Analysing uplink scheduling in mobile networks

A flow-level perspective

D. C. Dimitrova

Analysing uplink scheduling in mobile networks

A flow-level perspective

Desislava C. Dimitrova

Graduation committee:

Chairman:	prof. dr. ir. A. J. Mouthaan
Promoter:	prof. dr. J. L. van den Berg
Assistant promoter:	dr. ir. G. Heijenk

Members:

prof. dr. ir. B. R. H. M. Haverkort	University of Twente
prof. dr. ir. P. J. M. Havinga	University of Twente
prof. dr. ir. S. M. Heemstra de Groot	Delft University of Technology
prof. dr. R. D. van der Mei	Vrije Universiteit Amsterdam
prof. dr. T. Braun	Universität Bern
dr. MSc R. Litjens	TNO

CTIT Ph.D.-thesis Series No. 10-181
Centre for Telematics and Information Technology
University of Twente, P.O. Box 217, NL-7500 AE Enschede

ISSN 1381-3617
ISBN 978-90-365-3090-3

Publisher: Wöhrmann Print Service
Cover design: Desislava C. Dimitrova

ANALYSING UPLINK SCHEDULING IN MOBILE NETWORKS

A FLOW-LEVEL PERSPECTIVE

PROEFSCHRIFT

ter verkrijging van
de graad van doctor aan de Universiteit Twente,
op gezag van de rector magnificus,
prof. dr. H. Brinksma,
volgens besluit van het College voor Promoties,
in het openbaar te verdedigen
op woensdag 24 november 2010 om 15.00 uur

door

Desislava Cvetanova Dimitrova

geboren op 20 september 1981
te Plovdiv, Bulgarije

Dit proefschrift is goedgekeurd door:
prof. dr. J. L. van den Berg (promotor)
dr. ir. G. Heijenk (assistent-promotor)

Abstract

The main purpose of mobile networks is to enable customers that are located at arbitrary geographical locations to communicate with each other without the need of a physical connection. Within only two centuries mobile technology has evolved from analogue networks providing telephony services towards digital networks supporting larger variety of mobile services, enhanced coverage and higher data rates.

A key element of providing wireless connectivity is the management of the radio spectrum to be shared by the mobile users. The challenge in the spectrum management is to find a trade-off between efficiently using the network and at the same time providing the quality of service (QoS) requested by the users. In this task operators are strongly assisted by the radio resource management (RRM) mechanisms and scheduling in particular. Scheduling is responsible for the distribution of the available radio resource over the users that have requested service.

This thesis focuses on the performance of uplink scheduling schemes in mobile networks. We have dedicated special attention to the impact of flow-level dynamics, i.e. the random user behaviour regarding the initiation and completion of data flow transfers. Additionally, we investigate the possibilities to adopt relaying as a technique to boost performance and how it interacts with the scheduling mechanism.

In order to evaluate the impact of the flow-level dynamics on the scheduling performance we propose a novel hybrid analysis approach, which we consider to be an overall contribution of this thesis. The approach is a combination of analytical methods and simulation. Packet level details, playing on a small

time scale, are captured by analytical methods, while the flow-level behaviour, playing on a larger time scale, is simulated. This combination enables fast evaluation and comparison of the performance of different schedulers expressed in terms of flow throughputs and mean flow transfer times.

Our general conclusion is that flow dynamics can have significant impact on performance. For example, we show that some changes (benefits) in performance exhibit only during the interaction of flow transfers. Therefore, analysis of the system at flow level should be included in the analysis of mobile systems. The proposed hybrid analytical/simulation approach, which captures both packet- and flow-level behaviour, is very flexible and can be easily adapted to reflect changes in factors such as environmental conditions or technological specifications. These features make the approach very appropriate for application in other studies as well.

In our studies we have investigated the impact of various factors, e.g. inter-cell interference and individual user channel conditions, on the scheduling scheme. Based on our findings we conclude that a scheduling scheme should be carefully designed to consider, among others, requirements towards utilisation efficiency, user's QoS demands and the wireless environment. Furthermore, we also show that relaying leads to improved service provided, not only to users that directly make use of a relay station but also to users that communicate directly to the network.

Contents

1 Introduction

Ever since humans have learnt how to use the vocal cords to form meaningful sounds, communication has been a building pillar of society. Communication has integrated in every aspect of our personal and professional lives. It allows us to express our needs and feelings. It gives us the means to convey our desires and to animate our dreams. By communicating we are able to share our knowledge and experience with others in order to achieve a common goal.

The ability to communicate independently of the distance is one of the greatest achievements of humankind. Overcoming the geographical barriers enables the exchange of information among large and diverse groups of people leading to further and faster advances in many spheres of society among which economy and science. Communication networks, as the infrastructure supporting this information exchange, are equally important as the communication itself.

The first public communication network was the telegraph, which revolutionised the exchange of text messages during the late 1860s. Quickly afterwards, in the 1870s, the invention of the telephone enabled a whole new world of voice communication over distance. The next significant advance in land-based communication networks was in 1962 when the idea to create a computer network for the exchange of information between systems appeared. This first Internet-ancestor was called Advanced Research Projects Agency Network (ARPANET) and was developed by the United States Department of Defence during the 1960s. Almost thirty years later, during the 1990s, *Internet* reached the general public and began its triumphant march.

During the 1950s, mobile telephone networks, conveying information over the

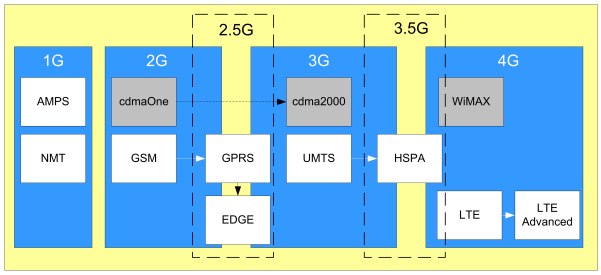


Figure 1.1: Evolution of mobile (cellular) networks - generations and corresponding technologies.

ether in the form of radio signals, started to appear. These networks provided telephone connectivity independent of location and therefore were expected to have high market potential. The first mobile phone networks were analogue, e.g. Advanced Mobile Phone System (AMPS), see [60], while later technologies evolved to work with the more robust digital signals. Digital cellular technologies such as Global System for Mobile communications (GSM), see [57, 60], brought along enhanced functionality, improved network capacity and more applications to enrich the customer's experience. The more attractive the new services, the more enthusiastic the customers become, the higher their demands towards mobile network operators. Today's customers want to be connected everywhere, at all time and to have Internet-like capabilities on their mobile phones. The cellular technologies currently in operation, such as Universal Mobile Telecommunications System (UMTS) and cdma2000, aim to meet customers' demands and yet seem to leave space for improvement.

To meet the increasing demands operators, manufacturers and researchers continuously try to upgrade current networks by finding new ways to overcome the existing technological challenges. One effort in this direction is the High Speed Packet Access (HSPA) technologies, which are an upgrade of the UMTS networks, see [1, 2]. Another effort is the Long Term Evolution (LTE) which is under ongoing development as the technology to extend the operation of the currently deployed UMTS/HSPA technologies. Figure 1.1 shows the evolution of mobile networks up to now and the corresponding technologies; the inheritance relation between technologies is indicated by an arrow. In white we have identified the most widespread technologies; LTE and LTE Advanced are expected to be the leading 4G technology.

In this thesis we address some of the challenges of mobile networking and in particular those related to network efficiency and performance optimisation. The capacity of a mobile network is determined mainly by the limited radio resource (frequency spectrum), which is why our research focuses on the radio access part of mobile networks. *Radio resource management* (RRM) mechanisms are defined to provide technical means for 'smart' (efficient) usage of the radio resource. Each mechanism, depending on its specific responsibility, offers distinct possibilities to optimise system operation. In this thesis we are especially interested in the so-called *scheduling mechanism*, which decides about how the radio resource is distributed over the users requiring service. More specifically, we will investigate the performance of scheduling strategies in various network scenarios, particularly taking into account the complicating effects of the randomness regarding the users' communication needs (time and location, duration). We concentrate on the communication direction from the users towards the mobile network, i.e. uplink.

The purpose of this first chapter is to introduce the research questions addressed in the thesis as well as to point out our main contributions. In order to understand the research, its relevance and novelty, the reader needs to be familiar with the basics of mobile networking and more specifically with certain technological aspects of their operation, which is done in Section 1.1. Particularly, in Sections 1.1.1 and 1.1.2 we identify several trends and challenges in the current development of cellular technologies; some of these challenges motivated our choice of research topics, which are discussed in Section 1.2. The research approach that we propose and use for analysis and evaluation is briefly described in Section 1.3. Section 1.4 presents the principal contributions of the thesis. Finally, in Section 1.5, the organisation of the thesis is outlined.

1.1 Mobile networks and services

In this thesis we focus on mobile cellular networks. For the sake of simplifying the textual expression, in the rest of the thesis we will speak about mobile networks leaving cellular out of the name.

A *mobile cellular network* is a radio network made up of a number of fixed-location transceivers (transmitter/receiver) known as a *cell site* or *base station* (BS). Each base station covers a particular geographical area referred to as radio cell (or just cell) and hence a network consists of multiple cells of finite size, see Figure 1.2. The finite cell size is due to radio transmission specifics, i.e. signal degradation in distance. The number of cells required to provide service over a

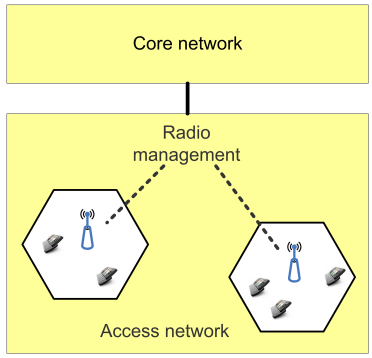


Figure 1.2: Mobile networks architecture - radio access part with cell structure and a core network connecting different radio access parts and providing access to e.g. the Internet.

particular geographic area depends on user density, landscape.

The base stations, together with some additional nodes also running radio resource management functionalities, are collectively referred to as *access network*, since they provide wireless access to the mobile users. Other administrative and network management functions are positioned in a *core network*, which also provides connectivity to external, e.g. land-based, networks. More details on a mobile network's architecture are provided in Chapter 2.

1.1.1 Trends

Mobile networks have been developed to provide uninterrupted wireless connectivity to (mobile) users at arbitrary geographic locations. Originally, the focus was on providing mobile telephony services. As the underlying cellular technologies changed from analogue, e.g. Nordic Mobile Telephone (NMT), to digital, e.g. GSM, also data based services became feasible. These services experienced unexpected popularity and became so favoured by customers that nowadays the majority of mobile traffic comes from non-voice Internet-like applications. As a result the development focus of current cellular technologies, e.g. UMTS/HSPA, as well as future ones, e.g. LTE, is on data and multi-media services. More details on cellular technologies are provided in Chapter 2.

The ever increasing *customer demand* for faster and more diverse services

and the subsequent eminent growth of data traffic fuels the engine of the mobile industry [1]. An interesting example in this context is the introduction of Apple's iPhone 3G in the USA in 2008 - the huge amount of traffic that the multitude of newly provided applications generated caused serious capacity depletion problems (mainly in big cities with large population of iPhone users) for the exclusive rights operator AT&T. This pushed AT&T to search new ways to provide the lacking capacity.

Another aspect of customer's behaviour, next to the request for higher rates that *network operators* need to deal with, is the demand for ubiquitous service availability. Mobile cellular networks have turned into a billions worth industry characterised by a complex structure, intricate regulations, a multitude of service operators and manufacturers and billions of customers.

1.1.2 Network efficiency and service quality

A major challenge ahead of network operators is how to fit, in the most efficient manner, the quickly increasing customer demand in their already limited radio resource (frequency spectrum). Resource allocation (division of the spectrum over the users) is further complicated by the necessity to deliver certain quality of service (QoS) for each user requesting service. These two requirements towards allocation strategies - efficient resource utilisation and QoS support - unfortunately seem to contradict with each other. Often the best utilisation of the available resource results in selective service of users, which contradicts with the idea of providing satisfactory QoS to all users. Hence, finding a trade-off between an acceptable QoS (benefits customers) and high efficiency of spectrum utilisation (benefits the operator) is a crucial task for a network operator.

In finding a good trade-off, as user traffic steadily increases, network operators are largely assisted by the already mentioned radio resource management mechanisms. They exercise direct or indirect control over the usage of the available network resources. Examples of RRM mechanisms are: admission control, which guarantees that only as many users are accepted for service as the network can actually support; power control that adapts the transmission powers to changing propagation/interference conditions in order to ensure successful signal reception while keeping a 'healthy' interference environment; and scheduling as the mechanism responsible for the distribution of the radio resource over

[1] An interesting contradiction is the case of UMTS, where operators and manufacturers were the driving force behind the network's evolution while customers were initially reluctant to switch toward the new technology.

the active users. Scheduling is a key RRM mechanism crucial for achieving high performance and improving utilisation. In this thesis we focus on scheduling.

The task of RRM in general and scheduling in particular is further complicated by several factors related to the user's behaviour and technological specifics of a mobile network and of the propagation channel. These factors and their effects should be carefully considered during the design and evaluation of RRM mechanisms in order to optimise the deployment of the mobile network. Below the main factors are shortly discussed.

User's behaviour/traffic load: Service requests from a large population of (potential) users occur at different locations and different time instants, and can be considered as random requests. Moreover, the duration of a service session (e.g. a phone call or a web browsing session), the traffic intensity during a session, etc. can vary strongly. This all makes that the users' demand for network resources is hard to predict and may fluctuate considerably over time.

Variation in channel quality: Another complicating factor is the (significant) variation in channel quality over the users, even if the same service is requested. This is a result of individual channel conditions, mainly affected by the user's location (i.e. path loss effect), user's mobility during flow transfer (i.e. shadowing effect) and changing environmental conditions in the radio channel (i.e. interference impact). Channel quality is important since it affects the decision how much resource to assign to a particular user in order to achieve the requested QoS.

Diversity of services: Future mobile networks should support a multitude of services (applications) characterised by distinct QoS requirements (e.g regarding throughput, transmission delay, etc.). In order to meet these distinct QoS requirements and achieve appropriate network efficiency, differentiation in resource allocation and traffic handling in the network is needed. Roughly spoken, services can be classified into elastic services, e.g. data services with relatively loose QoS requirements, and non-elastic services, e.g. voice or video with strict QoS requirements. For example, a voice call is very sensitive to delays while a data transfer can tolerate delays as long as the total information is exchanged.

1.2 Addressed research topics

As previously mentioned the focus of the work in this thesis is on uplink scheduling in mobile access networks; this topic is further introduced in Section 1.2.1. An additional topic considered in this thesis is relaying, i.e. the use of additional nodes (relay stations) in the access network to improve the communication

between users and base station. Relaying is briefly introduced in Section 1.2.2. We are interested in the actual performance improvements that can be achieved, hereby paying particular attention to the role of scheduling in a network setting with relaying.

1.2.1 Scheduling

Scheduling is the process of deciding how to distribute resources over a variety of possible tasks. In the case of mobile networks the resource is the available radio spectrum and the 'tasks' are the mobile users, who have data available to transmit or who want to download data. The scheduler is usually located at an access network entity, e.g. base station. It is the task of the scheduler to decide which users can be served, how much resource to assign to each of them and in what order to serve them. Scheduling decisions need to be taken on a frequent time bases (most often in the order of milliseconds) such that to quickly adapt to changes in the customers' demand or in channel quality. Therefore, from the scheduler's perspective, time is divided in slots and at the beginning of each slot scheduling decisions are re-evaluated.

Schedulers can differ significantly depending on how many system characteristics and information of the mobile users are used in the decision taking process. Comparison of a variety of schemes can be valuable for operators to decide which type of scheduler best suits their objectives regarding network performance and fairness. When evaluating the scheduling schemes it is essential to realise that mobile users have random behaviour (see Section 1.1.2), which significantly affects the performance. Including user behaviour in the performance analysis of scheduling schemes is an important issue in this thesis.

Much has already been done in the area of scheduling in cellular networks. In particular, researchers have optimised the resource efficiency of the scheduling strategy given certain pre-defined user performance objectives (regarding e.g. throughputs and delays) that have to be met, e.g. [29]. Specific issues that play an important role here are among others the impact of the traffic characteristics, see e.g. [53], and the radio channel characteristics (propagation, interference, etc), see e.g. [41, 55]. These scheduling studies have been performed in the context of a variety of cellular technologies, in particular UMTS, UMTS/HSPA and, more recently, LTE (some examples are [42, 66, 77]). Most research is done for downlink traffic scenarios since traffic in the *downlink* direction (from base station to mobile) used to dominate communication. However, currently a significant increase in *uplink* traffic (from mobile to base station) is observed, which is partly caused by the large interest customers show towards file sharing

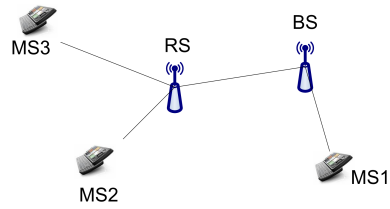


Figure 1.3: A relay-enabled cell.

and the popularity of social network applications. It is expected that this trend will continue.

Linking up with the increasing relevance of the uplink in mobile networks, and the relatively low attention paid to it in the existing literature, our focus in this thesis will be on uplink scheduling issues. Hereby, we are particularly interested in the impact of the random user behaviour (which we briefly addressed in Section 1.1.2) on the performance of different uplink scheduling schemes.

1.2.2 Relaying

Another issue we consider in this thesis is the application of *relaying* in mobile networks, in particular for the uplink. The idea of relaying is rather simple - position an intermediate device, a relay, which receives and forwards data between the user and the base station. Signal retransmission can be beneficial since signal strength weakens faster than linearly in distance. Hence, bridging the same distance by two transmission paths results in a lower path loss and increased signal quality. In fact, this is what relaying exploits. Figure 1.3 shows an example where users, depending on their position transmit via a relay or directly to the base station.

Relay deployment can be used to improve cell edge service, extend coverage beyond the cell edge or provide coverage behind obstacles. Anticipated advantages of using relaying, compared to extending the number of base stations in the network, are the low costs and easy deployment of relays - both resulting from the reduced technical complexity (and hence financial cost) of a relay station compared to a base station while using (in total) the same or even less transmission power.

The general concept of relaying has attracted much attention, e.g. [8, 25, 50]. The general conclusion is that indeed relaying leads to increased spectrum utilisation and higher data rates, from which both operators and customers

benefit. However, most studies concentrate on the sole contribution of the relay station(s) and neglect the role of the scheduling scheme, e.g. [64, 69]. Still the dynamic relation between scheduler and relay determines the eventual system performance and shapes opportunities to increase spectrum utilisation. It is our intention to investigate the interaction of the two mechanisms and to provide understanding how this knowledge can be used for the benefit of operators.

1.3 Research approach

The research approach adopted in this thesis has two particular aspects that we want to discuss in some more detail in the subsections below - one is the granularity level at which the considered systems are studied, and another one is the undertaken analysis approach.

1.3.1 Packet- and flow-level modelling

Roughly spoken we can differentiate between two granularity levels when observing traffic in mobile networks (and, in fact, in any modern communication network). On the finer level, termed packet level, we observe the individual data packets that are exchanged between a mobile user and its service base station. On the coarser level, referred to as flow level, packets become more or less transparent and the focus is on the traffic flows, comprising of a series of packets, generated by the network users. By analysing only packet-level behaviour one may overlook the impact of the flow-level dynamics, i.e. the continuous change in the number of ongoing flows as a result of the random initiation, in both time and space, of flow transfers, and their completion (see Section 1.1.2).

