoring agency)

1. Internal, Categorical Data Analysis, 25 hrs, Kaisheng Song (University of North Texas)
2. Internal, Come scrivere un articolo scientifico e come fare una presentazione, 20 hrs
3. Internal, Data mining: concetti e algoritmi, 20 hrs, Elena Maria Baralis (Politecnico di Torino)
4. Internal, Experimental modeling: costruzione di modelli da dati sperimentali, 35 hrs, Michele Taragna, Carlo Novara (Politecnico di Torino)
5. Internal, Metodi di ottimizzazione per problemi ingegneristici, 30 hrs, Maurizio Repetto (Politecnico di Torino)
6. Internal, Surrogate modeling: theory for the user, 20 hrs, Igor Simone Stievano (Politecnico di Torino)
7. Internal, Electrical load management, forecasting and control, 25 hrs, Gianfranco Chicco and Federico Piglione (Politecnico di Torino)
8. External, Electromagnetic interference in communication systems, 60 hrs, Luan Ahma, (University of Prishtina, Kosovo)
9. External, Big Data Algorithms and Applications to Traffic Measurements, 24 hrs, Pietro Michiardi, Antonio Barbuzzi (Eurecom, France), Joint mPlane and BigFoot summer school

C.3 **periodi di formazione esterni** al Politecnico di tutto il triennio **/ research periods spent abroad** (tipo di formazione, istituzione, luogo e durata) / (kind of activity, institution name/ place and duration).

Research and seminars, University of Prishtina, Prishtina, Kosovo, 12 months

D. Indicazione delle pubblicazioni e brevetti nel triennio / number of publications produced during the PhD period as a result of your activity (pubblicate e in corso: indicarne il numero e il tipo: riviste internazionali, congressi internazionali, riviste nazionali, congressi nazionali, capitoli di libri, brevetti, altro; riportare l'elenco completo seguendo la suddivisione sopra elencata in un file allegato.)/Published or submitted to international journals and conferences, national journals and conferences/ books chapters, etc. Indicate only the number

1 article published in an international journal
1 article submitted to an international journal
4 articles published in international conferences

Data 14.11.2016

Firma del dottorando
PhD student signature

VISTO: si approva

IL o I TUTORE/I

.....

========== A L L E G A T O ==========

ELENCO PUBBLICAZIONI SCIENTIFICHE IN ORDINE DI IMPORTANZA
SCIENTIFIC PUBLICATIONS LIST IN ORDER OF IMPORTANCE

DEL DOTTORANDO Zana Limani Fazliu

del XXVIII Ciclo negli anni 2013-2016

Indicare nell'ordine e separandole le pubblicazioni su: riviste internazionali, atti di congressi internazionali, riviste italiane, atti di congressi italiani/ Separate the various kind of publications in order of importance: international journals and conferences, National journals and conferences

Si usi lo standard tipo pubblicazioni IEEE con nome di TUTTI gli autori, titolo etc./ the IEEE standard should be used
Le pubblicazioni sono solo quelle fatte dopo l'inizio del dottorato./ only the publication done on the PhD research activity
Le pubblicazioni gia' accettate possono essere inserite nell'elenco/ you can insert accepted publications an published publications.
Le pubblicazioni in corso di revisione si devono mettere al fondo separandole dalle precedenti ed indicandole sotto la voce "Pubblicazioni in fase di revisione"/ submitted publication must be included at the end of the list indicatin them as "submitted"

List of publications

1. F. Malandrino, **Z. Limani,** C. Casetti, C.-F. Chiasserini, "Interference-Aware Downlink and Uplink Resource Allocation in HetNets with D2D Support," IEEE Transactions on Wireless Communications, vol. 14 n. 5, pp. 2729-2741, 2015.
2. **Z. Limani Fazliu**, C.-F. Chiasserini, G. M. Dell'Aera "Downlink Transmit Power Setting in LTE HetNets with Carrier Aggregation", IEEE WoWMoM, June 2016 (best paper award).
3. **Z. Limani**, C.-F. Chiasserini, G. M. Dell'Aera "Interference-Aware Resource Scheduling in LTE HetNets with Carrier Aggregation Support", IEEE ICC June 2015.
4. F. Malandrino, C. Casetti, C.-F. Chiasserini, **Z. Limani** "Fast Resource Scheduling in HetNets with D2D Support," IEEE INFOCOM, April 2014.
5. F. Malandrino, C. Casetti, C.-F. Chiasserini, **Z. Limani** "Uplink and downlink resource allocation in D2D-enabled heterogeneous networks," IEEE WCNC, April 2014.
6. **Z. Limani Fazliu**, C.-F. Chiasserini, G. M. Dell'Aera "Downlink Power Setting in Carrier Aggreagation Enabled Dense Networks", *submitted to IEEE Transactions on Wireless Networks.*