Since flow-level dynamics represent the change in the number of active users in the cell, they are relevant for the efficient utilisation of the common radio resource. Hence, in our opinion, the impact of flow dynamics on network's performance needs to be carefully considered. Surprisingly, this issue is not well studied - authors, e.g. [5, 6], mostly do not account for the random user behaviour, but assume a fixed constellation of active users in the network.

Therefore, in this thesis, we pay particular attention to incorporating the impact of the flow-level dynamics in the performance evaluation and comparison of different scheduling and relaying strategies. We show that by taking into account the random behaviour of the users in initiating and completing flow transfers, we are able to observe performance aspects of the studied mechanisms that remain hidden if only a packet-level analysis is performed. For example, as

we will later show, flow-level analysis for a relay enabled network showed that also users that do not use a relay benefit indirectly from the relaying.

1.3.2 Hybrid evaluation approach

Most studies adopt one out of two analysis approaches - either an analytical investigation or a simulation-based evaluation. Both types have their advantages and disadvantages. Analytical studies may provide insightful, explicit performance expressions and support fast evaluation. However, to achieve that one usually needs to significantly limit the level of system detail taken into account in the analysis. Simulation-based approaches generally incorporate many details of the studied mechanisms but at the cost of long simulation times.

In our opinion none of these two basic approaches meets our goal to model scheduling and relaying in a mobile network at both packet and flow level, while still allowing for fast evaluation. Therefore we propose a novel hybrid approach, combining simulation and analytical methods, which allows us to capture the most important system characteristics at both levels and considerably speeds up the performance evaluation compared to a 'simulation-only' approach.

1.4 Contributions

This thesis focuses on uplink scheduling in current and future mobile cellular networks, and in particular on how it can contribute to improving the service quality experienced by the users while trying to increase network efficiency. In addition, we also investigate the potential gains that relaying has to offer when combined with an appropriate scheduling scheme.

As an important overall contribution of the thesis we identify the undertaken hybrid performance analysis approach, combining simulation and analytical methods, which captures both the specifics of the studied network mechanisms (scheduler, relay) at packet level and the random behaviour of the mobile users regarding the generation of traffic flows to be handled by the network (i.e. flow-level dynamics). Taking into account the flow-level dynamics reveals effects that often remain hidden when the system is studied solely at the packet level. See also the discussion above in Section 1.3.

The research performed in this thesis is organised in three parts - one part on scheduling in UMTS/HSPA networks; one part on the cooperative operation of relaying and scheduling; and one part on scheduling in LTE networks. Below the main contributions of each part are discussed in some more detail.

1.4.1 Part I: Scheduling in UMTS/HSPA networks

This part is dedicated to the topic of resource management and scheduling for the uplink in a UMTS/HSPA mobile network. Several scheduling schemes are proposed and evaluated based on the newly developed hybrid analysis methodology, see above. The main contributions of the first part can be summarised as:

- Developing the hybrid analysis methodology briefly introduced in Section 1.3.2 that allows us to capture the effect of environmental, technological and user (random behaviour) factors on scheduler's performance. The methodology undertakes a modular approach towards the analysis of a scheduler, which make it very easy to adapt to the discussed network scenario (or technology).
- Analysing performance in a network with two coexisting types of traffic - 'plain' voice calls and data transmissions. Such a system is intriguing due to the fact that the two types of traffic are scheduled differently in order to ensure the tight QoS requirements (packet delays) of the voice traffic. A particularly interesting issue considered in the thesis is the impact of the 'high-priority' voice traffic on the performance of the data transmissions.
- Analysing the impact of the used scheduling scheme on the inter-cell interference generated by the users' transmissions, and how this in turn influences the performance. On the one hand, we show that depending on the scheduling scheme a cell affects the performance of its neighbours differently, i.e. inter-cell interference is scheduler specific. On the other hand, we also show that the performance of various schedulers is influenced by inter-cell interference differently.

1.4.2 Part II: Relaying

As explained in Section 1.2.2 the purpose of a relay is to bridge the distance between mobile user and base station in order to improve service quality, e.g. achieve higher data rates. As another participant in the radio interface relay transmissions have to be served by the scheduler as well, which asks for modifications of the scheduling schemes. The main contributions of Part II are summarised as:

- Identifying practically feasible relay-enabled scheduling schemes, and evaluating and comparing their performance at flow level.

- Determining how several relay deployment aspects, e.g. relay location and transmit power, affect user performance.
- Evaluating the effect of relaying on inter-cell interference and the consequences of that for the performance of mobile users.

1.4.3 Part III: Uplink scheduling in LTE networks

LTE adds an extra dimension to scheduling - it allows not only scheduling in the time domain, as in UMTS/HSPA, but also in the frequency domain as will be explained in Chapter 2. This provides more flexibility in resource allocation, but brings also a higher level of complexity. For investigating the flow-level performance of different LTE scheduling strategies the hybrid analysis approach developed in Part I for one-dimensional (time-)scheduling strategies (in the context of UMTS/HSPA) is extended to the situation with two degrees of freedom in LTE. As the main contributions of Part III we identify:

- Evaluating the performance of various LTE packet scheduling strategies. The results of our investigations here emphasise the particular importance of incorporating flow-level dynamics in the analysis for getting the right insights in the schedulers' performance.
- Analysing the consequences of several essential, technological system limitations of LTE networks (e.g. maximum number of users that can be simultaneously scheduled in a time slot) on the performance of the various schedulers.
- Studying the effects of LTE scheduling strategies on inter-cell interference, and how this influences the performance at flow level. Our findings provide useful insights, which are valuable, as we also show, for the design of interference mitigating scheduling schemes.

1.5 Outline

This thesis spans over nine chapters. Firstly, the current Chapter 1 presents an overall introduction to the research area and the relevance of the performed research. The next chapter, i.e. Chapter 2, provides a description of the main technical features of the investigated cellular technologies and mechanisms as far as needed for the understanding of the work presented in this thesis.

Subsequently, the chapters covering the three research parts that we described in the previous section are presented. In particular, Part I (Uplink scheduling in UMTS/HSPA networks) consists of Chapter 3, which is mainly dedicated to the introduction of the hybrid performance analysis approach and its application to basic UMTS/HSPA single cell scenarios, and Chapter 4, where the effects of inter-cell interference in network scenarios with multiple cells are studied. Part II (Relaying) is built up of Chapters 5 and 6. In Chapter 5 the focus is on the performance analysis and comparison of different relay-enabled scheduling schemes; Chapter 6 is mainly dedicated to investigating the impact of relay deployment aspects like e.g. the relay's location in the cell and its transmit power. Chapters 7 and 8 together form the last part of the thesis, i.e. Part III (Uplink scheduling in LTE networks), where Chapter 7 considers the basic single cell scenario and Chapter 8 focuses on the interaction between scheduling and inter-cell interference in multiple-cell scenarios. Finally, in Chapter 9, we summarise our work and give outlook towards future research perspectives.

The work presented in Chapter 3 is an extension of [13], which was presented at WWIC 2008. The research of Chapter 4 is based on [12] and [16], which appeared at WiMob 2008 and NGMAST 2009 respectively. Chapter 5 combines our studies from [17] and [24]; the former was included in the proceedings of WMNC 2009 while the latter in the proceedings of European Wireless 2010. Chapter 6 is based on work initially presented at the 3rd ERCIM Workshop on eMobility, see [14], and extended in [15], appearing at ICUMT 2009. Finally, Chapter 7 is an extension of [18]. This was presented at the 4th ERCIM Workshop on eMobility. Currently, we prepare the work of the last two chapters 7 and 8 for submission.

While writing we have tried to structure the thesis in such a way that each chapter can be read on its own without relaying too much on other chapters. Therefore, there is some overlap among the chapters.

2 Mobile cellular networks

2.1 Introduction

In Chapter 1 we briefly introduced the technologies that emerged in the evolution of mobile cellular networks. The current chapter further elaborates on these in order to establish familiarity with mobile cellular networks and their major functional elements. The focus is on those mobile networking technologies and those functionalities (mechanisms) that are important for understanding the work in this thesis. Furthermore, common notation, used throughout the thesis, is introduced.

First, in Section 2.2 we introduce the particular evolution stages of mobile cellular communications. This will help the reader to position the research topics of this thesis - scheduling and relaying - in the context of cellular network development. Subsequently, in Section 2.3 we give a general introduction to some basic mobile networking aspects related to our research, including network architecture, radio resource management mechanisms and signal propagation theory. Sections 2.4 and 2.5 explain these aspects of the particular mobile technologies of our interest in more detail. The focus of these sections is on scheduling and its relation to other network functionalities.

2.2 Evolution

We will discuss three 'families' of cellular technologies - GSM, UMTS and LTE - each having a distinct approach towards radio resource management. The technologies are presented in chronological order of their appearance in the arena of mobile networks.

2.2.1 GSM

The Global System for Mobile communications (GSM), developed by the European Conference of Postal and Telecommunications Administrations (CEPT), is the most widely adopted second generation (2G) technology. 2G systems replaced the initial first generation (1G) networks, which were analogue, mainly due to two advantages. 2G technologies are significantly more efficient in spectrum usage, which is beneficial for coverage, and they are also more robust since the original analogue signal is transmitted in digital form.

An interesting event in the development of GSM was the introduction of the Short Message Service (SMS). Its surprisingly high popularity among customers indicated that data services have high potential for large-scale adoption. This phenomenon was one of the factors behind a new research effort, dedicated to transforming existing (GSM) networks to provide higher flexibility in the support of data services. Two such modifications - General Packet Radio Services (GPRS) and Enhanced Data Rates for GSM Evolution (EDGE) - were developed for GSM. Both enhancements are data-traffic oriented and incorporated improvements to the GSM radio interface, which delivered higher data rates, higher spectral efficiency and lower cost of operation. Designed as backward compatible solutions both GPRS and EDGE are technological extensions of GSM, which kept upgrading costs for operators low.

2.2.2 UMTS

A more fundamental step towards increasing data rates and coverage was the development of UMTS, a completely new third generation (3G) technology specified by 3GPP's (3rd Generation Partnership Project) Release '99. UMTS is the 3G technology adopted most widely, including large telecom markets such as all European countries, Japan, Korea, the USA and China. The development of UMTS was a common effort of the various parties forming 3GPP - standardisation bodies, network operators and equipment manufacturers. Since 3GPP is responsible for 3G standardisation in general, it ensured alignment with other

approved 3G technologies such as cdma2000. Although the take-off of UMTS was troublesome due to the high introduction costs (expensive spectrum licences), currently UMTS is thriving. Even more so, operators introduced network upgrades commonly known as High Speed Packet Access (HSPA) in order to meet the increasing customer demands regarding data rates and capacity, see [29]. Furthermore, HSPA is more efficient in the use of the available spectrum and in the support of (multimedia) data services.

HSPA is a development effort of 3GPP and includes two technological enhancements - High Speed Downlink Packet Access (HSDPA) and Enhanced UpLink (EUL). HSDPA was standardized first in Release 5 of the UMTS standard, see [1], and addressed performance improvements on the *downlink* - the communication direction from base station to the users. Later 3GPP's Release 6, see [2], introduced EUL, also known as High Speed Uplink Packet Access (HSUPA), as the counterpart on the *uplink* - the communication direction from users to base station. Both HSPA technologies can be regarded as a refined link-layer UMTS-based technology. The most prominent difference with 'plain' UMTS is the sharing of the network resource among data users, see [29]. Sharing the medium has the potential for higher flexibility and efficiency of resource utilisation, which may be beneficial given the variable packet size typical for data applications. Generally, HSPA shares infrastructure with UMTS networks but it also moves some resource management functionality, e.g. scheduling, to network entities (base stations) closer to the end users. This also is motivated by a desirable higher flexibility in resource assignment.

2.2.3 LTE

Originally, cellular network technologies, e.g. GSM, were developed for the service of circuit-switched traffic, i.e. each users receives a dedicated radio channel for its transmission, see [57]. However, for the service of Internet-like traffic packet-switching is more appropriate since it provides more flexible resource allocation, which depends on the actual user demand. In order for these basic technologies to support packet-switched traffic enhancements were added (GPRS for GSM and HSPA for UMTS).

The first mobile cellular technology solely developed for packet-switched operation is LTE - the latest phase in the evolution of cellular networks. The larger available spectrum, e.g. up to 20 MHz, enables increased system capacity of LTE systems. What also makes LTE special is its ability to allocate radio resource in both frequency and time domain, which additionally benefits the extra capacity that LTE can provide. In fact the possibility to schedule also in the

frequency domain is an attractive choice due to the orthogonality between frequency carriers, i.e. users in the same cell sending at different frequencies do in principle not interfere with each other. Later on in Section 2.5.3 we will discuss how this is relevant for scheduling.

LTE is an initial step towards fourth generation (4G) technologies but is not yet fully capable to satisfy the data rate requirements set by 3GPP towards a 4G technology, see [6, 51]. 4G should provide Internet-like connectivity to both stationary and mobile users. Therefore, 3GPP has initiated a new effort, termed LTE Advanced, whose aim is to deliver on all requirements set towards a 4G technology, see [22, 51].

2.3 Basic concepts

In this thesis we focus on the role of scheduling for the uplink in UMTS/HSPA and LTE networks. Scheduling is the mechanism that decides how to distribute the radio resources over the users that have data to send (in the case of uplink). Its operation is related to other resource management mechanisms as well. Therefore, in the rest of the chapter we shortly present basic concepts of UMTS/HSPA and LTE radio networks relevant for our research and introduce in detail the scheduling mechanism. Note that UMTS/HSPA networks are already deployed and hence technical specifications are available. Therefore, our research on scheduling in the context of UMTS/HSPA is guided by these specifications. Although there are already first deployments LTE is not yet wide spread and the standard is still under development which leaves more design freedom.

In this section we start with a general discussion on the architecture and the main functionalities of a cellular radio network, including radio resource management mechanisms. Subsequently, in Section 2.4 and 2.5 we will zoom in on UMTS/HSPA and LTE respectively.

2.3.1 Cellular network architecture

In Chapter 1 we briefly commented that a mobile network consists of an access network and a core network, see Figure 1.2. The main tasks of the core network are switching and routing of traffic, connectivity management to external networks as well as mobility management, see [28]. Furthermore, the core network also performs administrative tasks such as storing user profiles and keeping statistics, which can be used for billing, performance evaluation

and network planning. Connectivity among the different core network entities as well as to the access network is typically realised via high capacity wire-line communication links.

The access network provides wireless connectivity to mobile stations (MSs) over set of radio channels (carriers). A specific network entity, e.g. base station (BS), is responsible for the assignment of radio resource over the mobile stations. Besides maintaining communication channels access network entities also perform tasks such as signal processing, evaluation of the communication channel and radio transmission and reception as well as mobility management, e.g. handover. The access network operates based on radio access techniques, which we discus below.

2.3.2 Multiple access schemes

A multiple access scheme allows several users to use the same medium to transmit their information. The most commonly used multiple access techniques in contemporary mobile networks are time division multiple access (TDMA), frequency division multiple access (FDMA) and code division multiple access (CDMA), see [60]. In TDMA, adopted by GSM, mobile stations use the same frequency band and their access to the channel is organised in successive time slots. Simultaneous transmissions in the same time but at different frequencies is supported by FDMA which allocates specific narrow frequency band(s) to a mobile station (MS). Concurrent transmissions in time and frequency are possible with a CDMA technique, as e.g. adopted by UMTS, in which case transmissions of different MSs are uniquely identified and efficiently retractable by distinctive channelisation codes. More detailed description of the schemes can be found in [57, 60].

With the development of beyond 3G technologies the orthogonal frequency-division multiple access (OFDMA) technology gained attention. OFDMA, see [30, 62], divides the available wide-band spectrum into orthogonal sub-carriers and a data stream is distributed over several sub-carriers. This strategy is beneficial for decreasing the negative impact of a single carrier with prolonged bad channel conditions. Main advantage of OFDMA-based systems, such as LTE, is offering higher data rates (as result of assigning several sub-carriers to the same user) and avoiding intra-cell interference of simultaneous transmissions due to sub-carrier orthogonality. Furthermore, OFDMA offers significant flexibility in resource allocation by means of opportunistically selecting sub-carriers with good propagation conditions. However, OFDMA poses certain implementation challenges (related to power conversion) at the mobile terminal, which is

disadvantageous in an uplink scenario.

Single Carrier Frequency Division Multiple Access (SC-FDMA) does not have the particular implementation issues of OFDMA and is therefore preferred for operation in the uplink. The lower complexity is the result of an additional signal processing step that is done before the user data is mapped to the sub-carriers. Similarly to OFDMA the radio spectrum is divided in sub-carriers and multiple sub-carriers allocation is possible. However, if a single user is assigned multiple sub-carriers they need to be contiguous, which resembles the concept of a single carrier (explaining the name of the technology). For more details we refer to [30], Chapter 4.

2.3.3 Radio signal propagation

The communication between mobile devices and the access network takes place over the radio interface. Since the radio signal propagates in the ether, its reception at the base station is vulnerable to radio environmental factors, most notably to path loss, fading and interference. Consequently, these factors are of significant importance for the spectrum usage efficiency, system capacity, service quality, etc. Hence, their effect should be carefully considered in the analysis of the performance of a radio access network.

Path loss is the reduction in power density of a radio signal as it propagates through space. It is the result of physic phenomena such as free-space loss, absorption and refraction, and it depends on the distance between transmitter and receiver. Several path loss models can be used for the planning of mobile networks, depending on the environment, e.g. urban city vs. rural. In this work we have chosen to apply the Cost 231 Hata path loss model for urban setting, see [28]. According to the model, the path loss $L(d_i)$, expressed in dB, is given by

$$L(d_i) = L_{fix} + 10a \log_{10}(d_i), \tag{2.1}$$

where L_{fix} is a parameter that depends on system parameter such as antenna height (of both base station and mobile) and carrier frequency, a is the path loss exponent (typically in the range of 2 to 4) and d_i is the distance between the communicating devices.

Taking into account the path loss, with given transmit power, the power at the receiver (expressed in dB) can be derived as

$$\text{Received power} = \text{Transmit power} - \text{Path loss}.$$

With a path loss exponent a higher than one this translates to faster than linear

decrease in signal strength (power) over the distance. More detailed discussion on the topic is provided in Chapter 3.

Fading reflects the variation in amplitude and phase that a radio signal experiences over its propagation path, see [60]. When the variations are a result of the transmitted signal travelling multiple paths, an effect caused by signal reflection, we talk about fast fading. Shadowing, on the other hand, is a consequence of the presence of obstacles on the signal path. User mobility leads to changes in both the fast fading and the shadowing profiles.

Other factors that affect the signal quality at the receiver are *thermal noise*, denoted N, and *interference*, denoted I. Major interference components are: intra-cell interference from parallel transmissions in the same cell and inter-cell interference from signals outside the cell. Given a signal strength S at the receiver, the quality of the (radio) communication channel is described by the signal-to-interference-plus-noise ratio ($SINR$), i.e. $SINR = S/(I + N)$. The experienced $SINR$ determines the data rate r that can be achieved on that channel. An idealistic relation between experienced $SINR$ and data rate is given by the well-known Shannon formula according to which $r = BW \log_2(1 + SINR)$, where BW is the channel bandwidth, see [30]. In a practical system however this data rate can not be achieved and corrections need to be introduced to describe system and implementation losses, e.g. [3].

2.3.4 Radio resource management

One of the main challenges in mobile networks is to utilise the limited radio spectrum as efficiently as possible while still providing the QoS requested by the users. As mentioned in Chapter 1, in mobile networks the management of the radio spectrum is the responsibility of the radio resource management mechanisms, which decide on radio transmission issues such as applied transmit power, scheduling order, channel allocation, etc. RRM is a necessary element of any mobile network technology which allows the selection of an appropriate radio network parameters considering the traffic load, user locations, QoS requirements, propagation conditions, etc. In the circumstances of node mobility RRM becomes even more important. Of the many RRM mechanisms we will discuss admission control, rate adaptation, power control and, most prominently, packet scheduling. For a broader overview of RRM please consult [28, 30, 60].

Admission control

Admission control ensures that at any time instant, under normal operation, there are no more active users in the network than the available radio resource can support. Admission control is executed any time an additional user requests service, an existing user wishes to upgrade its negotiated data rates or before handover occurs. Requests that could result in overbooking of the available radio resource such that the required QoS levels can not be achieved are rejected by the network.

Admission control mechanisms indirectly determine the number of active users in the cell and therefore provide the input of the scheduling schemes. In the context of our work we lay the focus on scheduling and thus assume that admission control operates correctly. Hence, we assume it to be transparent for the operation of the scheduling scheme. Therefore we will not discussed it in further detail.

Rate adaptation

The eventual goal of a mobile operator is to maximise the total system capacity and to provide to its users the negotiated QoS level given the current propagation conditions. External factors such as interference or fast fading can have negative impact on the received signal and thus data rate. This negative impact can be partly diminished by applying suitable levels of channel coding and modulation.

In *channel coding* additional redundancy information is added to the original data in order to make it more robust to disturbances in the radio communication channel. *Modulation* is the process of 'translating' an analogue or digital signal to a signal that can be conveyed over a physical medium. This modulated signal carries the original information in the form of signal 'states', where each 'state' is uniquely identified by a particular signal characteristic, i.e. amplitude, phase or frequency. For example, in amplitude modulation different signal amplitudes correspond to different 'states'. The range of uniquely identifiable signal 'states' (in the example amplitudes) determines the modulation order.

3G mobile networks adopt an *adaptive modulation and coding* (AMC) technique. The main idea of AMC is to dynamically change the modulation and coding scheme (MCS) depending on the transmission channel conditions of a particular user. Hence, the MCSs applied by different users are independently selected. The better the quality of the transmission channel the higher the MCS because errors are less probable to occur (requiring less redundancy information) and a larger range of signal 'states' can be distinguished (support high

modulation order).

AMC is a key technique that enables operators to achieve high spectral efficiency. It is initiated by the receiving side on a-priori known time frame and is based on feedback information on channel quality from the transmitting side. In the context of our research topic (scheduling) we assume that an appropriate MCS is always chosen and will not further elaborate on that. More details will be provided only when necessary for the understanding of the performed research. For a more extensive discussion on AMC please consult [60].

Power control

Power control is the mechanism responsible for the initial selection and subsequent maintenance of an appropriate transmit power level. Its main task is to find a 'trade-off' transmit power level for each individual user. On the one hand, it should select sufficiently high transmit powers in order to ensure successful reception at the receiver. On the other hand, it tries to minimise interference levels by limiting the transmit powers. The minimum required received power (and a corresponding $SINR$) is eventually determined by the QoS requirements of the particular application, see Section 2.3.3. Transmit powers, higher than the level that is absolutely necessary to reach the desired $SINR$, would only increase interference without significantly benefiting user performance. The variation in propagation conditions (including interference) turns power control into an important RRM component.

In systems with shared spectrum, higher interference (in the cell and outside) disturbs other (concurrent) transmissions and effectively decreases the total cell throughput. In such networks, e.g. UMTS/HSPA, power control, if applied, can lead to significant capacity improvements. OFDMA-based networks, e.g. LTE, are less susceptible to interference due to orthogonal frequency carriers. Therefore, for them the benefits of power control are less.

The importance of power control for downlink and uplink differs as well. Generally in both cases power control is beneficial for interference management. Specifically for the uplink however battery consumption is another driver for its adoption. Since mobile devices operate on batteries with finite capacity it is advantageous to maximise the lifetime. By trying to maintain the lowest transmit power sufficient for successful reception power control can positively influence battery lifetime.

Although power control is not the focus of the research in this thesis, we consider it important as it interacts with the scheduling in each of the studied mobile cellular technologies. This will be further discussed in Section 2.4.3 for

the case of UMTS/HSPA and in Section 2.5.3 for LTE.

Scheduling

The last RRM functionality we would like to discuss is scheduling. The main objective of a scheduler in a mobile network is to assign the available radio resource as efficiently as possible over the active users while satisfying their QoS requirements. Hence, it is the task of the scheduler to decide, among others, which users to serve, in what order to serve them and how much resource to assign to each user. Scheduling decisions are based on e.g. requested QoS, and channel conditions. Note that scheduling plays a prominent role in particular in shared channel (packet switched) systems, e.g. UMTS/HSPA; in systems with pre-determined fixed-size channel allocations, e.g. GSM, there is much less freedom and flexibility in resource allocation.

Typically channel access is organised in time in order to accommodate all active users. This is done on the base of time slots of fixed duration size, which are termed transmission time intervals (TTIs) in the networks we consider. The shorter the TTI the more often scheduling decisions are taken and the faster the scheduler may adapt to changing channel conditions. Obviously, schedulers that can quickly react to changes in the radio channel can exploit the radio resources more efficiently than slower ones.

Different classifications of scheduling schemes are possible depending on the chosen classification criteria. An important distinction is based on the role of channel conditions for the scheduling decisions. A scheduler that assigns resources only to users that can make best use of it, i.e. with favourable channel conditions, belongs to the group of *channel-aware*, or opportunistic, scheduling. Although such scheduler can efficiently use the (radio) resource and optimise the total cell throughput, it could starve users with poor channel conditions, particularly when these conditions prevail, e.g. when the user is stationary. An exception is the proportionally fair class of schemes that aim to improve fairness while still exploiting channel variation, see [35]. As belonging to the group of channel-aware schemes a proportional fair scheduler tries to maximise the total cell throughput but it adds an additional requirement - that each user is guaranteed a level of fairness. The fairness issue does not arise at all for *channel-oblivious* schedulers, which serve all active users independently of their channel conditions. Schemes from this group are however inherently less efficient since they assign resources also to users that cannot optimally use them. Among the advantages of channel-oblivious schedulers are the low complexity, straightforward implementation and fairness. At the same time opportunistic

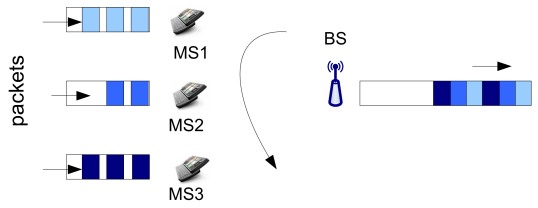


Figure 2.1: Round Robin scheduling scheme.

scheduling (especially on the uplink) is not expected to deliver large gain. The reason for the latter is the limited transmit power of a mobile device, which suggests that performance limitations are often posed by the devices themselves and not the propagation environment.

All scheduling schemes, studied in this thesis, belong to the group of channel-oblivious schedulers and, more specifically, are a type of Round Robin (RR) scheme. Round Robin serves all users with non-empty buffers (active users) in circular order and gives each user equal opportunity to access the radio channel. An example with three users is shown in Figure 2.1. Note, that fair channel access does not necessarily imply that the scheduling scheme yields equal bit rates to the different MSs. The actual data rates depend on individual channel conditions, e.g. MS's location in the cell and corresponding path loss. RR benefits from simplicity of implementation and fair channel access but, since it is a channel-oblivious scheduler, is characterised by non-optimal spectral efficiency.

Another classification of schedulers, relevant for our research, is based on the number of simultaneous transmissions within a TTI - single-user policy versus parallel access. As previously discussed, transmitters in the uplink, i.e. mobile stations, are battery-constrained power-limited devices. Bearing that in mind, assigning the total radio resource to a single MS might be inefficient if the MS is not able to fully utilise the resource. Therefore, in such scenario simultaneously serving several MSs in the same TTI will mostly have higher utilisation efficiency. Given that the spectrum is shared, concurrent transmissions in systems with non-orthogonal transmission channels, e.g. UMTS/HSPA, cause interference among the transmissions. As result certain signal degradation is introduced, which implies a trade-off between resource utilisation efficiency and user service. One of the research objectives of this thesis is to find out which of the two strategies is more advantageous in the uplink of shared spectrum networks without (UMTS/HSPA) and with (LTE) channel orthogonality.

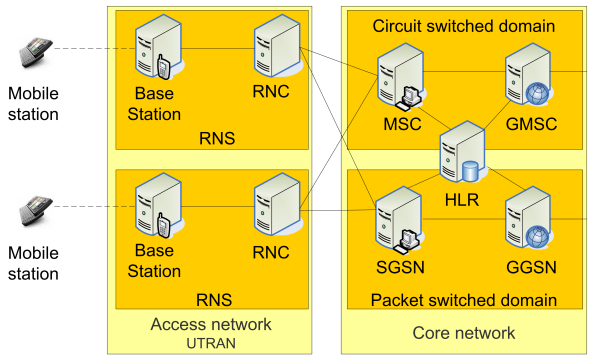


Figure 2.2: UMTS/HSPA network architecture, based on [29].

2.4 UMTS/HSPA cellular technologies

In this section we briefly outline the basic UMTS/HSPA network architecture, the associated radio access technology and some RRM functionalities with the main focus on HSPA. Scheduling is discussed in more detail in order to establish a base for the scheduling schemes, analysed in the upcoming Chapters 3 and 4.

The HSPA technology is an upgrade of the UMTS standard, intended to provide higher data rates, to improve cell capacity and decrease delays. As such they build on top of the basic UMTS network but introduce several architectural and operational modifications. The main difference in terms of functionality with a 'plain' UMTS network is the sharing of the communication channel resource among users. Furthermore, the channel access is organised on a time scale of 2 ms as opposed to the 10 ms in UMTS. Other relevant changes in HSPA are moving scheduling decisions to the base station and introducing faster physical layer retransmissions. More details on the operation and functionalities of UMTS/HSPA can be found in [29].

2.4.1 Network architecture

The UMTS/HSPA network architecture is presented in Figure 2.2. The core network comprises a circuit switched domain, serviced by the Mobile Switching Centre (MSC), and a packet switched domain, serviced by the Serving GPRS Support Node (SGSN), see [28]. The connectivity to external networks is provided by the Gateway MSC (GMSC) and the Gateway GPRS Support Node

(GGSN) respectively. Both domains use the administrative support of the Home Location Register (HLR), which stores for each customer information such as the subscribed services and billing profile.

The access network, formally termed UMTS Terrestrial Radio Access Network (UTRAN), is organised in one (or more) Radio Network Systems (RNS). A RNS consists of a single Radio Network Controller (RNC) and multiple base stations (BSs) also termed NodeBs. The RNC is responsible, among others, for admission control, congestion control and handover management. In HSPA a base station is given along with the standard air interface processing, e.g. channel coding, and certain operations related to power control, also responsibilities such as scheduling and dynamic resource allocation, see [28]. This is done baring in mind that a BS is closer to the users, which decreases communication delays and enables faster reaction to changing propagation conditions.

2.4.2 Multiple access scheme

Mobile stations communicate directly to the BS via the air interface by means of an access technology. The access technology for UMTS/HSPA is based on Wideband Code Division Multiple Access (WCDMA). Wideband comes from the fact that the data is spread over a wideband (5 MHz) carrier. This is beneficial for the robustness of the signal to fading persistently appearing over a single frequency. The spreading of the signal happens at the sender by multiplying the information flow with a spreading code, which converts the narrow-band information flow to a wide-band chip 'flow' at a chip rate of r_{chip} =3.84 Mchips/sec. The chip rate is generally much higher than the data rates that are supported. Given the data rate r requested by a user and the $SINR$ experienced on the channel, assigned to this same user, the resulting energy-per-bit-to-noise ratio E_b/N_0 can be calculated as:

$$\frac{E_b}{N_0} = \frac{r_{chip}}{r} SINR. \tag{2.2}$$

The ratio $\frac{r_{chip}}{r}$ of the chip rate r_{chip} and the data rate r of the actual data signal is termed *processing gain*.

WCDMA supports two basic modes of operation - Frequency Division Duplex (FDD) and Time Division Duplex (TDD), see [28]. In FDD uplink and downlink communications operate in separate 5 MHz frequencies bands, i.e. paired spectrum bands, with a guard band between them. In TDD only one 5 MHz

band is time-shared between uplink and downlink. In this thesis we assume the most widely deployed FDD mode of operation.

In 'plain' UMTS networks the information sent towards or coming from a particular user is carried by the dedicated transport channel (DCH). A DCH is exclusively assigned to a specific user and the resource reserved by that particular channel cannot be used by any other user. There are two sides to such channel assignment strategy. Although a certain data rate can be guaranteed no flexibility of resource allocation is supported. For example, data traffic is generally coming in bursts, meaning that at times the resources allocated to a particular channel (user) remain unused.

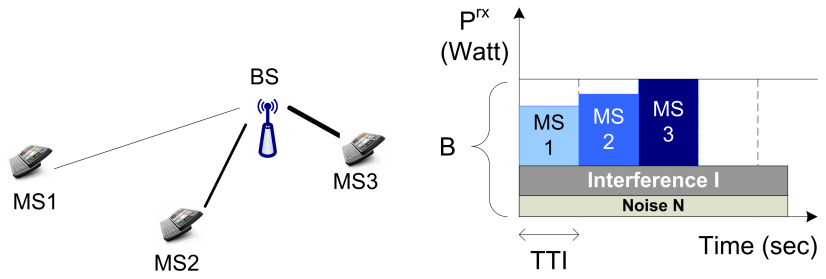
Allocation efficiency is improved in HSPA technologies due to faster resource management performed by the BS and sharing of the channel access. HSDPA introduces new data channel termed HS-DSCH (High Speed Downlink Shared CHannel) that is shared among the users. In EUL however each user receives an Enhanced DCH (EDCH) channel but all users share the access to BS (effectively resulting in a shared channel resource situation). Independently of the approach, sharing the channel access provides to the mobile network more flexibility into adapting to the user demands leading to improved spectrum utilisation. Recalling the above example, in HSPA a user with an empty transmission buffer does not get allocated any resources but they are offered to another user with data present in the buffer.

2.4.3 RRM mechanisms

In UMTS/HSPA the radio resource mechanisms are placed in the UTRAN entities (NodeB and RNC) and in the MS. MSs participate in the resource assignment on the uplink by delivering information on buffer occupation and channel quality. This information is used by the access network entities to perform resource allocation. The allocation decision is taken by the access network since it has information on ongoing traffic, channel conditions and available network resource for its belonging base stations.

Power control

In UMTS/HSPA networks power control for the uplink ensures that the transmit power used by the mobile is sufficient for the successful reception of its signal at the base station. In fact power control consists of several complex mechanisms and is performed on a frequent, short time-scale basis. In our study of scheduling for UMTS/HSPA networks, we assume that power control operates correctly and



(a) Cell population (b) Radio resource structure

Figure 2.3: EUL scheduling - an example with three mobile stations. The cell population, see (a), and radio resource utilisation by user signals and interference, see (b), are presented.

we do not look into the details of it. In particular, the transmit powers of mobile stations are adapted according to the instructions of the particular scheduling scheme, the experienced interference and the propagation conditions.

Scheduling

Scheduling decisions in UMTS/HSPA networks are taken at a scheduling period (TTI) of 2 ms and are responsibility of the base station. Decreasing the TTI from 10 ms in UMTS to 2 ms in HSPA enables faster resource reallocation (result of changes in the user's resource request) and quicker response to variations in the channel quality (results of changes in the propagation environment). Additionally, reaction time is won by moving scheduling from the RNC to the base station, which decreases the time necessary to communicate the scheduling decisions to the mobiles stations.

In HSDPA the base station is the transmitter and it typically has sufficient power capacity to fully use the total channel resource, available during a TTI, for the service of a single user. Therefore, a one-by-one scheduling strategy seems to be favourable because of low intra-cell interference. An HSDPA scheduler takes decisions such as which user to serve within a particular TTI, depending on the user's QoS requirements and channel conditions.

Scheduling in the EUL is inherently different. Unlike the BS, a battery-limited mobile station needs to preserve well its power capacity in order to increase operation lifetime. Furthermore, the transmit power capacity of a mo-

bile station is limited as well (and lower than that of a base station). As a result, typically a mobile station is not capable to fully use the total channel resource at the BS. In EUL the common radio resource is the totally supported received power at the base station, which we term *budget* B, see Figure 2.3(b). This budget is occupied by the 'useful' received powers from data users and by interference I as well as thermal noise N at the base station. The interference I includes intra-cel interference from non EUL transmissions within the cell and inter-cell interference from neighbour cells. Each TTI the total budget is available for the scheduler to serve users. Figure 2.3 shows an example of budget occupation for three users. Each user has different distance to the base station and hence received powers differ. All users however experience the same total interference, which in the presented examples is assumed to be time invariant.

The limited ability of a mobile to use the total budget suggests that a single-user scheduling strategy, if applied to EUL, may be inefficient and service in parallel might prove to be better. Concurrent transmissions however suffer from mutual interference (intra-cell interference) causing MSs to boost their transmit power (the increase being limited by the maximum transmit power) to overcome this interference. Hence, an optimisation of the scheduler may require a trade-off, i.e. fewer transmissions are more advantageous due to lower intra-cell interference but they may lead to unused radio resource This phenomenon only strengthens the negative effects of the cell on the radio resource available at its neighbours, i.e. increases inter-cell interference. The main goal of a scheduler is thus to improve network utilisation and MS's performance while keeping interference (in the cell and outside it) low.

The various, distinctively different approaches towards distributing the common resource over the active users turns scheduling in the EUL into an intriguing research topic. In the research presented in the later Chapters 3 and 4 we will discuss several possible scheduling strategies, depending on the desired performance objective.

2.5 Long Term Evolution - LTE

In this section we will concentrate on the second cellular technology discussed in this thesis - LTE. Again we begin with the network architecture and continue to briefly introduce the radio channels supporting wireless connectivity. The focus however falls on the radio resource management and in particular on scheduling.

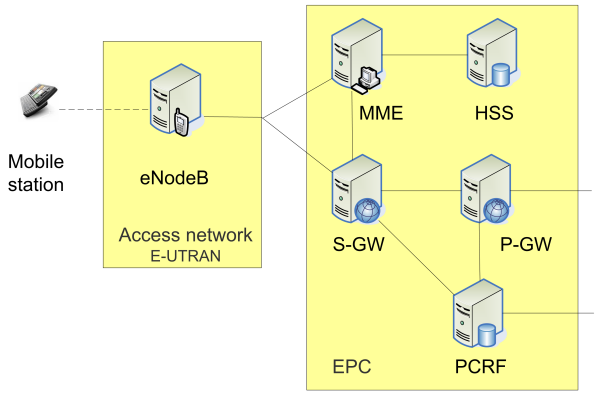


Figure 2.4: LTE-enabled network architecture.

2.5.1 Network architecture

Figure 2.4 shows, at high level, the network architecture of an LTE-enabled network with System Architecture Evolution (SAE) as the core network and Evolved UTRAN (E-UTRAN) as the access network. The most prominent difference with UMTS/HSPA in terms of functionality, which also affects the network architecture, is that 4G networks are oriented towards packet-switched services and therefore the core network does not include a circuit-switched domain.