CREDITS TABLE – SUMMARY

- Credits from exams		=>	50
- Papers accepted or published in international journals	num 1 x 20	=>	20
- Papers accepted or presented in international conferences	num 4 x 10	=>	40
- Papers accepted or published in national journals	num ___ x 10	=>	0
- Papers accepted or published in national conferences	num ___ x 5	=>	0

Patents accepted num.___ submitted num. 1

- Papers submitted at international journals	num 1
- Papers submitted at international conferences	num ___
- Papers submitted at national journals	num ___
- Papers submitted at national conferences	num ___

Bibliography

[1] Cisco White Paper, "Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2015 - 2020", February 2016.

[2] J. Andrews, "Seven ways that HetNets are a cellular paradigm shift," IEEE Communication Magazine, 2 March 2013.

[3] J.G.Andrews, S.Buzzi, W.Choi, S.V.Hanly, A.Lozano, A.C.K.Soong, J.C.Zhang, "What will 5G be?", IEEE Journal on Selected Areas in Communications, 2014.

[4] T. S. Rappaport, G. R. MacCartney, M. K. Samimi and S. Sun, "Wideband Millimeter-Wave Propagation Measurements and Channel Models for Future Wireless Communication System Design," in IEEE Transactions on Communications, vol. 63, no. 9, pp. 3029-3056, Sept. 2015.

[5] K.I. Pedersen, F. Frederiksen, C. Rosa, H. Nguyen, L. G. U. Garcia, and Y. Wang, "Carrier Aggregation for LTE-Advanced: Functionality and Performance Aspects", IEEE Communications Magazine, June 2011.

[6] F. Malandrino, Z. Limani, C. Casetti, C.-F. Chiasserini, "Interference-Aware Downlink and Uplink Resource Allocation in HetNets with D2D Support," IEEE Transactions on Wireless Communications, vol. 14 n. 5, pp. 2729-2741, 2015.

[7] Z. Limani Fazliu, C.-F. Chiasserini, G. M. Dell'Aera, "Downlink Transmit Power Setting in LTE HetNets with Carrier Aggregation", IEEE WoWMoM, June 2016 (best paper award).

[8] Z. Limani, C.-F. Chiasserini, G. M. Dell'Aera,"Interference-Aware Resource Scheduling in LTE HetNets with Carrier Aggregation Support", IEEE ICC June 2015.

[9] F. Malandrino, C. Casetti, C.-F. Chiasserini, Z. Limani, "Fast Resource Scheduling in HetNets with D2D Support," IEEE INFOCOM, April 2014.

[10] F. Malandrino, C. Casetti, C.-F. Chiasserini, Z. Limani, "Uplink and downlink resource allocation in D2D-enabled heterogeneous networks," IEEE WCNC, April 2014.

[11] Z. Limani Fazliu, C.-F. Chiasserini, G. M. Dell'Aera, E. Hamiti, "Downlink Power Setting in Carrier Aggregation Enabled Dense Networks", submitted to IEEE Transactions on Wireless Networks.

[12] 5G-Infrastructure Public-Private Partnership 2013. [Online].