The main SAE component, termed Evolved Packet Core network (EPC), does not include a circuit switched domain and hence no MSC- and GMSC-like entities. Generally the functionality of EPC is the same as the functionality of the packet switched domain in UMTS's core network. However, the distribution of responsibilities over the elements is quite different. The Serving Gateway (S-GW) and the Packet Data Network Gateway (P-GW), see Figure 2.4, are both defined for the user plane. S-GW manages all connectivity and switching for downlink and uplink in the user plane of its connected base stations while P-GW provides outside connectivity to external packet data networks. These two entities - S-GW and P-GW - communicate with the Policy and Charging Resource Function (PCRF) in order to ensure the appropriate handling of a particular user in terms of provided quality of services and applied charging. The Mobility Management Entity (MME) is the central control element in the EPC and is responsible for authentication and security, mobility management

and managing subscription profile and service connectivity, see [30]. The last element to discuss is the Home Subscription Server (HSS), which is in fact a database with the permanent information about each subscriber of the particular network. More detailed description of the core networks is provided among others in [30, 62].

The access network, termed (E-UTRAN), consists only of base stations, which in the context of LTE are termed eNodeB, see [30, 62]. Hence, there is no RNC-like entity. An eNodeB is responsible for all radio functionality and implements the RRM mechanisms performing tasks such as resource allocation, scheduling traffic according to required QoS and monitoring resource usage, see [30]. Additionally it also participates in mobility management, securing the radio messages and bearer handling. When necessary, e.g. in the event of handover, the serving eNodeB communicates to its neighbouring eNodeBs that are the potential new serving nodes for the connection in handover.

2.5.2 Multiple access schemes

The radio access technology, chosen by 3GPP for the LTE uplink, is based on SC-FDMA, see Section 2.3.2. The radio spectrum is organised into sub-carriers with spacing of 15 kHz. However, the minimum resource in the frequency domain that can be allocated is a group of 12 sub-carriers and is thus equal to 180 kHz, see Figure 2.5 and [30]. A group of 12 sub-carriers we term *sub-channel* and each user can receive multiple contiguous sub-channels up to the current bandwidth limit of 20 MHz. After the user data is mapped to the sub-carriers the resulting signal is fed to the time domain signal generation, which forms the SC-FDMA signal, see [30], Chapter 4. The smallest scheduling unit, considering both frequency and time domain is termed *resource block* (RB). More details on radio resource management and scheduling are presented later in Section 2.5.3.

User data from the mobile stations and its associated control information are carried by the Uplink Shared Channel (uplink-SCH). The control information over a particular connection can be classified into data-associated signalling (concerning the data processing, e.g. transport format) and non-data-associated signalling (concerning the connection status, e.g. scheduling requests). Another channel relevant for uplink transmissions is the Physical Downlink Control CHannel (PDCCH), see [62], that carries instructions from the base station towards the mobile, among others, about its allocated resource (number of sub-carriers) and power setting.

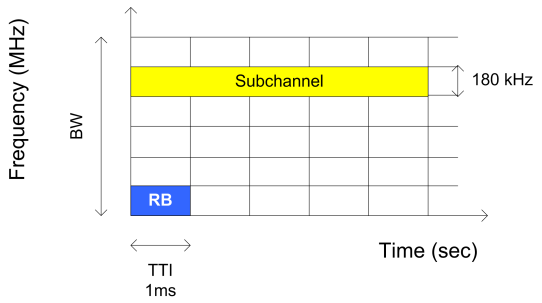


Figure 2.5: Radio resource structure in LTE networks.

2.5.3 RRM mechanisms

The LTE access network consists only of base stations, which implement all radio resource management mechanisms. Power control and particularly scheduling are the most relevant ones for our research in Chapters 7 and 8.

Power control

Due to the orthogonality of sub-carriers in LTE uplink interference among the carriers used in the same cell is avoided. However, interference between cells is still an issue. Recall that the purpose of power control in LTE uplink is to regulate the transmit powers in order to mitigate inter-cell interference. In fact, in LTE the power control mechanism at the BS informs the mobile station about the necessary received power spectral density, i.e. the power per resource block, such that eventually the requested QoS can be provided. Based on this information, the mobile terminal chooses itself a suitable transmit power per RB. In this selection process the mobile station considers the number of resource blocks (allocated by the scheduling scheme) over which the transmit power is distributed, the path loss in its communication channel and potentially fast fading.

Scheduling

A key feature of packet scheduling in LTE networks is the possibility to schedule users in two dimensions, viz. in time and frequency, see Section 2.5.2 and [30, 62, 51]. In SC-FDMA, as previously discussed, the radio spectrum is divided into in principle perfectly orthogonal sub-carriers. Hence, simultaneous transmissions

on sub-carriers in the same cell do not interfere with each other, which allows the scheduler to use also the frequency dimension.

The maximum total bandwidth currently specified for LTE is 20 MHz, while smaller bandwidths, e.g. 5 MHz and 10 MHz, are also possible. As we discussed in Section 2.5.2 the aggregate bandwidth is divided in sub-channels with a bandwidth of 180 kHz, see Figure 2.5. A sub-channel is the smallest allocation unit in the frequency domain. In the time dimension, the access to the sub-channels is organised in time slots of 0.5 ms. Two slots of 0.5 ms form a TTI (Transmission Time Interval). The smallest scheduling unit in LTE is the intersection of a 180 kHz sub-channel with a 1 ms TTI. In this thesis we refer to the smallest scheduling unit as a resource block[1]. The total number of RBs available to assign within a TTI depends on the chosen total system bandwidth, e.g. 10 MHz bandwidth corresponds to 50 RBs. Note that part of the bandwidth is lost to control channels and guard bands.

While in CDMA-based systems, e.g. UMTS/HSPA, users compete for a share in the available uplink noise rise, in OFDMA-based networks, e.g. LTE, transmissions compete for a share in the set of orthogonal sub-carriers (RBs). The allocation of resource blocks is managed by the scheduling mechanism in the base station each TTI. In principle a user can receive more than one RB, in which case its transmit power is divided over the allocated RBs. However, it is important to realise that there is a minimum transmit power per RB that is necessary for a successful reception at the base station. The issue is discussed in more detail in Chapter 7.

[1] Note that in fact a RB, also termed physical resource block (PRB) has dimensionality of 180 kHz and 0.5 ms. Hence, each TTI the scheduler assigns in fact groups of 2 RBs. Working with such terminology however complicates discussion and can lead to confusion.

Part I

Uplink scheduling in UMTS/HSPA networks

3

Flow-level analysis of packet scheduling for UMTS/EUL

3.1 Introduction

In Chapter 2 we have explained the role and challenges of packet scheduling being one of the main radio resource management mechanisms in 3G/4G mobile networks. The various design choices that can be made ask for careful evaluation and comparison of different scheduling strategies regarding their impact on the network capacity and the performance observed by the users, see the discussion in Section 2.3.4. In this first part of the thesis we will consider these issues in more detail in the specific context of UMTS/EUL.

As argued in the previous, introductory chapters, one of our main focus points in the evaluation of packet scheduling schemes is to capture the effect of the flow-level dynamics, i.e. the initiation of random size flow transfers by the network users at random time instants and locations, which causes the number of ongoing flow transfers to vary over time. In other studies available in the literature this aspect is mostly not considered. In the present chapter, and in Chapter 4, we intend to fill this gap and provide more insight into the impact of the random user behaviour on the performance of different uplink scheduling schemes. First, considering a single-cell scenario, we develop an analysis approach that combines analytical methods and simulation and enables fast evaluation of the performance measures of interest. Next, in Chapter 4, the study is extended to the situation with multiple cells and the focus is on the

question how the inter-cell interference, and its impact on the performance, depends on the specific scheduling scheme deployed in the network.

In the remainder of this section we will first give an overview of related work on the analysis of EUL scheduling schemes, see Section 3.1.1. Next, we briefly explain the main ideas of our analysis approach (Section 3.1.2) and present the main contributions of this chapter (Section 3.1.3). Finally, in Section 3.1.4, the organisation of the rest of the chapter is explained.

3.1.1 Literature

Most EUL performance studies in the literature are based on dynamic system simulations, see e.g. [27, 39, 55]. The underlying simulation models incorporate many details of the channel operation and traffic behaviour, but running the simulations tends to require a lot of time. Analytical modelling avoids time-demanding issues by abstracting from system details yet allowing the same (or even more) qualitative insights into the system performance. Unfortunately, in order to keep analysis feasible, such studies make sometimes unrealistic assumptions.

Independently of the evaluation approach, EUL studies which consider the flow dynamics are not easily found. An interesting study applying a simulation-oriented approach is [66]. The authors evaluate the performance of several schedulers which combine knowledge on channel quality and desired service goal, e.g. realised minimum data rate. Although insightful [66] is purely simulation-oriented study and do not provide analysis of the schedulers. Another study that similarly to our approach uses Markov models to represent the flow dynamics in a single cell is [21].

Several interesting examples of studies with analytical modelling are [35] and [53]. In [53] the authors examine the co-existence of voice and data traffic on the uplink with two approaches towards parallel transmissions for data traffic. Disadvantage of the research is the assumptions that the number of ongoing flows is fixed and that MSs have unlimited transmitted power capacity. A fixed number of users is also assumed by [35], which provides a very well worked out and optimal but complex scheduling solution. Though the authors also provide several simplified scheduling schemes for practical deployment they still might be difficult to implement.

There are only a few analytical studies on EUL performance that capture both scheduling detail and flow-level dynamics. Interesting references here are [21, 41] and [43]. In particular, [41] presents a rather complete study on the feasible service region supported by EUL and it takes into account for flow

dynamics and several relevant interference sources. However, the main goal of the authors is to compare resource management strategies, e.g. centralised vs. distributed, rather than particular scheduling schemes. The authors of [43] analyse two specific (rate-fair) scheduling disciplines including flow-level performance metrics but with the unrealistic assumption that the transmit powers of all mobiles are sufficient to reach the maximum bit rate.

In the present chapter we extend the model in [43] to the practical situation that the transmit power of the users is a limiting factor. In addition, we will also consider scenarios with two types of users: besides EUL (EDCH) users generating 'elastic' data traffic flows we assume the presence of 'plain' UMTS (DCH) voice users that share the same radio resources, see [53].

3.1.2 Modeling and analysis approach

Our modelling and analysis approach is based on time scale decomposition and captures both the system behaviour at packet level and the flow-level dynamics. In the packet level analysis we incorporate details of the scheduler's behaviour, e.g. power control, as well as characteristics of the transmission medium, e.g. channel conditions. In particular, characteristics such as received powers and realised data rates are then evaluated. The analysis on the packet level is for a fixed number of flows. The changing number of ongoing flows is accounted for in the flow-level analysis. As flow dynamics we identify flow initiations and completions due to the random user behaviour. The nature of these events allows the considered single cell system with data and voice flows to be modelled by a continuous-time Markov chain (CTMC). From the steady-state distribution of the Markov chain the performance measures of interest, i.e. mean flow transfer times and fairness, can be derived.

For some special cases, when the form of the Markov model allows it, explicit expressions for the steady-state distribution can be obtained. However, for more complex models closed-form expressions are not available. In these cases, standard techniques for deriving the steady-state distribution can be used, e.g. numerical solution of the balance equations or simulation of the Markov chain. Section 3.6.3 explains more on how Markov models can be efficiently simulated requiring relatively small running times.

Our proposed hybrid approach has several advantages: (i) by applying packet level analysis we are able to take into account the specifics of both scheduling scheme and environment; (ii) flow-level analysis captures the user's behaviour, which contributes to a more realistic modelling of the network; (iii) it supports fast evaluation. Furthermore, the approach is flexible since changes in the

scheduler or the environment require basically only modifications at the packet level.

3.1.3 Contributions

In this chapter we compare the performance of different EUL scheduling schemes. We are particularly interested in the influence of the random user behaviour leading to constantly changing number of ongoing flows. Our goal is to quantify performance measures such as mean flow transfer times and fairness, expressing how the performance depends on the MS's location in the cell. We concentrate on a single cell scenario with two types of traffic - 'plain' UMTS voice traffic and EUL data traffic. Voice traffic requires constant bit rate and minimum transmission delays. Therefore, a voice call is served with priority and is being served over a DCH channel, see Section 2.4.2. The EUL traffic is of best effort type and adapts to the available resources left over by the voice traffic. EUL flows are served over the more flexible EDCH channel. From now on we will use the terms voice call and data flow to distinguish between the two types of flows with their corresponding channels.

The main contributions of this chapter are:

- Developing a novel, hybrid analytical/simulation approach for fast evaluation of EUL scheduling schemes; it captures both the scheduling details at packet level and the impact of the flow-level dynamics.
- Providing analysis of different EUL scheduling schemes. The results stress the importance of modelling the flow-level dynamics.
- Evaluating the mutual influence data flows and voice calls have on each other and more specifically how voice traffic affects the performance of data traffic.
- Deriving several practical recommendations for deployment based on observed performance behaviour under different network settings.

3.1.4 Outline

The rest of the chapter is organised as follows. In Section 3.2 we first describe the scheduling schemes that are evaluated followed by a description of the model is Section 3.3. The general analysis methodology is presented in Section 3.4 while the analysis adapted to each of the schemes is discussed in Section 3.5.

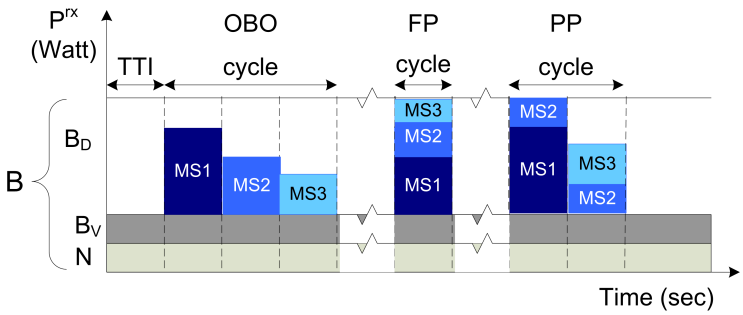


Figure 3.1: RR scheduling schemes. EUL users far away from the base station (MS3) have lower received power than EUL users closer by (MS1 and MS2).

Section 3.6 elaborates on several design and modeling issues relevant for the complete understanding of the numerical results presented in Section 3.7. Finally, in Section 3.8 we will summarise the work presented in this chapter.

3.2 Scheduling schemes

The purpose of a scheduler, as stated in Section 2.3.4, is to distribute the available channel resource over the flows, present in the system. An ultimate scheduler should provide maximised resource utilisation, high quality of service, e.g. high data rates, and equal treatment to all MSs, i.e. fairness. In fact very often a tradeoff is necessary. For example, assigning service only to the users that best can use the budget (have high received power) results in low fairness because other uses (with low received power) are barely given service.

In order to compare the effects of a chosen tradeoff we selected three scheduling schemes for evaluation, each scheme representing a major scheduling decision in terms of number of simultaneous transmissions. The three schemes are termed one-by-one, partial parallel and full parallel and are described in more detail below.

All schemes belong to the Round Robin family where mobile stations receive fair channel access independently of the actual channel conditions (channel oblivious scheduling), see Section 2.3.4. Note that fair channel access does not necessarily imply that a scheduling scheme yields equal bit rates for the different MSs; rather, each user is given equal chance to access the radio channel. The

experienced bit rates depend on the received powers, which differ based on the location of the MS (and hence on the experienced path loss) and on the applied scheduling scheme.

Recall that the channel access in UMTS/EUL is shared among all ongoing flows and is organised in time slots (TTIs), see Section 2.3.4. The total time, expressed in number of TTIs, necessary to serve all ongoing flows once we define as *(transmission) cycle*. We can expect that the different schemes will have different cycle lengths.

Further, the EUL channel resource is the total interference increase at the base station, which we will term from now on *(available) total budget* B. Only part of that budget is available for EUL data traffic since the rest is occupied by interference, e.g. thermal noise. Simultaneous transmissions from 'plain' UMTS voice calls are also interfering with EUL data transmissions and occupy a part, *voice budget* B_V, of the total budget. Eventually, the resource available for the service of EUL data flows we term *(available) data budget* B_D. Due to the distinct management approach taken by each scheduler the interference sources for each of the three schemes differ and hence the data budget B_D changes as well. More details on interference is provided later in Section 3.3.

We will now proceed to describe each of the considered schemes.

3.2.1 One-By-One scheme

In a one-by-one (OBO) scheme a single mobile station (data flow) is allowed to use the total data budget B_D during a single TTI and it can transmit at maximum power level, see Figure 3.1, [8]. As a result the length of the transmission cycle is easily defined as the number of currently ongoing flows. OBO is a strategy well known among operators and a relatively easy one to implement. Several studies have dedicated research on the scheme, e.g. [55].

A single transmission benefits from high achievable rate due to the lack of interference from other MSs. A MS with favourable channel conditions, i.e. low path loss, can fully utilise the available data budget on its own, which optimises throughput, see [53]. However, in the circumstances that a MS does not have the capacity to fully utilise the channel resource the scheme leads to under-utilisation. What part of the resource is unused depends on the user's position in the cell and on the experienced path loss. Thus the performance of the OBO scheme is highly sensitive to the channel conditions of the particular MSs and hence channel utilisation is a variable measure.

The particular ordering of the MSs to be served, e.g. in descending or ascending order of the received powers, is irrelevant for the performance within

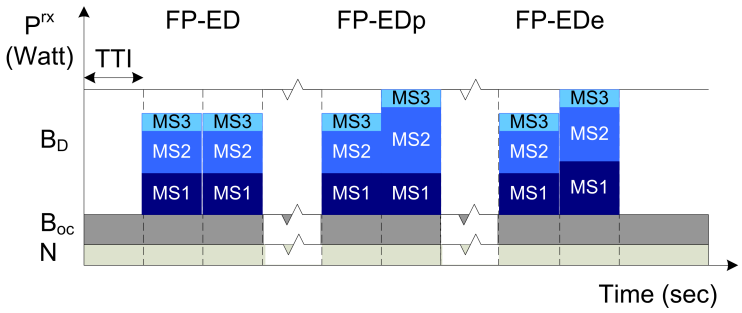


Figure 3.2: FP implementation with different approach towards power assignment. EUL users far away from the base station (MS3) have lower received power than EUL users closer by (MS1 and MS2).

the cell. We will show that in Section 3.5.1.

3.2.2 Full Parallel scheme

Another scheme discussed in the literature is the (full) parallel scheduler (FP). Its aim is the total channel resource utilisation. In this scheme all active users are given simultaneous access in one TTI, see Figure 3.1, [8], yielding cycle length of one TTI. Although FP achieves full channel utilisation, it also introduces a new interference source - all MSs transmitting together generate intra-cell interference each other.

Depending on the number of MSs and their received power two situations are possible. When the total resource requested by the MSs is lower than the data budget B_D no full channel utilisation is possible due to insufficient demand. In this situation each mobile station can use its maximum transmitted power. If the total requested resource is equal or higher than the available data budget full utilisation can be achieved. However, it may be that, in order for all transmissions to fit in the budget, the transmitted powers have to be decreased.

Several approaches to decrease the powers are possible. A FP scheme with proportional power division is one possibility in which the decrease portion is proportional to the original transmitted power. In particular, MSs with high transmitted power will have to decrease more and vice versa. Another possibility is an equal division when, as the name suggests, each MS receives equal portion of the total available power budget at the base station. We term this

strategy FP with equal division (FP-ED), see Figure 3.2. In case of equal power division a situation might occur when some MSs receive bigger power portion than they can actually make use of. Hence, part of the budget will be left unused while there may be other MSs, which are able to utilise the leftover. Therefore we propose two implementations that attempt to solve the issue. The FP-ED scheme with priority (FP-EDp), shown in Figure 3.2, takes the unused budget and assigns it to MSs that have extra transmitted power starting from the furthest located one. The second scheme - FP-ED with equality (FP-EDe) - redistributes, again equally, the left over budget over all MSs that have extra transmitted power. This is an iterative process and stops when all budget has been distributed. Note that the differences between the two implementations (FP-EDp and FP-EDe) and FP-ED (without reassignment of unused budget) show only when the total requested resource is higher than the total available data budget.