Available:http://5g-ppp.eu/
[13] METIS-II: Mobile and wireless communications Enablers for Twenty-twenty (2020) Information Society-II, https://metis-ii.5g-ppp.eu/.
[14] M. Agiwal, A. Roy, and N. Saxena, "Next Generation 5G Wireless Networks: A Comprehensive Survey", IEEE Communications Surveys & Tutorials, Vol. 18, No. 3, Third Quarter 2016.
[15] GSMA Intelligence, "Understanding 5G:Perspectives on future technological advancements in mobile," White paper, 2014.
[16] S. Chen and J. Zhao, "The requirements, challenges, and technologies for 5G of terrestrial mobile telecommunication," IEEE Communications Magazine, vol. 52, no. 5, pp. 36-43, May 2014.
[17] G. P. Fettweis, "The tactile Internet: Applications and challenges," IEEE Vehicular Technology Magazine, vol. 9, no. 1, pp. 64-70, Mar. 2014.
[18] S. Landstrom, A. Furuskar, K. Johansson, L. Falconetti and F. Kronestedt, "Heterogeneous networks - increasing cellular capacity", Ericsson Review, 1/2011.
[19] S. Dimatteo, P. Hui, B. Han; V. O. K. Li, "Cellular Traffic Offloading through WiFi Networks", IEEE International Conference on Mobile Ad-Hoc and Sensor Systems, 2011.
[20] X. Lin, J. G. Andrews, A. Ghosh, "A comprehensive framework for device-to-device communications in cellular networks," http://arxiv.org/ abs/1305.4219.
[21] S. Sesia, I. Toufik, M. Baker (Eds.), LTE - The UMTS long term evolution: From theory to practice, Wiley, 2009.
[22] S. Deb, P. Monogioudis, J. Miernik, J.P. Seymour, "Algorithms for enhanced inter-cell interference coordination (eICIC) in LTE HetNets," IEEE/ACM Transactions on Networking, 2014.
[23] 4G Americas, "Understanding 3GPP Release 12: Standards for HSPA+ and LTE Enhancements", February, 2015.
[24] 3GPP, "TR 36.808 Evolved Universal Terrestrial Radio Access (E-UTRA); Carrier Aggregation; Base Station (BS) radio transmission and reception", Release 10, 2012.
[25] Qualcomm Technologies, Inc. Presentation, "Delivering on the LTE Advanced promise: Faster, better mobile broadband experience", March, 2016 [Online]. Available: https://www.qualcomm.com/documents/lte-advanced-evolving-and-expanding-new-frontiers
[26] M. Iwamura, K. Etemad, M.H. Fong, R. Nory and R. Love, "Carrier Aggregation Framework in 3GPP LTE-Advanced", IEEE Communications Magazine, August 2010.
[27] X. Lin, J. G. Andrews and A. Ghosh, "Modelling, Analysis and Design for Carrier Aggregation in Heterogeneous Cellular Networks", IEEE Transactions on Communications, September 2013.

[28] C. Jha, K. Sivanesan, R. Vannithamby, and A. T. Koc, "Dual connectivity in LTE small cell networks," in Proc. IEEE Globecom Workshops (GC Wkshps), Austin, TX, USA, Dec. 2014, pp. 1205-1210.

[29] M. Kamel, W. Hamouda, and A. Youssef, "Multiple association in ultra-dense networks," in Proc. IEEE Int. Conf. Commun. (ICC), Kuala Lumpur, Malaysia, May 2016.

[30] Y. Wang, K. Pedersen, P. Mogensen, and T. Srensen, "Resource allocation considerations for multi-carrier LTE-advanced systems operating in backward compatible mode," in 2009 IEEE 20th Int. Symp. Personal, Indoor and Mobile Radio Commun. , pp. 370 -374, Sept. 2009.

[31] Qualcomm White Paper, "LTE Advanced: Heterogeneous Networks," 2011.

[32] H. Elsawy, E. Hossain, D. I. Kim, "HetNets with cognitive small cells: user offloading and distributed channel access techniques," *IEEE Comm. Mag.*, vol. 51, no. 6, 2013.

[33] M. Peng, C. Wang, J.Li, H. Xiang, and V. Lau, "Recent Advances in Underlay HeterogeneousNetworks: Interference Control, Resource Allocation, and Self-Organization", IEEE Communication Surveys and Tutorials, Vol. 17, No. 2, 2015.

[34] G. Boudreau, J. Panicker, N. Guo, R. Chang, N. Wang, S. Vrzic, "Interference coordination and cancellation for 4G networks," *IEEE Commun. Mag.*, vol. 47, no. 4, 2009.

[35] 3GPP Std. Rel. 10, "Enhanced Inter-Cell Interference Control (ICIC) for non-Carrier Aggregation (CA) based deployments of heterogeneous networks for LTE," RP-100383, June 2013.

[36] G. Fodor, E. Dahlman, G. Mildh, "Design aspects of network assisted device-to-device communications," *IEEE Comm. Mag.*, vol. 50, no. 3, 2012.

[37] L. Xingqin, J. Andrews, A. Ghosh, R. Ratasuk, "An overview of 3GPP device-to-device proximity services," *IEEE Comm. Mag.*, vol. 52, no. 4, 2014.

[38] L. Lei, Z. Zhong, C. Lin, X. Shen, "Operator controlled device-to-device communications in LTE-advanced networks," *IEEE Wireless Comm.*, vol. 19, no. 3, 2012.

[39] M. Zulhasnine, C. Huang, A. Srinivasan, "Efficient resource allocation for device-to-device communication underlaying LTE network," *IEEE WiMob*, 2010.