3.2.3 Partial Parallel scheme

An attempt to avoid the drawbacks of OBO and FP but to preserve their advantages is a combined scheme termed partial parallel (PP). According to it multiple active users can simultaneously access the channel in one TTI, see Figure 3.1. Hence, the PP scheme can achieve total channel utilisation by assigning maximum transmitted powers to MSs but at the cost of a total service time generally longer that one TTI. How many users are served in a single TTI depends on the available power budget and on the received powers of the particular MSs, selected for transmission. For example, high received powers results in few users per TTI and vice versa. Therefore, the cycle length in PP varies depending on the number and location of active users.

The PP scheduler selects MSs for transmission in the same TTI until the cumulative requested resource reaches the available data budget B_D. All users are assigned their maximum transmitted powers. Should the last MS selected for service in the TTI not fit with its complete received power its transmission is divided over two consecutive TTIs.

A PP scheme introduces intra-cell interference as well. However, the number of interfering sources within one TTI, i.e. transmitting MSs, is smaller when compared to a FP scheme.

Also in the case of partial parallel several scheduler orderings of the MSs are possible but they, just as in OBO, do not influence the performance within the cell, see Section 3.5.3. Note that this conclusion is correct in the light of the assumption that transmitting in one TTI at the maximum power delivers the

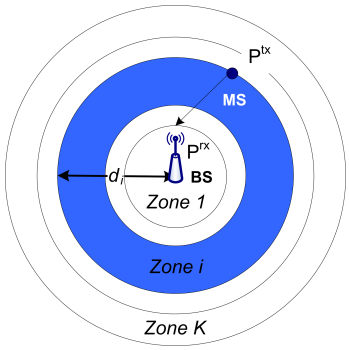


Figure 3.3: Single cell model - cell division into K zones.

same result as transmitting in two TTIs but at lower powers.

In a preliminary qualitative performance comparison of the three scheduling schemes we expect the PP scheduler to perform best. Which of the other two schedulers is best primarily depends on the available data budget: if the budget is relatively low, the OBO scheduler is expected to outperform the FP scheduler since it experiences no interference from other EUL users; if the budget is high, the FP scheduler is likely to better utilise the available budget. In terms of operational complexity and computation the three schedulers differ - OBO being the least complex and PP the most. However, compared to other EUL functionalities, e.g. power control, the level of complexity is relatively low.

3.3 Model

In this section we describe the network scenario we evaluate along with the assumptions underlying the later analysis. The model of a single cell presented in this section is generic for the whole thesis. Each of the models in Chapter 4 to 8 are based on this model but appropriately adapted to reflect the specifics of the corresponding scenario. These modifications (extensions) will be explicitly announced.

System level view

We consider a single cell with an omnidirectional base station. Uplink traffic originates in 'plain' UMTS mobile stations generating voice traffic and EUL-enabled mobile stations generating data traffic. We assume that a mobile station generates a single flow.

As illustrated in Figure 3.3, the cell is split up into K concentric zones. Zone i is characterized by a distance d_i to the base station and a corresponding path loss denoted $L(d_i)$, $i = 1, ..., K$, see Equation (2.1). The distance d_i is measured from the outer edge of the zone. All MSs within the same zone i are assumed to have the same distance d_i to the base station. Such division allows us to capture the impact of user's position on EUL traffic performance. Voice flows are not influenced by the distance since they require a fixed data rate of 12.2 kbps, which can be realised from any point of the cell. Hence, all voice calls have the same received power at the base station.

The system can be described by a full system state $\underline{n} \equiv (n_1, n_2, ..., n_K, n_V)$ where n_i describes the number of ongoing EUL flows in zone i, $i = 1, ..., K$, and n_V is the total number of voice calls. With n we denote the total number of ongoing data flows, i.e. $n = \sum_{i=1}^{K} n_i$.

The total received interference budget at the base station is denoted B and can be derived from an operator-specified noise rise target η, i.e. $B = \eta N$. Additionally, we set up a voice budget B_V which to model the presence of voice calls in the cell. At any time, EUL data flows may fully use that part of the total budget B that is not claimed by thermal noise and voice traffic. Hence, the data budget B_D can be derived as $B_D = B - (N + B_V)$. Note that the data budget B_D is implicitly dependent on the number of ongoing voice calls n_V.

For a data flow interference is all signals that degrade its own reception. In the current research as possible interference is identified: *(i)* the thermal noise level N; *(ii)* the self-interference modelled by parameter ω, which is due to the effects of multi-path fading; *(iii)* the interference $I_D(\underline{n})$ originating from EUL data flows; and *(iv)* the interference $I_V(n_V)$ originating from voice calls.

User level view - data flows

Data flows are generated according to a spatially uniform Poisson arrival processes with rate λ. As a direct consequence of the uniformity assumption, the probability q_i that a generated EUL data flow appears in zone i, is calculated as the ratio of the area of zone i and the total cell area, such that the data flow arrival rate λ_i in zone i is equal to λq_i, $i = 1, ..., K$. The data flow size is

exponentially distributed with mean F (in bits).

All mobile stations have the same maximum transmit power $P_{\max}^{tx}$ but different maximum received power at the base station $P_{i,\max}^{rx}$ due to the zone-dependent path loss. No MS mobility is modelled. The bit rate at which a MS is served depends on the experienced signal-to-interference-plus-noise ratio $SINR = S/(I + N)$ at the receiver. Assuming a prefixed E_b/N_0 (energy-per-bit to interference-plus-noise-density ratio) requirement, the attainable bit rate an be derived from Equation (2.2) and is equal to $r = (r_{chip} SINR)/(E_b/N_0)$, where $r_{chip} = 3840$ kchips/s denotes the system chip rate, see [28]. Note that in practise data rates are not continuous but belong to a set of discrete values. The signal level S is determined by the MS's transmitted power and the zone-dependent path loss.

User level view - voice calls

Voice calls are characterised by a constant bit rate and hence require the same received power at the base station P_V, regardless of the specific location of the voice traffic source. During the calculation of P_V we assume a worst case scenario in which the noise rise target η is fully utilised, which is indeed the objective of the EUL scheduler and hence a rather harmless assumption. Given a P_V the above-mentioned voice budget B_V is easily translated to a maximum number of admissible voice calls M_V, i.e. $M_V = \lfloor B_V/P_V \rfloor$, where M_V is increasing in B_V. Since we are interested in the performance of data traffic it suffices to keep track of the total number of voice calls in the cell. Voice calls are generated according to a spatially uniform Poisson arrival processes with rate λ_V. The voice flow duration is exponentially distributed with mean τ (in seconds).

The fixed rate voice flows are treated with priority by the scheduler and are admitted as long as the predefined voice budget B_V allows it. If the B_V is reached no newly arriving voice flows are accepted. Note that it is not necessary that all B_V is utilised. Consequently, the dynamics of the voice flows, i.e. flow initiation and completion, can be described by an $M/M/m/m$ queuing model (Erlang loss model), which is independent of the EUL data flow dynamics. For this Erlang loss model explicit expressions are known that relate the traffic load and the channel capacity to the blocking probability.

3.4 Analysis approach

Our approach to evaluating the performance of the data traffic consists basically of three steps, see also the high level description in Section 3.1.2. The first two steps capture the details of the scheduler's behaviour in a given state of the system. In particular, in the first step the data rate at which a scheduled EUL user can transmit is determined; the second step determines the users average throughput by accounting for the repetition rate at which the user is scheduled for transmitting data. In the third step these throughputs and the rates at which new 'plain' UMTS and EUL users become active are used to create a continuous time Markov chain describing the system behaviour at flow level. From the steady-state distribution of the Markov chain performance parameters, such as mean file transfer time of a user, can be calculated. For deriving the steady-state distribution standard techniques can be used, e.g. numerical solution of the balance equations or simulation. As the jumps in the Markov chain only applies to the initiation or completion of flow transfers, simulation of the Markov chain appears to be a very attractive option (note that the packet level details are captured in the transition rates which are calculated analytically) and does not suffer from the long running times of the detailed system simulations used in many other studies. For some special cases (scheduling strategies) the Markov chain's steady-state distribution can be solved explicitly yielding closed-form expressions for the performance measures, which will be given where applicable.

In this section we present the generic steps in the analysis when applied to the model described in Section 3.3. This generic analysis approach forms the basis of the modelling approaches applied in the upcoming Chapters 4 to 8. Subsequently in Section 3.5 we provide scheduler-specific expressions incorporating the details of each scheme.

3.4.1 Received powers

Given a maximum transmit power P_{max}^{tx}, the received power P_i^{rx} of any MS from zone i can be derived from the specific path loss $L(d_i)$, converted to linear units, as follows:

$$P_i^{rx} = \frac{P^{tx}}{L(d_i)}. \tag{3.1}$$

The transmitted power should be such that it does not cause overload of the available budget, thus it is limited by either the maximum power of the mobile

station (for far MSs) or by the available B_D (for close by MSs):

$$P^{tx} = min(P^{tx}_{max}, B_D L(d_i)). \tag{3.2}$$

We apply the Cost 231 Hata path loss model, namely $L(d_i) = 123.2 + 10a\log_{10}(d_i)$ (in dB) where a is the path loss exponent and d is the distance in kilometres, see Section 2.3.3 and [28]. Note that the path loss expression is given in dB while in the rest of the analysis a linear scale is used.

3.4.2 Instantaneous rate

The data rate achieved by a flow from zone i considering only the transmission channel conditions, i.e. received power and interference, is the *instantaneous rate* $r_i(\underline{n})$. Hence, this is the rate realised during one TTI. The instantaneous rate is zone-specific i.e. depends on the distance d_i, and differs per scheduler. Each scheduler has a particular strategy towards power assignment and simultaneous transmissions leading to a specific intra-cell interference pattern. This pattern depends on the current system state $\underline{n}$ and hence so does the instantaneous rate. We define $r_i(\underline{n})$ of a data flow as a generalisation of [28] eq. (8.4):

$$r_i(\underline{n}) = \frac{r_{chip}}{E_b/N_0} \cdot \frac{S}{I+N}. \tag{3.3}$$

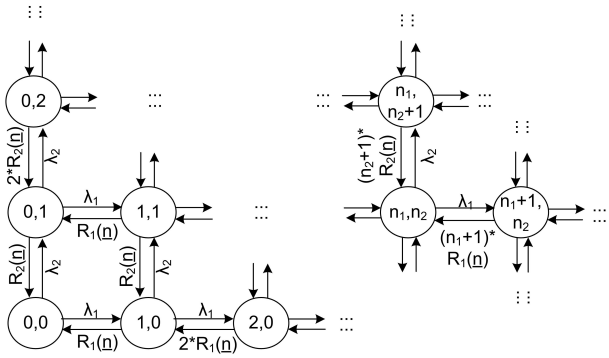
Taking into account the particular interference sources considered in the current evaluation scenario, see Section 3.3, we can rewrite:

$$r_i(\underline{n}) = \frac{r_{chip}}{E_b/N_0} \cdot \frac{P^{rx}_i}{I_D(\underline{n}) - \omega P^{rx}_i + I_V(n_V) + N}. \tag{3.4}$$

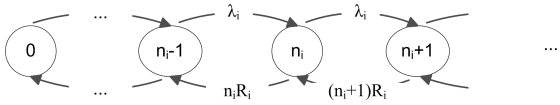
$I_D(\underline{n})$ is the total received power of all data flows served in the same TTI. By definition it includes received power of the flow that we currently evaluate. Therefore, in order to model self-interference properly we need to subtract a fraction ω of the own signal from $I_D(\underline{n})$. $I_D(\underline{n})$, for the different scheduling schemes, is given by

$$I_D(\underline{n}) = \begin{cases} P^{rx}_i & \text{for OBO,} \\ min\{\sum_{i=1}^{K} n_i P^{rx}_i, B_D\} & \text{for PP and FP.} \end{cases}$$

Note that the interference I_D can never exceed the available budget B_D at the base station, which could potentially happen only in the PP and FP schemes.



(a) Two-dimensional Markov chain



(b) Dimension i

Figure 3.4: Markov chain modelling: (a) An example for cell division in two zones, each represented by a dimension in the Markov chain; (b) transitions in a single dimension i.

3.4.3 State-dependent throughput

Knowledge of $r_i(\underline{n})$ is not sufficient to determine a flow's throughput in system state $\underline{n}$ since a flow often has to wait several TTIs between actual data transmissions. The result is a decreased actual transmission rate, which we term *state-dependent throughput* $R_i(\underline{n})$. More precisely, $R_i(\underline{n})$ is the average transmission rate that an active call achieves during one scheduling cycle, given that the system is and remains in state $\underline{n}$. Denoting with $c(\underline{n})$ the cycle length, which depends on the number of ongoing EUL data flows and the applied scheduling scheme, we have

$$R_i(\underline{n}) = \frac{r_i(\underline{n})}{c(\underline{n})}. \tag{3.5}$$

3.4.4 Flow throughput

Cell dynamics are determined by the arrival rate of new flows and the rate at which flows are served. To reflect these events in our analysis we use continuous time Markov chains which provide natural mapping to the system behaviour. A state in the Markov model corresponds to the full system state $\underline{n} = (n_1, n_2, ..., n_K, n_V)$, i.e. the distribution of the EUL data flows over the different zones in the cell and the total number of on-going voice flows. New flow arrivals are modelled by forward jumps in the model and flow completions by backward jumps. The eventual Markov model has $K+1$ dimensions, with K dimensions representing the data flows evolution and an additional dimension for the voice calls evolution. Each of the K data dimensions is unlimited in the number of admissible calls, while the 'voice' dimension is limited to M_V simultaneous calls, see Figure 3.4. The transition rates of the Markov model are as follows:

$$\begin{array}{llll} \underline{n} \to (n_1, ..., n_i + 1, ..., n_K, n_V) & \text{at rate} & \lambda_i & \text{(EUL flow arrival)}, \\ \underline{n} \to (n_1, ..., n_i, ..., n_K, n_V + 1) & \text{at rate} & \lambda_V & \text{(voice call arrival)}, \\ \underline{n} \to (n_1, ..., n_i - 1, ..., n_K, n_V) & \text{at rate} & \frac{n_i}{F} R_i(\underline{n}) & \text{(EUL flow completion)}, \\ \underline{n} \to (n_1, ..., n_i, ..., n_K, n_V - 1) & \text{at rate} & \frac{n_V}{\tau} & \text{(voice call completion)}. \end{array} \tag{3.6}$$

Once we have constructed the Markov model of a scheduler we can determine its steady-state distribution and derive parameters such as flow throughout or mean flow transfer times. In some special cases, when the resulting Markov model belongs to a class of already well-studies models, the steady-state distribution is provided by closed-form expressions. Otherwise, in the more generic case, the distribution can be found by simulating the model, which we explain how to do in Section 3.6.3.

3.5 Scheduler-specific analysis

The generic approach described in Section 3.4 can be used to analyse various schedulers. Applied to the three scheduling schemes of our interest the approach results in three different Markov models that can be classified as complex processor sharing type of queuing models. The differences are a consequence of the different expressions of $r_i(\underline{n})$ and $c(\underline{n})$. Due to the specifics of the schedulers, the Markov chains generated here are too complex for the steady-state distribution to be obtained analytically. Therefore we have chosen to simulate the Markov model in order to find the steady-state distributions.

3.5.1 OBO scheme

In case of OBO scheduling only one data flow may be scheduled per TTI. Hence the scheduled flow may in principle utilise the entire data budget B_D but may very well be limited by its own maximum transmitted power, see Equation (3.2).

Having only a single active flow per TTI yields a cycle length of $c(\underline{n}) = n$ where $n = \sum_{j=1}^{K} n_j$ (for state $\underline{n}$). Under OBO scheduling the sources of interference are thermal noise, interference from voice traffic and self-interference. $I_D(\underline{n})$ in this case consists only of the own signal power, which allows Expression (3.4) to be rewritten and the resulting state-dependent throughput $R_i(\underline{n})$ becomes:

$$R_i(\underline{n}) = \frac{r_{chip}}{E_b/N_0} \cdot \frac{P_i^{rx}}{(1-\omega)P_i^{rx} + I_V(n_V) + N} \cdot \frac{1}{n}. \tag{3.7}$$

Expression (3.7) shows that $R_i(\underline{n})$ depends on the current state $\underline{n}$ only via $c(\underline{n})$ and n_V. This also clarifies why the particular ordering of the MSs does not affect the final performance.

Substituting the arrival rate of the Poisson process and the state-dependent throughputs we can construct the Markov model of the OBO scheme. Due to the dependency of $R_i(\underline{n})$ on the interference from voice calls the model is not trivial and its steady-state distribution is more easily derived by simulation. In the special case where $\lambda_V = 0$ however, the Markov model effectively becomes a multi-class $M/M/1$ processor sharing (PS) model, which we will shortly describe below.

M/M/1 processor sharing model

When there are no voice calls present in the system the Markov model of the OBO scheduler appears to have the same form as the Markov chain describing an M/M/1 processor sharing model with multiple customer classes, see e.g. [10]. Each class actually corresponds to one of the zones in the cell model, see Section 3.3. The model is of processor sharing type due to the equal sharing of the access (in terms of time) to the EUL data budget among the active users. In more detail, the correspondence between the two models is seen from the similarities of the (form of the) transition rates of the Markov chains.

In particular, given the condition $n_V = 0$ the state-dependent throughputs which represent the transition rates in the Markov chain become:

$$R_i(\underline{n}) = \frac{r_{chip}}{E_b/N_0} \cdot \frac{P_i^{rx}}{(1-\omega)P_i^{rx} + N} \cdot \frac{1}{n}, \tag{3.8}$$

where $i = 1, ..., K$. Hence $R_i(\underline{n})$ depends only on the total number of active users in the cell. Upon substituting the Poisson arrival rates λ_i and the state-dependent throughputs $R_i(\underline{n})$ as transition rates in the model, and mapping them to the M/M/1 PS model we obtain the following expression for the steady-state distribution (cf. the results in [10]).

$$Pr\{N_1 = n_1 ... N_k = n_k\} = (1 - \rho) \cdot \frac{(n_1 + ... + n_k)!}{n_1! ... n_k!} \cdot \prod_{i=1}^{K} \rho_i^{n_i},$$

where ρ is defined as $\rho = \sum_{i=1}^{K} \rho_i$ and $\rho_i = \lambda_i / r_i(\underline{n})$. From the steady-state distribution, the mean flow transfer time T_i of a MS in zone i can be obtained via Little's formula. see e.g. [26]. After some calculations we find:

$$T_i = \frac{F}{r_i(\underline{n})} \cdot \frac{1}{1 - \rho}. \tag{3.9}$$

3.5.2 FP scheme

Under FP scheduling, an active user transmits in each TTI and therefore the cycle length $c(\underline{n})$ is equal to 1 for all states $\underline{n}$. Hence the state-dependent throughput is equal to the instantaneous rate (calculated for the appropriate received power level), i.e. $R_i(\underline{n}) = r_i(\underline{n})$. In terms of budget utilisation we distinguish between two cases. In the first case the number of data flows is such that the sum of their (received) maximum powers is lower than the data budget B_D. In this case the interference $I_D(\underline{n})$ comprises contributions from all ongoing data flows, i.e. $I_D(\underline{n}) = \sum_{i=1}^{K} n_i P_i^{rx}$. The second case is when the summed maximum received power from all data flows is higher than the data budget. Since all MSs are assigned to transmit in parallel, the transmitted power levels have to be decreased such that the summed received powers fit in B_D. The resulting received power levels P_i^{rx} are derived from the maximum received power via a proportional decrease:

$$P_i^{rx} = \frac{P_{i,\max}^{rx}}{\sum_{i=1}^{K} n_i P_{i,\max}^{rx}} \cdot B_D, \tag{3.10}$$

so that $I_D(\underline{n}) = B_D$. Using P_i^{rx} we can rewrite Expression (3.5) for this second case of FP scheduling as

$$R_i(\underline{n}) = r_i(\underline{n}) = \frac{r_{chip}}{E_b/N_0} \cdot \frac{P_i^{rx}}{B_D - \omega P_i^{rx} + I_V(n_V) + N}. \tag{3.11}$$

Since $R_i(\underline{n})$ depends on the full state information $(n_1, n_2, ..., n_K, n_V)$ the resulting Markov chain model for the FP scheme is also a complex processor sharing type of queueing model, for which we apply Markov chain simulation to obtained the performance results.

Power assignment

As discussed in Section 3.2 we propose a FP scheme with proportional power division and three implementations of a FP scheme with equal power division, namely FP-ED, FP-ED with priority (FP-EDp) and FP-ED with equality (FP-EDe). Recall that the schemes with equal power division initially divide the budget in equal portions, whose size B_D/n is determined by the number of active users n, and each user gets assigned one such portion. Depending on the match between the budget portion and the received power of a user two situations are possible. In the first case all users have received power such that they can fully use their assigned portion, i.e. $P_i^{rx} \geq B_D/n$. In the second case this condition does not hold for at least one user, i.e. $P_i^{rx} < B_D/n$, leaving $U(\underline{n})$ part of the budget unused. The size of $U(\underline{n})$ depends on how many users cannot fill up their portion and on their received powers. $U(\underline{n})$ can be derived as:

$$U(\underline{n}) = \sum_{j=1}^{n} \left(\frac{B_D}{n} - P_j^{rx} \right). \tag{3.12}$$

Each scheme adopts a different approach towards sharing $U(\underline{n})$ over the users. FP-ED leaves $U(\underline{n})$ as it is and by this does not exploit all available resources even if other users, let's term them *candidates*, could use the extra resource. Both FP-EDp and FP-EDe apply a more advanced strategies and try to fully use the budget. In FP-EDp the candidate most far from the base station is given the whole $U(\underline{n})$ while FP-EDe equally divides $U(\underline{n})$ over all candidates. The redistribution of $U(\underline{n})$ in both schemes is in fact an iterative process.

Accounting for the specific power assignment in the different FP schemes with equal division, we can write for the received powers:

$$P_i^{rx} = min \left(\frac{B_D}{n} + E(n), B_D \right), \tag{3.13}$$

where

$$E(n) = \begin{cases} 0 & \text{for FP-ED and FP-EDp,} \\ U(\underline{n}) & \text{for FP-EDp, farmost user,} \\ \dfrac{U(\underline{n})}{n_{cand}} & \text{for FP-EDe,} \end{cases}$$

where n_{cand} is the number of candidates. Consequently Equation (3.11) still applies but with the power levels given by Equation (3.13).

3.5.3 PP scheme

As the PP scheduler allows parallel transmissions of multiple EUL data flows in a single TTI, $I_D(\underline{n})$ comprises the interference contributions from all scheduled data flows in a TTI. Consequently, the instantaneous rate depends on the state $\underline{n}$, see Equation (3.3). The opportunity for simultaneous transmissions results in a shorter cycle than under OBO scheduling which can be expressed as the ratio of the aggregate resource requested by all ongoing EUL flows and the available data budget B_D. The cycle length is then given by

$$c(\underline{n}) = \max\left\{1, \frac{\sum_{i=1}^{K} n_i P_i^{rx}}{B_D}\right\}. \tag{3.14}$$

In the PP scheduler, similar to FP, we distinguish between two utilisation cases. In the first case the user demand is lower than the available data budget, i.e. $\sum_{i=1}^{K} n_i P_i^{rx} < B_D$, each active user sends in each TTI. Hence, the cycle length is $c(\underline{n}) = 1$. That simplifies the state-dependent throughput expression to

$$R_i(\underline{n}) = \frac{r_{chip}}{E_b/N_0} \cdot \frac{P_i^{rx}}{I_D(\underline{n}) - \omega P_i^{rx} + I_V(n_V) + N}. \tag{3.15}$$

In the second case, when the demand is equal or higher than the data budget, i.e. $\sum_{i=1}^{K} n_i P_i^{rx} \geq B_D$, the state-dependent throughput is equal to

$$R_i(\underline{n}) = \frac{r_{chip}}{E_b/N_0} \cdot \frac{P_i^{rx}}{B_D - \omega P_i^{rx} + I_V(n_V) + N} \cdot \frac{B_D}{\sum_{i=1}^{K} n_i P_i^{rx}}. \tag{3.16}$$

Note that, the particular ordering of the MSs does not affect performance since the intra-cell interference from other data transmissions stays the same.

The resulting flow-level Markov chain model is a complex processor sharing type of queuing model with transition rates that depend on the detailed state distribution $(n_1, n_2, ..., n_K, n_V)$.

Table 3.1: Cell dimensioning in zones

Zone number	1	2	3	4	5	6	7	8	9	10
Zone radii (km)	0.97	0.98	0.99	1.02	1.06	1.11	1.20	1.39	1.49	1.65
Bit rate (kbps)	4096	3584	3072	2560	2048	1536	1024	512	384	256

3.6 Preliminaries for numerical studies

In this section we elaborate on several modeling issues which form the basis for the numerical results presented later in Section 3.7. In particular, in Section 3.6.2, we argue on the appropriate number of zones used for cell modelling while in Section 3.6.3 discusses our methodology to simulate Markov model in Matlab environment. But first, in Section 3.6.1 all traffic and system parameters are set and the performance measures are introduced.

3.6.1 Parameters settings

In the numerical experiments we work with (predetermined) system chip rate r_{chip} of 3840 kchips/s, a thermal noise level N of -105.66 dBm and a noise rise target η at the base station of 6 dB. From these parameters the total received power budget B can be calculated: $B = \eta N$. A self-interference of 10% of the own signal is considered, i.e. $\omega = 0.9$. The assumed path loss model is given by $L(d) = 123.2 + 35.2 \log_{10}(d)$ (in dB), e.g. [28].

In order to capture the effect of the distance on performance we divide the cell into $K = 10$ zones. This particular choice is motivated in Section 3.6.2. Given an E_b/N_0 target of 1.94 dB for EUL transmissions, a maximum transmission power of $P^{tx}_{\max} = 0.125$ Watt and a worst case interference level (where the received power budget B is fully used), we applied straightforward link budget calculations to determine the zone radii corresponding to a set of ten bit rates[1] between 256 kbit/s (zone 10) and 4096 kbit/s (zone 1), see Table 3.1. EUL flows are characterised by mean file size $F = 1000$ kbit and a default aggregate rate at which new flow transfers are initiated $\lambda = 0.4$.

UMTS users are assumed to generate voice calls with requested bit rate of 12.2 kbit/s, an activity factor of 50% and an E_b/N_0 of 5.0 dB. The mean duration τ of the voice calls is 120 seconds. Given a noise rise target of 6 dB, this translates to a received power P_V of $0.0729 \cdot 10^{-14}$ Watt. A budget B_D of 70% of the totally available channel resource B was used as a default value, implying a maximum of 77 simultaneous voice calls. Given a target blocking

[1] These bit rates are realised given that a user transmits on its own.

probability of 1%, this translates to a supported voice traffic load of about 62 Erlang, and hence $\lambda_D \approx 0.52$ calls/s.

The comparison of the three scheduling schemes is based on performance parameters such as mean file transfer time for zone i and fairness. The fairness is evaluated with the fairness index used by Jain, see [33]:

$$F = \frac{(\sum_{i=1}^{K} T_i)^2}{K \sum_{i=1}^{K} (T_i)^2}, \tag{3.17}$$

where T_i denotes the mean flow transfer time for users in zone i, $i = 1, ..., K$.

3.6.2 Cell modelling with zone division

The cell division into zones is an essential decision since it allows us to differentiate between MSs' positions and feasible data rates in order to model a realistic scenario. Just as important it is to decide on a representative number of zones K. We have performed experiments with 5, 10, 15, 20 and 40 zones and two different loads - medium and close to saturation - to find an optimal number of zones. A large number of zones supports finer differentiation in the MS's location but also leads to increased modelling complexity and simulation time. Depending on the research goals, a trade-off between granularity and complexity of modelling is necessary.

In order to find an appropriate modelling choice, we have compared the relative difference in mean flow transfer time when we increase the number of zones. We prefer relative to absolute difference since it gives the improvement registered by each increase in K. An appropriate number of zones is found when further increase in K leads to a negligible relative difference in the mean flow transfer times.

The general trend is that increasing the number of zones leads to decrease in the relative difference. The initial improvement of 41.45% when changing from 5 to 10 zones reduces to only 3.15% when changing from 20 to 40 zones. We believe that further increase in the number of zones would not have a considerable contribution. The results show that 20 zones is an optimal choice for a wide range of loads - it provides satisfying granularity while still keeping simulation time manageable. We have chosen to work with 10 zones - the modelling and simulation efforts are considerably less and granularity level is satisfying unless very high loads are used.

3.6.3 Simulation of flow-level Markov chain

A Markov model with simple-form transition rates can be analytically described by explicit mathematical expressions and a steady-state distribution. More complex models are challenging for a purely analytical approach, in which cases simulation of the Markov model can be a solution. In order to obtain a steady-state distribution we created a generic simulator for multi-dimensional Markov chains by using MATLAB (`http://www.mathworks.com/`). We actually simulate state transitions taking into account the transition rates from the flow-level Markov model, cf. [4]. In particular, given the currently visited state $\underline{n}$, we calculate the total rate out of the state, denoted by $Q(\underline{n})$, as:

$$Q(\underline{n}) = \sum_{i=1}^{K} \lambda_i + \sum_{i=1}^{K} R_i(\underline{n}) + \lambda_D + R_D(n_D),$$

where $R_D(n_D) = \frac{n_D}{\tau}$. Subsequently, the transition probabilities are calculated as:

$$\begin{aligned} Pr\{\text{forward jump in dimension i}\} &= \lambda_i/Q(\underline{n}), i = 1, \ldots K, \\ Pr\{\text{backward jump in dimension i}\} &= R_i(\underline{n})/Q(\underline{n}), i = 1, \ldots K, \\ Pr\{\text{forward jump in dimension i}\} &= \lambda_D/Q(\underline{n}), i = V, \\ Pr\{\text{backward jump in dimension i}\} &= R_D(n_D)/Q(\underline{n}), i = V. \end{aligned}$$

An uniform sample is used to determine which transition takes place. Before moving to the newly chosen state we record the time spent in the current state, taken as a sample from an exponential distribution with mean $1/Q(\underline{n})$.

By applying this iterative process for one million state transitions and collecting data on the total time spent in each visited state, we derive the steady-state distribution. Experiments showed that working with one million state transitions is sufficient to generate trustworthy results, i.e. 95% confidence intervals of about 3% of the obtained mean.

The combined approach of mathematical analysis and Markov model simulation allows fast evaluation of the required performance measures. For example, the models considered in the present chapter require less than five minutes simulation time for a single graph (based on one million simulated state transitions). More complex models, e.g. six cell system with accounting for inter-cell interference, see e.g. Chapter 4, stretch up to an hour. All simulations are performed in a MATLAB environment under MS WindowsXP Professional operating system on a Compaq nc6320 machine with Intel(r) Core(TM)2 T5600 processor of 1.83 GHz and 1.99 GB RAM.

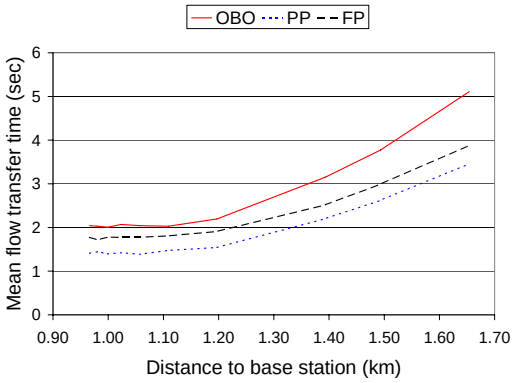


Figure 3.5: Impact of the mobile station's location, i.e. distance to the base station, on the mean flow transfer times.

3.7 Numerical results

This section presents all performance evaluation scenarios we have examined. All results are based on the default settings of Section 3.6.1. Sections 3.7.1 to 3.7.4 aim to observe the impact of various factors on the mean flow transfer times of EUL data flow. In Section 3.7.5 we compare the scheduling schemes in the delivered fairness and finally the performance differences between several implementations of a full parallel scheduler are examined in Section 3.7.6.

3.7.1 Impact of MS location

Figure 3.5 shows, for each of the three schedulers, the mean flow transfer time as a function of the MS's distance from the base station. The default system and traffic parameters from Section 3.6.1 apply. As expected, the partial parallel scheduling scheme outperforms the other two schemes, the full parallel scheme performs second best and the one-by-one scheme shows the worst performance. For all three schedulers, when moving away from the base station, the mean flow transfer times remain more or less constant until a distance of about 1.1 km. Apparently, MSs in these central zones have received powers such that they can fill up the total data budget B_D and hence realise the same, maximally possible data rate. At larger distances the effect of the increasing path loss,

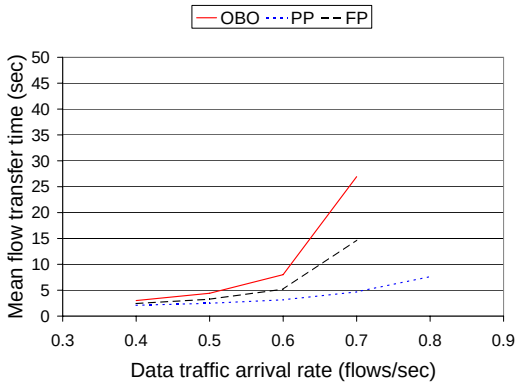


Figure 3.6: Impact of the EUL arrival rate λ on mean flow transfer times.

consequently lower attainable received powers and hence lower bit rates becomes clearly visible in the rapidly increasing mean flow transfer times.

Note, that even for those MSs close to the base station, for which $P^{rx} = B_D$ holds, the PP scheme performs considerably better than the OBO (and FP) scheme. In OBO MSs at the cell edge cannot fill the budget B_D on their own and thus waste channel resources compared to PP. As the channel access is fairly shared among all users in the cell such MSs also have a disadvantageous effect on the throughputs obtained by the MSs at the centre of the cell.

3.7.2 Impact of effective EUL load

In Figures 3.6 and 3.7 the effective EUL load is varied in two distinct ways. In Figure 3.6, the aggregate EUL flow arrival rate λ is varied directly. In Figure 3.7, given a fixed EUL flow arrival rate, the available capacity for EUL transfers is varied, which effectively corresponds to an inverse variation of the EUL load. As we will see, the key difference between these distinct approaches is visible in the relative performance of the OBO and FP schedulers.

In Figure 3.6 the mean flow transfer time (appropriately averaged over all zones in the cell) is given as a function of the EUL flow arrival rate λ. All other default parameters hold. The performance difference between the schedulers observed in Figure 3.5 becomes even more pronounced when λ increases. In particular, the mean flow transfer time for the OBO (and also FP) scheme

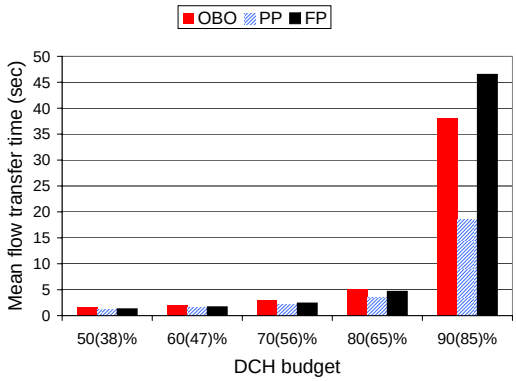


Figure 3.7: Impact of the voice budget B_V, and implicitly on the available EUL data budget B_D, on mean flow transfer times.

increases very rapidly when λ becomes larger than 0.6 flow initiations/sec, while the system becomes saturated for λ's between 0.7 and 0.8. The growth of the mean flow transfer time under PP scheduling however remains moderate.

It is particularly interesting to consider how the performance gain of PP over the OBO and FP schedulers depends on the available EUL budget. It is expected that when the available EUL budget is small enough to be filled up by a single user, OBO is more efficient than FP, since it yields lower intra-cell interference, higher $SINR$ ratios and hence higher bit rates. In order to investigate this we varied the voice budget B_V between 50 and 90%, which leads to decrease in the available EUL budget B_D and hence increase in the EUL load. For each distinctive value of B_V we determined the voice traffic arrival rate λ_V such that ensure a blocking probability of 1%. In Figure 3.7 the x-axis indicates the reserved voice budget B_V followed, between brackets, by the (average of the) actually used budget by the voice calls.

Figure 3.7 plots, for each of the schedulers, the resulting mean flow transfer times of the EUL flows versus the voice budget B_V. For all three schedulers, the general trend is that decreasing the EUL budget leads to increase in mean flow transfer times. For small values of the voice budget the performance of PP and FP (and to a lesser extent also OBO) is quite similar. In such situation there is sufficient data budget such that the cycle length is often one TTI and thus the schemes (PP and FP) handle flows effectively in the same way. For

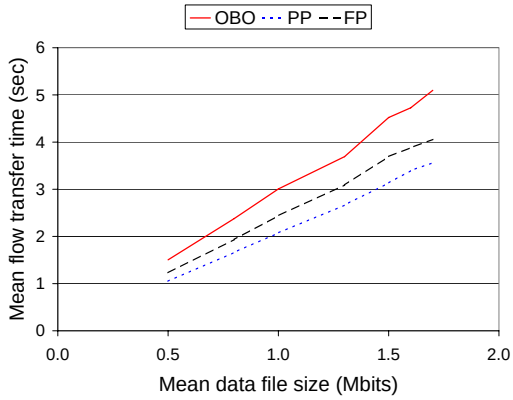


Figure 3.8: Impact of EUL flow size on mean flow transfer times.

higher (voice) loads the performance gain of PP (compared to FP and OBO) increases, while for the highest considered (voice) load, the OBO scheme indeed performs better than FP.

In conclusion, the PP scheduling scheme performs always better than (or at least as good as) the OBO and FP schemes. The relative performance of OBO and FP depends on particular load of voice traffic - with low load FP is better, with high load - OBO is better.

In the course of simulations we also observed a particularly interesting tendency - with increasing system arrival rate OBO is the first scheme that will reach system instability, followed by FP and PP. This observation leads to two very important conclusions. First, since the three schedulers behave as three different system models there is not a single arrival rate (load), beyond which the system becomes unstable, applicable to all three. Second, the relatively inefficient traffic handling in the OBO and FP schemes apparently leads to a considerable reduction of the cell capacity compared to the PP scheme.