[40] C.-H. Yu, K. Doppler, C. B. Ribeiro, O. Tirkkonen, "Resource sharing optimization for device-to-device communication underlaying cellular networks," *IEEE Trans. on Wireless Comm.*, vol. 10, no. 8, 2011.

[41] F. Malandrino, C. Casetti, C.-F. Chiasserini, "A fix-and-relax model for heterogeneous LTE-based networks," *IEEE Mascots*, 2013.

[42] L. Lei, X. Shen, M. Dohler, C. Lin, Z. Zhong, "Queuing models with applications to mode selection in device-to-device communications underlaying cellular networks," *IEEE Trans. on Wireless Communications,* vol. 12, no. 12, 2013.

[43] L. Song, D. Niyato, Z. Han, E. Hossain, "Game-theoretic resource allocation methods for device-to-device (D2D) communication," *IEEE Wireless Comm. Mag.*, 2014.

[44] X. Lin, J. G. Andrews, A. Ghosh, "A comprehensive framework for device-to-device communications in cellular networks," `http://arxiv.org/abs/1305.4219`.

[45] L. G. U. Garcia, K.I. Pedersen and P. E. Mogensen, "Autonomous Component Carrier Selection: Interference Management in Local Area Environments for LTE-Advanced", IEEE Communications Magazine, September 2009.

[46] H. Wang, C. Rosa, and K. Pedersen, "Analysis of Optimal Carrier Usage for LTE-A Heterogeneous Networks with Multicell Cooperation", IEEE GLOBECOM, 2013.

[47] Y. Wang, K. I. Pedersen, T. B. Sorensen and P. E. Mogensen, "Carrier Load Balancing and Packet Scheduling for Multi-Carrier Systems", IEEE Transactions on Wireless Communications, May 2010.

[48] K. Sundaresan and S. Rangarajan, "Energy Efficient Carrier Aggregation Algorithms for Next Generation Cellular Networks", IEEE ICNP 2013.

[49] H-S. Liao, P-Y. Chen and W-T. Chen, "An Efficient Downlink Radio Resource Allocation with Carrier Aggregation in LTE-Advanced Networks", IEEE Transactions on Mobile Computing, October 2014.

[50] T. Heikkinen, "A Potential Game Approach to Distributed Power Control and Scheduling," *Computer Networks,* 2006.

[51] M. Xiao, N.B. Shroff, E.K.P. Chong, "A Utility-based Power-control Scheme in Wireless Cellular Systems," *IEEE/ACM Transactions on Networking,* 2003.

[52] D. Monderer, L. S. Shapley, "Potential Games," *Games and Economics Behaviour,* vol. 14, no. 1, pp. 124–143, 1996.

[53] P. Dubey, O. Haimanko, A. Zapelchelnyuk, "Strategic Complements, Substitutes and Potential Games," *Games & Economic Behavior,* 2004.

[54] L. Duan, J. Huang, B. Shou, "Economics of Femtocell Service Provision," *IEEE Trans. on Mob. Comp.,*, vol. 12, no. 11, 2013.

[55] F. Pantisano, M. Bennis, W. Saad, M. Debbah, M. Latvaaho, "Improving Macrocell-Small Cell Coexistence Through Adaptive Interference Draining," *IEEE Trans. on Wireless Comm.*, vol. 13, no. 2, 2014.

[56] Z. Zhang, L. Song, Z. Han, W. Saad, "Coalitional Games with Overlapping Coalitions for Interference Management in Small Cell Networks," *IEEE Transactions on Wireless Communications*, 2014.

[57] S. Guruacharya, D. Niyato, E. Hossain, D.I. Kim, "Hierarchical Competition in Femtocell-Based Cellular Networks," *IEEE Globecom,* 2010.

[58] L. Diez, G.-P. Popescu, R. Aguero, "A Geometric Programming Solution for the Mutual Interference Model in HetNets," *IEEE Communications Letters,* 2016.

[59] Z. Ding, S. M. Perlaza, I. Esnaola, H. V. Poor, "Power Allocation Strategies in Energy Harvesting Wireless Cooperative Networks," *IEEE Transactions on Wireless Communications,* vol. 13, no. 2, pp. 846–860, 2014.

[60] M. Lin, N. Bartolini, T. La Porta, "Power Adjustment and Scheduling in OFDMA Femtocell Networks," *IEEE Infocom,* 2016.

[61] N. Sapountzis, T. Spyropoulos, N. Nikaein, U. Salim, "Optimal Downlink and Uplink User Association in Backhaul-limited HetNets," *IEEE Infocom,* 2016.