3.7.3 Impact of EUL flow size

We have observed the schedulers behaviour for a range of mean flow sizes - from 0.5 to 1.7 Mbits. Figure 3.8 presents the results for the three schedulers. Increasing the flow size leads to an expected increase in the mean flow transfer time due to the bigger amount of data that has to pass through the system. OBO,

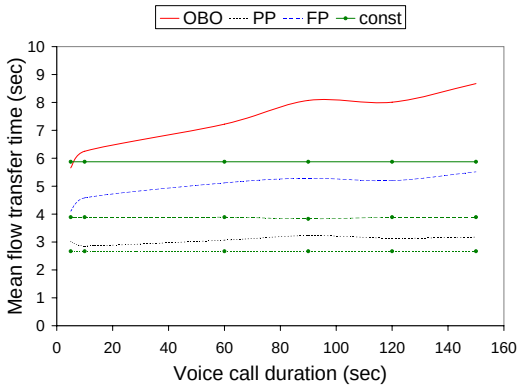


Figure 3.9: Impact of voice calls durations on mean data flow transfer times.

as the most inefficient scheme in terms of budget utilisation, shows fast degradation. Its curve is the steepest and displays fast tendency towards infinity, which is an indication of the system approaching instability. The same process but to a lesser degree can be observed for the FP scheme. Once more the optimal assignment of TTIs and transmit powers places PP as the best performing scheme.

3.7.4 Impact of voice call duration

Theory on processor sharing queuing systems shows that a constant service rate C is more beneficial for system performance than a varying service rate with mean C, see e.g. [49]. In particular, slow variations have a strong negative effect on performance. Very fast variations lead to system performance close to the value realised under constant rate. Hence, we could expect that mean data flow transfer times increase when the duration of the voice calls increases.

In order to examine the impact of the duration of voice calls we compared EUL performance in terms of mean flow transfer times aggregated over the whole cell, for two scenarios. In the first scenario we set the number of voice calls fixed (infinitely long calls) such that 70% of the total budget is used by the voice calls. In the second scenario we considered finite voice calls occupying on average 70% of the total budget. Voice calls have data rate of 12.2 kbit/s; we set the call duration at (5, 10, 20, 40, 60, 80, 100, 120, 150) seconds and chose the

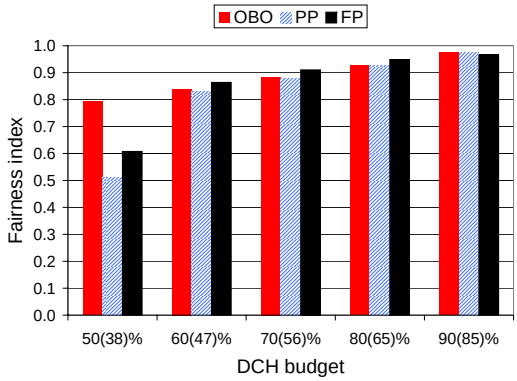


Figure 3.10: Fairness index.

voice call arrival rates such that the 70% budget occupation was realised. For each of these settings we ran extensive simulations of one million state changes, i.e. flow initiations or completions, of voice calls.

Figure 3.9 shows the results for each of the three schedulers with the two examined scenarios. Indeed, as we expected, the graphs for constant and variable service rate lie close to each other for short voice calls and further apart for long voice calls. Note that in the case of constant service rate the realised mean flow transfer time, which is in fact a single values, is represented as a line in order to ease comparison with the variable service rate over the examined range.

The general trend in performance is the same for all three scheduling schemes but the specific impact of the voice call duration on each scheme is different. Recall that OBO and FP have longer mean flow transfer times than PP, therefore long periods of low service rate. Hence, long voice calls have stronger negative influence on a OBO and FP scheme than on PP scheme.

3.7.5 Fairness issues

Beside in mean flow transfer times we are particularly interested in the fairness the three schedulers show with respect to the service they offer to MSs at different locations in the cell. The fairness can be defined in different ways. We have used the fairness index applied by e.g. Jain [33], which, in the present context,

is defined in Equation (3.17). The maximum value of the fairness index equals one, which refers to a perfectly fair scenario in which the mean flow transfer times are the same for all zones. The smaller the fairness index the larger the (relative) differences among the transfer times in the different zones.

Figure 3.10 shows the fairness results for the three schedulers. Again in the setting of changing voice budget, see Section 3.7.2, the fairness index is calculated. The general impression is that the fairness performance is more or less the same for the three schemes. However, for the case that the voice budget/load is small (i.e. the available EUL data budget is relatively large) the OBO scheme appears to be significantly more fair than the two other schemes, particularly than the PP scheme. This can be explained as follows. Under PP scheduling, MSs near the base station (with a high maximum received power) are more likely to be served alone compared to MSs at the cell edge, which are mostly served in parallel with others (because they are unable to utilise the available EUL budget on their own). Consequently, remote MSs experience lower $SINR$ ratios and hence lower bit rates explaining the relatively large differences in mean flow transfer times of MSs close to the base station and users at the cell edge.

Comparing OBO and FP, the former provides greater degree of fairness. The reason leis in the intra-cell interference that both schemes deliver. In OBO MSs are scheduled in a one by one fashion eliminating interference from other EUL users; in FP such interference is present due to parallel transmissions. Certain unfairness originates in the fact that MSs have different received powers (due to location-specific path loss) and stations with low received powers are more affected by the interference than stations with high received powers. A more fair situation would be where each MSs experience interference comparable to its own received power level.

When the voice budget B_V increases (i.e. available EUL budget decreases) the fairness of the schedulers equalises. This is due to the fact that when the available EUL budget becomes smaller, the maximum (received) power that can be achieved by the MSs at different locations is less and less a limiting factor for remote MSs, and hence the disadvantageous effect that such MSs suffer from added intra-cell interference vanishes.

3.7.6 Impact of FP implementation

In order to evaluate the importance of the exact implementation of a FP scheduler we observed the mean flow transfer times registered by MSs from different

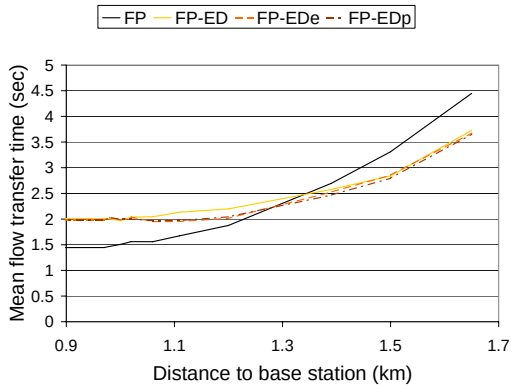


Figure 3.11: Performance of implementations of a FP scheme.

locations in the cell with default scenario settings for each of the four FP implementations described in Section 3.2.

The results, presented in Figure 3.11, show that there is a negligible difference between the FP implementations which apply equal budget division. A detailed study of the results, i.e. TTI based evaluation, revealed that the effective unused EUL budget after the first division is in fact negligibly small. Since these three implementations differ exactly in the management of the unused budget, we can justify the similarities in performance.

Furthermore, Figure 3.11 shows that the graph of the initial FP with proportional division is steeper than the other three which can be explained with the different fairness of the schemes. FP with proportional power division favours MSs located closer to the base station while the equal division schemes try to serve all MSs equally, i.e. they are more fair. Therefore remote MSs under FP with equal division improve in performance but on the cost of close by MSs.

The choice of implementation is defined by the goal of the scheme - providing the best possible service or keeping fairness among the MSs. Interestingly, even FP with proportional division, which delivers the poorest fairness among all FP implementations, is still more fair than the PP scheme.

We have also evaluated the fairness index for each of the FP implementations. The results for all four FP implementations for a range of EUL traffic arrival rates are plotted in Figure 3.12. The reserved voice budget is fixed at 70%. What first draws the attention is the clear disadvantage of the proportional power

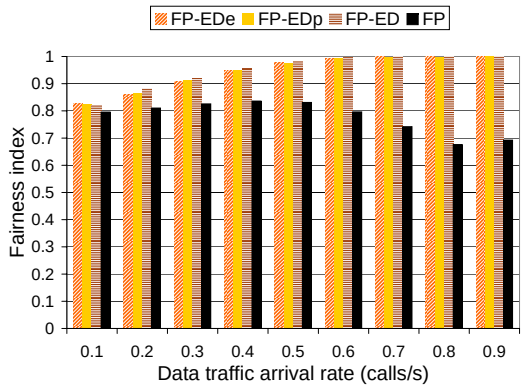


Figure 3.12: Fairness index for FP implementations.

division approach, i.e. FP, in terms of fairness. All schemes with equal power division are more fair since they assign to each MSs, independently of location, an equal part of the data budget. Further differences in the performance of these schemes is not observed. As the load of EUL traffic increases so does the average number of ongoing EUL flows. Consequently, situations of budget under-utilisation occur less often and hence FP-EDe and FP-EDp converge de facto to a FP-ED schemes when all MSs are able to fill up their assigned portions, explaining the increasing fairness. The FP schemes show degradation in the fairness index since higher load implies bigger power cuts of which remote MSs suffer more than close by users.

3.8 Conclusions

In this chapter we focused on packet scheduling for serving data traffic on the enhanced uplink for UMTS. In particular, we presented three scheduling schemes and evaluated their performance for a diversity of traffic settings. The schemes differ in the management of parallel transmissions, namely service a single, multiple or all ongoing flows in a scheduling interval.

The presented study concentrates on a single cell scenario with different coexisting traffic types - 'plain' UMTS voice calls and EUL data flows. We were able to examine how these traffic types influence each other's performance and derived several practical recommendations for deployment. In the following

Chapter 4 we will concentrate on the interaction of multiple cells and the impacts of it on performance.

Performance evaluation was achieved by developing a methodology which captures both packet level details and flow-level dynamics. An important part of the methodology is the application of Markov chains to model flow-level behaviour, i.e. the initiation and completion of flow transfers. Our modelling and analysis approach supports fast evaluation and easy modification of the system details. The main performance measures in our study are mean flow transfer time and fairness (i.e. the ratio of the mean flow transfer times for MSs at different locations in the cell).

To summarise our findings: (i) the partial parallel scheduling strategy, which combines several transmissions in a TTI but preserve the maximum achievable received powers, performs best; (ii) keeping a small resource reservation for EUL data traffic can significantly improve its performance without seriously degrading the service offered to 'plain' UMTS voice traffic; (iii) the proposed scheduling schemes do not exhibit significant differences in fairness; and (iv) implementation details of a scheduler, e.g. regarding the composition of flows for parallel transmissions, do not considerably affect the performance at flow level and can be therefore considered irrelevant.

4 Scheduler-dependent inter-cell interference modelling

4.1 Introduction

In the previous Chapter 3 we analysed and discussed the performance of various packet schedulers in a single cell scenario. It was argued that flow-level dynamics have significant impact on performance and should be taken into account during evaluation. We expect this argument to hold even stronger for a scenario with multiple cells.

It is our intention in this chapter to investigate the mutual influence neighbour cells have on each other through the generated inter-cell interference and how this depends on the applied scheduling scheme. Again we pay special attention to the importance of flow-level dynamics. The same basic scheduling schemes as introduced in Chapter 3 are considered, namely one-by-one, full parallel and partial parallel.

Since inter-cell interference is a central research topic in this chapter, we only consider elastic data traffic from EUL users. Working with a single traffic class allows us to focus on a single issue - inter-cell interference in this case - and to identify the specific cause of a particular behaviour. We have undertaken a modular approach to our investigations. In particular, first we consider a simplified network model of two neighbour cells from which we can gain significant insights on the inter-cell interference process. Second, this knowledge is used to develop an approach to evaluate a more complex multi-cell network scenario.

Independently of the complexity of the discussed scenario the same combined packet and flow-level analysis, which was introduced in Section 3.4, is applied. Certain modification of the approach were required to allow the correct representation of inter-cell interference and its effects on performance. These modifications are discussed in more detail in Section 4.4.4.

4.1.1 Literature

Inter-cell interference is a naturally occurring phenomenon in cellular systems with common radio resource such as UMTS/HSPA and as such its impact on performance and system planning has been widely researched with few examples being [27, 40, 63]. However, we are particularly interested in considering the inter-cell interference process in combination with the particular scheduling scheme (applied on the uplink in our case). Furthermore, flow dynamics, i.e. the constantly changing number of users, should be included in the analysis since we expect it to significantly affect performance.

Studies that provide analysis of scheduling schemes considering both environmental factors and flow dynamics are not easy to find. One such study that is similar to ours is [43]. It tries to capture the impact of inter-cell interference for a multi-cell system. In particular flow-level performance metrics are analysed for two (rate-fair) scheduling disciplines assuming that the transmit powers of all mobiles are sufficient to reach the maximum bit rate, such that (in contrast to our present study) the users' performance does not depend on their location in the cell. In a later research, e.g. [41, 42], the same authors consider a multi-cell scenario and choose for a log-normal approximation of the inter-cell interference process. However, this model does not capture the impact of the particular scheduling scheme used in the network.

4.1.2 Contributions

The current chapter presents a study on the impact of inter-cell interference on EUL data traffic performance. The analysis includes modelling approaches, system representation issues and performance evaluation. The eventual goal is to identify and quantify the dependency of the inter-cell interference on the applied scheduling scheme, and to investigate its impact on the users' performance. In order to accomplish this goal we have applied basically the same hybrid analytical/simulation approach as used in Chapter 3 capturing both packet-level details and flow-level dynamics.

The central contribution of the chapter is the evaluation of the combined effects that packet scheduling and inter-cell interference have on mobile stations performance. The other significant contributions are:

- Evaluating how the inter-cell interference process, generated by a single cell, depends on the applied packet scheduler.
- Determining what is a good modelling strategy to capture the mutual effect that neighbour cells have on each other in terms of transmit power assignment and hence inter-cell interference levels. We show that this mutual effect can be well captured by one or two iterations of the transmitted power recalculation.
- Evaluating the impact of the resulting, scheduling-specific inter-cell interference processes and power assignments (see above) on the users' performance in terms of e.g. flow transfer times.

4.1.3 Outline

In the rest of the chapter we first address, in Section 4.2, the scheduling schemes and then, in Section 4.3, the considered system model(s). Later in Section 4.4 the analytical model, which is used to evaluate performance, is explained. Subsequently, the discussion of the numerical results is presented in Section 4.5 and Section 4.6 for the single neighbour cell and the multi-cell scenario respectively. Finally, Section 4.7 shortly summarises the presented research.

4.2 Considered scheduling schemes

In Chapter 3 we proposed and evaluated the performance of three packet schedulers which provide equal opportunity to users to use the EUL channel, namely one-by-one (OBO), full parallel (FP) and partial parallel (PP). To shortly summarise the schemes - OBO serves a single user at a time; FP schedules all active MSs simultaneously; and the PP scheme serves several but not necessary all MSs in the same TTI. For more detailed description of the schedulers and their transmit power assignment policy we refer to Section 3.2.

Figure 4.1 shows how each scheme serves the active users over time as well as the budget consumption (by users and interference). Each scheduler differs in the policy of assigning transmitted power and in the approach towards parallel transmissions. Hence, the inter-cell interference generated towards neighbour

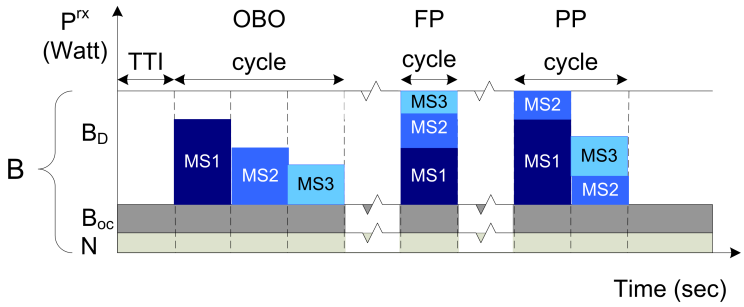


Figure 4.1: Scheduling schemes under study.

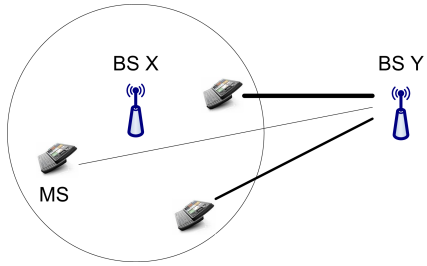


Figure 4.2: Effects of the mobile station location on inter-cell interference.

cells is also scheduler-specific. Furthermore, it dynamically changes in time as the number of active users varies. To reflect the impact of these changes on the system an adaptive reservation B_{oc} is introduced as part of the total available channel resource B that is reserved to cope with inter-cell interference. Consequently, after the interference is subtracted from the total budget B, the channel resource available for EUL transmissions B_D effectively decreases.

Besides being scheduler-specific the inter-cell interference is also implementation specific. In Chapter 3 we showed that the particular implementation, i.e. the order in which MSs are served, does not influence (significantly) the performance within the cell. This however is not the case with inter-cell interference. If several MSs are scheduled together their distances to the base station of a neighbour cell can significantly differ. This is illustrated in Figure 4.2 - a MS which is relatively close to base station Y generates a much higher inter-cell interference than a MS far from it. Note that in the FP since all users are served

together for a given number of users the interference is robust to the particular user ordering.

In order to investigate the effects of the particular implementation strategy we consider two extreme implementations of the PP scheduler. The implementations are chosen such that to provide us with upper and lower bounds on the performance. In both implementations first all users are ranked according to their induced inter-cell interference levels (from high to low). The *'worst'* PP scheme, denoted PP-w, selects the users to serve in a TTI following the ranking. Therefore, users with similar power levels often are served in the same TTI. This is particularly damaging when the users are located near the cell edge of cell X close to cell Y.

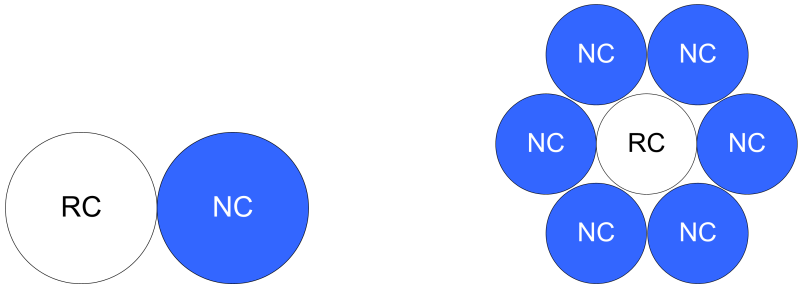
The *'optimal'* PP schemes, denoted PP-o, applies a 'smarter' approach trying to level the total interference in a TTI. It achieves that by selecting MSs in alternating order of their induced inter-cell interference levels, e.g. highest, lowest, second-highest, etc.

The effect of the scheduler implementation is examined only for modelling scenario I with a single neighbour cell. We will show later, in Section 4.5, that eventually the flow-level performance is not considerably influenced by the implementation. Therefore, modelling scenario II includes a single PP implementation.

4.3 Model

The evaluation of the effects of inter-cell interference requires the modelling of a multi-cell system. Such models however may be very complex and require long evaluation times. High model complexity also makes it often difficult to identify the cause of a particular behaviour due to multiple contributing factors. We have chosen to apply a modular approach. First, we analyse a basic modelling scenario, termed scenario I, which consists of only two cells - a reference cell, where performance is evaluated, and a neighbour cell, where inter-cell interference originates. Subsequently, a more realistic scenario of six neighbour cells, termed scenario II, is evaluated. Scenario I allows us to more easily detect performance trends and to identify their origin, while scenario II provides us more realistic performance figures.

In order to examine in sufficient detail the inter-cell interference we need to apply a proper modelling of the cell. The modelling should account for the effect that distance has on inter-cell interference levels and hence should allow us to differentiate in user's position. This is relevant for the modelling of a neighbour



(a) Scenario I (b) Scenario II

Figure 4.3: System modelling scenarios.