[62] E. Yaacoub, A. Imran, Z. Dawy, A. Abu-Dayya, "A Game Theoretic Framework for Energy Efficient Deployment and Operation of Heterogeneous LTE Networks," *CAMAD,* 2013.

[63] K. Yang, S. Martin, T.A. Yahiya, J. Wu, "Energy-efficient Resource Allocation for Downlink in LTE Heterogeneous Networks," *IEEE VTC,* 2014.

[64] S. Samarakoon, M. Bennis, W. Saad, M. Debbah, M. Latva-aho, "Ultra Dense Small Cell Networks: Turning Density Into Energy Efficiency," *IEEE Journal on Selected Areas in Communications* , 2016.

[65] A. Y. Al-Zahrani, F. R. Yu, M. Huang, "A Joint Cross-Layer and Colayer Interference Management Scheme in Hyperdense Heterogeneous Networks Using Mean-Field Game Theory," *IEEE Transactions on Vehicular Technology,* 2016.

[66] H.-H. Nguyen, W.-J. Hwang, "Distributed Scheduling and Discrete Power Control for Energy Efficiency in Multi Cell Networks," *IEEE Communication Letters,* 2015.

[67] G. Yu, Q. Chen, R. Yin, H. Zhang, G.Y. Li, "Joint Downlink and Uplink Resource Allocation for Energy-efficient Carrier Aggregation," *IEEE Transactions on Wireless Communications,* 2015.

[68] Qualcomm, "Creating a digital sixth sense with LTE Direct", July 2014.

[69] ITU-R, "Guidelines for evaluation of radio interface technologies for IMT-Advanced", *Report ITU-R M.2135-1,* Dec. 2009.

[70] "C-RAN - The road towards green RAN", China Mobile Research Institute White Paper, 2011.

[71] Freescale white paper, "Next-generation wireless network bandwidth and capacity enabled by heterogeneous and distributed networks," July 2013.

[72] D. Bladsjo, M. Hogan, S. Ruffini, "Synchronization Aspects in LTE Small Cells," *IEEE Communication Magazine,* 2013.

[73] 3GPP RP-122009, "Study on LTE device to device proximity services," in 3GPP TSG RAN Meeting #58, Dec. 2012.

[74] 3GPP Technical Specification 36.213, "LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures," Release 11, 2013.

[75] J. Meinilä, P. Kyösti, L. Hentilä, T. Jämsä, E. Suikkanen, E. Kunnari, M. Narandžić, "D5.3: WINNER+ final channel models," *Wireless World Initiative New Radio WINNER+*, 2010.

[76] D. Martín-Sacristán, J. F. Monserrat, J. Cabrejas-Peñuelas, D. Calabuig, S. Garrigas, N. Cardona, "3GPP long term evolution: Paving the way towards next 4G," *Waves*, 2009.

[77] W. B. Powell, *Approximate dynamic programming*, Wiley, 2011.

[78] 3GPP Technical Report 36.814, "Further advancements for E-UTRA physical layer aspects," 2010.

[79] D. J. Watts, *Small worlds: The dynamics of networks between order and randomness,* Princeton University Press, 1999.

[80] FP7 IP EARTH project, "Deliverable D2.3: Energy efficiency analysisof the reference systems, areas of improvements and target breakdown," https://www.ict-earth.eu/.

[81] A. Ghosh, N. Mangalvedhe, R. Ratasuk, B. Mondal, M. Cudak, E. Visotsky, T. A. Thomas, "Heterogeneous cellular networks: From theory to practice," *IEEE Comm. Mag.*, vol. 50, no. 6, 2012.

[82] 3GPP Technical Report 36.942 V12.0.0, 2014.

[83] B. Soret, H. Wang, K.I. Pedersen, C. Rosa, "Multicell Cooperation for LTE-Advanced Heterogeneous Network Scenarios," *IEEE Wireless Communications,* 2013.

[84] http://www.laboratorioroma.it/Banca dati/08_zone urbanistiche.htm

[85] http://senseable.mit.edu/manycities/.

[86] L. Bracciale, M. Bonola, P. Loreti, G. Bianchi, R. Amici, A. Rabuffi, CRAWDAD dataset roma/taxi, traceset: taxicabs, http://crawdad.org/roma/taxi/20140717/taxicabs, doi:10.15783/C7QC7M, 2014.

[87] Ericsson white paper, "It all comes back to backhaul," 2012.