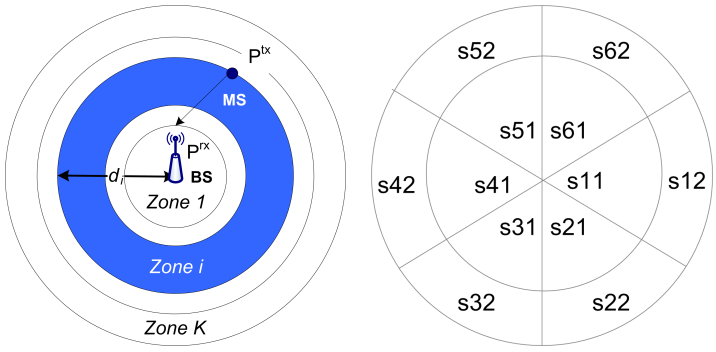
cell. For the modelling of the reference cell we apply the same single cell model, which we described in Section 3.3.

4.3.1 System Models

Scenario I discusses a system of one reference cell (RC) and one neighbour cell (NC), see Figure 4.3(a). *Scenario II* considers a system of seven cells - a reference cell and six neighbour cells - as shown in Figure 4.3(b). A second tier of cells is not considered since we focus on the performance in the reference cell. The distance between the base stations of neighbouring cells is double the cell radius. Each cell is served by a base station with an omnidirectional antenna.

In both scenarios for the reference cell we apply the single cell model from Section 3.3; the cell in divided into K concentric *zones* with equal areas, which allows to account for the effect of user's location. Zone i is characterized by a distance d_i to the base station and corresponding path loss denoted by $L(d_i)$, $i = 1, ..., K$, e.g. Figure 4.4(a).

In order to account properly for inter-cell interference, the basic single cell model had to be extended such that to include finer granularity of user's location. Hence, a neighbour cell is additionally split up in S *sectors* also of equal size. The intersection of zones and sectors in the neighbour cell determines equally sized *segments* characterized by a distance d_{ij} and corresponding path loss $L(d_{ij})$ to the base station of the reference cell where $i = 1, ..., K$, $j = 1, ..., S$. Figure 4.3(b) shows an example for $K = 2$ and $S = 6$.



(a) Zones (b) Segments

Figure 4.4: Cell modelling of the reference and neighbour cell.

4.3.2 Assumption

We distinguish between assumptions at system level and assumption at user level. The assumptions made for the reference cell are the same as the ones presented in Section 3.3. Additional assumptions are made to account for the inter-cell interference. The same holds for the assumptions made about the users.

System level

Similar to Section 3.3, the total received interference budget B at the base station of a RC or a NC can be derived from an operator-specified noise rise target η as $B = \eta N$. However, in order to cope with anticipated inter-cell interference I_{oc} we introduce a scheduler-specific reservation B_{oc}. Since inter-cell interference varies in time, the reservation B_{oc} is a dynamically changing variable, see Section 4.4.2. In particular, we assume that B_{oc} is adapted instantaneously at state changes of the neighbour cell. Note that B_{oc} is a modeling parameter to represent the real measure I_{oc}. The specific choice of the value of B_{oc} will be discussed later on.

In our model, the total interference in the cell comprises several distinct components: *(i)* the thermal noise level N; *(ii)* the self-interference modelled by parameter ω, which is due to the effects of multi-path fading; *(iii)* the interference I_D originating from simultaneously served EUL uses in the reference cell; and *(iv)* the inter-cell interference I_{oc} originating from EUL flow transmissions

in the neighbour cell(s). Note that ongoing 'plain' UMTS traffic is not taken into account. Eventually, the data budget B_D available to EUL data traffic can be derived from the total budget B as $B_D = B - N - B_{oc}$.

Each cell can be uniquely described, in a particular time instant, by *the system state* $\underline{n}$. In the reference cell the state represents the number of active users and their distribution over the cell zones, $\underline{n} \equiv (n_1, n_1, \cdots, n_K)$, where n_i is the number of users in zone i, $i = 1 \cdots K$. Due to the additional division of the NC in sectors the state describes the population of users in the segments, i.e. $\underline{n} \equiv (n_{11}, n_{12}, \cdots, n_{21}, n_{22}, \cdots, n_{KS})$, where n_{ij} represents the number of active flows in a segment ij, $j = 1 \cdots S$.

User level

EUL data flows are generated according to a spatially uniform Poisson arrival process with rate λ. The flow size is exponentially distributed with mean F (in kbits). As a direct consequence of the uniformity assumption and the choice of equally sized zones, the probability that a data flow appears in zone i, is $1/K$ and equal for all zones, so that the EUL flow arrival rate per zone λ_i is equal to λ/K, $i = 1, \cdots, K$. These are the same assumptions made also in Section 3.3. The arrival rate per segment in the NC however can be calculated as $\lambda_{ij} = \lambda/KS$, where the total arrival rate in all segments belonging to the same zone is $\lambda_i = \sum_{k=1}^{S} \lambda_{ik}$.

All mobile stations have the same maximum transmitted power $P_{\max}^{tx}$ but different corresponding received power at the base station $P_{i,\max}^{rx}$ due to the zone-dependent path loss. No user mobility is considered.

Neighbour cell

It is assumed that a neighbour cell behaves independently of the reference cell. In particular, from the point of view of the reference cell, the inter-cell interference generated by the neighbour cell is an autonomous process determined by the evolution of the number of active users in each of the segments of the neighbour cell and the applied scheduling scheme. Later on in Section 4.4.5 we will show the base for this assumption and provide results supporting its correctness.

4.4 Analysis

For each of the scheduling schemes the analysis starts with determining the inter-cell interference caused by the neighbour cell, see Section 4.4.2. Next, the flow-level behaviour of the reference cell is analysed while taking into account the inter-cell interference from the neighbour cell, see Sections 4.4.3 and 4.4.4. During the analysis we need to consider the mutual influence that neighbour cells have on each other in terms of power allocation and how this interaction is best modelled. Our investigations on the issue are presented in Section 4.4.5.

Both steps of the analysis are based on the flow-level modelling and analysis approach for a single cell scenario, which we discussed in Section 3.4. We will now start (in Section 4.4.1) with briefly summarising the main points of the analysis, while for full details we refer back to Chapter 3.

4.4.1 Preliminaries - single cell analysis

The analysis of a single cell begins at packet level with determining the *instantaneous rate* and continues to include the impact of flow dynamics by first accounting for the number of active users (captured by the state-dependent throughput) and later for the change in this number (captured by Markov models).

The instantaneous rate $r_i(\underline{n})$ of a scheduled user in zone i, see also Section 3.4.2, captures the effects of factors such as interference (both inside and outside the cell) and user's position (reflected by location-specific path loss). In contrast to the scenario discussed in Chapter 3 in this chapter we take into account inter-cell interference but abstract from the presence of UMTS voice calls. Therefore Equation (3.4) takes the modified form of:

$$r_i(\underline{n}) = \frac{r_{chip}}{E_b/N_0} \cdot \frac{P_i^{rx}}{I_D(\underline{n}) - \omega P_i^{rx} + B_{oc} + N}, \tag{4.1}$$

where the index i, $i = 1, ..., K$, indicates the zone to which a user belongs. Note the new term B_{oc} which captures the impact of inter-cell interference and as we will discuss later is a dynamically changing parameter. The derivation of the received power P_i^{rx} is presented in Section 3.4.1, Equation 3.1.

In order to account for the number of active users and their effect on how often a user realises its instantaneous rate, we introduced in Section 3.4.3 the state-dependent throughput $R_i(\underline{n})$. It captures the scheduler-specific manage-

ment in the time domain by considering the cycle length $c(\underline{n})$:

$$R_i(\underline{n}) = \frac{r_i(\underline{n})}{c(\underline{n})}. \tag{4.2}$$

Similar to Chapter 3 we can apply Markov chains to model the change in number of users. A state in the chain corresponds to a particular user population and its distribution over the cell. The transition rates are derived from the arrival rate λ of the Poisson process and the state-dependent throughputs, see Expressions (3.6).

Following, the steps sketched above the behaviour of the reference cell can be described by a K-dimensional continuous time Markov chain (CTMC). In a similar way, a neighbour cell, due to the more detailed division in segments is represented by a $K \times S$ dimensional CTMC. Eventually, depending on the discussed scenario, we construct a complex multi-dimensional Markov model describing the behaviour of both the reference and the neighbour cell(s), see Section 4.4.4. Simulation of the Markov chain in order to find its steady-state distribution is the more time-efficient option than trying to find a closed-form solution. How a Markov chain can be simulated in a fast, computationally efficient way is described in Section 3.6.3.

4.4.2 Inter-cell interference process

The inter-cell interference I_{oc} generated by a NC at the base station of the RC is a stochastic process. In order to model it correctly we need knowledge on the interference values and the probability with which they exhibit. The possible I_{oc} values depend on in which segments of the NC the active MSs appear. Given a distance d_{ij} and a corresponding path loss $L(d_{ij})$, we can calculate the inter-cell interference generated by a user in segment ij in the NC. It is in fact the received power of that user at the BS in the RC, which we derive by rewriting Equation (3.1):

$$I_{oc,ij} = \frac{P^{tx}}{L(d_{ij})}. \tag{4.3}$$

Note that the inter-cell interference changes at two time scales - on TTI level and on flow-level. Hence its value changes from TTI to TTI but also at state transition, i.e. flow arrival or departure.

For a given state $\underline{n}$ of the neighbour cell the inter-cell interference generated during each TTI of a scheduling cycle (for the duration of that state) is fully determined. As an example, Figure 4.5 shows, for one particular system state,

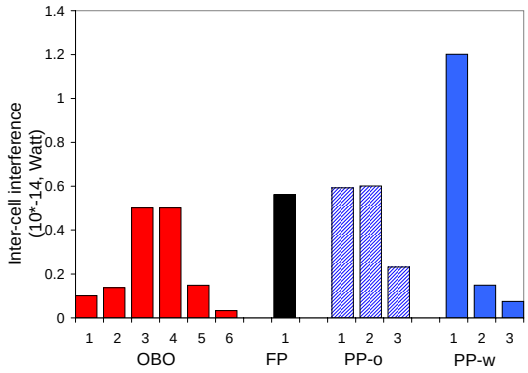


Figure 4.5: Inter-cell interference process at the TTI level - example for six users.

the interference from each of the schedulers, see Section 4.2, generated during the TTIs of the transmission cycle. For the FP scheme the cycle length equals one, and hence the inter-cell interference for a given system state is a repeating single value. For the OBO, PP-o and PP-w schemes, the inter-cell interference may vary within a cycle, depending on the locations of the currently served user(s). Note that for a particular system state, the maximum level generated by the PP-w scheme within a cycle is the highest registered by all schemes. On the one hand, PP-w serves the users with the highest induced inter-cell interference together. Hence, due to the single user policy, the OBO scheme can never generate higher interference level than the accumulated effect of the PP-w. On the other hand, PP-o is already designed to avoid the high interference levels that the PP-w can cause.

The probability of each I_{oc} value can be statistically determined from simulation by applying straightforward calculations. For each system state we record the interference levels caused by the currently active users. Additionally, also the time the system spent in the particular state is known. Hence, accounting for the cycle length, we can calculate how long each exhibited value has been present. Dividing by the total simulation period, it is easy to derive the probability that these same values appeared in the system.

4.4.3 Accounting for inter-cell interference

As we showed in the previous Section 4.4.2 the inter-cell interference varies over time. Therefore the choice of the reservation B_{oc} to cope with it is crucial for the performance of mobile stations. In Section 4.3 B_{oc} was introduced as a dynamically changing reservation, which is adapting to the changes in the inter-cell interference process. Recall that the interference process changes at both TTI and flow level. Capturing the changes at flow level is straightforward while capturing them at the TTI level requires much more computational time.

To cope with the variation of the interference at TTI level in our model a single value is selected being representative for the whole transmission cycle. This value is taken for the computation of the reservation level B_{oc} during a given state. Depending on the particular scenario, i.e. scenario I or B, the representative value is selected differently, which is discussed below. At flow level, at any state change in a neighbour cell, the representative value needs to be recalculated, and hence the reservation B_{oc} to be adapted.

In order to select an appropriate representative value we need knowledge on the inter-cell interference generated during each TTI. In the model we assume this knowledge to be instantaneously available. Although a rather optimistic assumption, it provides a lot of insight on the inter-cell interference process and its impact on performance.

In practice it is challenging to receive such precise information and a more realistic approach would be necessary. For example, if we select to work with the maximum realised interference, the base station can periodically measure the received interference over a pre-defined period of time and compare the maximum value with the already registered value from a previous measurement. If the new value is higher the base station should update the reservation B_{oc}. Depending on the chosen approach to select a representative value the described procedure might differ slightly.

Below we will describe in more detail how the inter-cell interference variation is dealt with in each of the discussed scenarios.

Scenario I

Knowing the state in the neighbour cell, we can determine the inter-cell interference distribution over the transmission cycle. Keeping that in mind, there are multiple possibilities to select the size of the reservation B_{oc}, for the duration of the particular state, depending on the desired performance.

Choosing a reservation equal to the maximum inter-cell interference genera-

ted in the scheduling cycle (i.e. the 100% percentile) is the safest approach - the total power budget will never be exceeded. However, a significant part of the budget might then (unnecessarily) not be accessible for the mobile stations. Consequently, mean flow transfer times increase, while part of the budget is unused during TTIs with lower than the maximum interference levels. On the contrary, making no reservation at all (i.e. the 0% percentile) does not affect the power budget available for EUL data traffic in the reference cell, but it increases the risk that the total budget B is exceeded. This could cause degradation of the offered quality of service (QoS) via increased block error rates. A compromise option is to set the reservation at a particular percentile value between 0% and 100% of the inter-cell interference generated by the neighbour cell.

Since each scheduling scheme has its specific power and TTI assignment we could also expect that even if selecting the same percentile value for B_{oc} the actual reservation of the budget will be different.

Scenario II

In principle the performance in the reference cell is influenced by each state change in any of the neighbour cells. With the complex network of six cells in Scenario II this makes both finding a percentile of the I_{oc} distribution and evaluating the whole cell system a very time consuming and difficult to manage task. To avoid complications we have decided to simulate in full detail only one neighbour cell (*full-detail NC*) while the other neighbour cells are represented by an average interference value (*abstract NC*). For the full-detail NC two approaches to set B_{oc} are chosen. In the first case, we use the average inter-cell interference generated during a particular state (over the corresponding transmission cycle). As a second option B_{oc} is set at the maximum realised value for the same particular state (transmission cycle).

Taking an average value brings the risk that the real interference might be often higher than the reservation B_{oc}, which potentially can lead to exceeding the budget. Using the maximum value avoids the problem but it also implies that there will be over-reservation. This is especially true for users located at the edge of the NC close to the RC, which generally have long service times and generate high inter-cell interference. Despite the shortcomings of both approaches they allow us to evaluate the system for two extremes concerning the realised I_{oc}, leading to upper and lower bounds of the performance. Furthermore, as we will show later, some schedulers are not sensitive to the chosen method.

4.4.4 Performance of mobile stations

The combined flow-level behaviour of the reference and neighbour cell(s) can be modelled by a multi-dimensional continuous-time Markov chain. The data rates calculated at the packet level - arrival rate from the Poisson process, see Section 4.3, and the state-dependent throughput, see Section 4.4.1 - map to the transition rates in the Markov chain according to Equation (3.6). The specific effect of the scheduling scheme and the impact of inter-cell interference are captured by the state-dependent throughputs. The exact form of the Markov model for the whole system depends on the scheduling schemes and on the discussed scenario. Since in Scenario I a RC and a NC are considered, the Markov model has $K + K \times S$ dimensions, see Section 4.4.1. In the case of Scenario II, since for five of the six NCs the induced inter-cell interference is represented by average value, only the reservation B_{oc} needs recalculation but the form of the Markov model of Scenario II is the same as the model in Scenario I.

Due to the complex form of the Markov model simulating is the more appropriate option to find the steady-state distribution. Subsequently, we can derive the mean flow transfer time realised by a user in a particular zone. In addition, we also determine the overall *budget excess probability*, i.e. the fraction of time that the total received power at the RC's base station ('useful' signals plus interference) exceeds the target level B. Such situation occurs if the actual inter-cell interference exceeds the anticipated reservation level B_{oc}. In order to calculate the overall budget excess probability we need to measure it first for each state of the Markov chain, visited during simulation. The probability is measured over a (single) scheduling cycle and equals the ratio of the number of TTIs in which B is exceeded and the total number of TTIs in the cycle. Appropriately weighting the budget excess probability in a state with the state duration gives us the overall budget excess probability.

4.4.5 Mutual influence of NC and RC

In a deployed network neighbour cells affect each others resource management and power allocation in particular. Due to the shared utilisation of the same radio spectrum in all cells of the network, in UMTS/EUL this interaction can negatively influence the available power budget. Hence the mutual impact neighbour cells have on each other needs to be examined and appropriately modelled (if necessary) for correct system representation. This mutual influence is in fact a dynamic process of transmit power adaptation by the mobile stations which

reaches equilibrium when all users are satisfied by their assigned powers (can guarantee their transmissions). In our model such dynamic adaptation can be represented by an iterative process of transmit power calculation. Ideally, the iterations stop when the equilibrium powers are found. However, a big number of iterations could hinder evaluation by unacceptably increasing computation time.

We have evaluated the changes in power levels as a result of iterative power adaptation in the context of scenario II. Based on the results we conclude how crucial it is to include the iterative process in the model. The investigation is done considering Scenario II, i.e. a RC with six NCs. Our findings are presented in Table 4.1. We have performed four independent experiments, which differ in the total average inter-cell interference $I_{oc,av}$ generated by all six NCs[1]. Each experiment covers 600 000 system states and for each state an iterative power adaptation between the reference cell and one neighbour cell is run. At each iteration step the change in power allocation is recorded.

Eventually, we were able to determine (i) the probability $Pr(\delta P > x)$ that the power change at the second iteration step exceeds a certain threshold x; and (ii) the maximum recorded power change $\max\delta P$ taken over all state changes. As Table 4.1 shows $\max\delta P$ is considerable in all experiments for both PP and FP. However, these large changes seem to occur for specific, rare state changes; thus the overall probability of a power change to be bigger than 5% is negligible. Given that the power change decreases in the iteration number, based on the results we consider it sufficient to account only for the first iteration, i.e. the influence of the NCs on the RC.

Although we simulate state changes in a single neighbour cell the investigation is representative for the whole system because in practice every change (independently in which neighbour cell) leads to recalculation of the power levels.

4.5 Numerical results scenario I

In this section we present a quantitative evaluation of the inter-cell interference and its impact on the performance of the scheduling schemes for the model discussed in Scenario I. As performance parameters we considered budget excess probability and mean flow transfer time. Additionally we discuss the inter-

[1] The average value of the NC, where the state changes dynamically, in the wighted average value of all inter-cell interference instances generated during simulation. In the other five abstract NCs a pre-set average value is used.

Table 4.1: Impact of first iteration on received power levels.

$I_{oc,av} * 10^{-14}$	Watt	2.2	3	3.4	4.7
Partial parallel					
max δP	%	6%	12%	22%	40%
$Pr(\delta P > 5\%)$	%	0.00%	0.01%	0.04%	2.00%
Full parallel					
max δP	%	10%	21%	39%	73%
$Pr(\delta P > 5\%)$	%	0.02%	0.13%	0.19%	45%

$I_{oc,av}$	mean inter-cell interference
max δP	maximum power change
$Pr(\delta P > 5\%)$	probability the power change to be bigger than 5%

cell interference process. For the simulations we extended the model of the MATLAB simulator we introduced in Chapter 3.6.3 such that to accommodate the simulation of flow dynamics in the neighbour cell. The result generation for a single scheme and a particular simulation set up took typically about 10 minutes.

4.5.1 Parameter settings

In the numerical experiments we apply a system chip rate r_{chip} of 3840 kchips/s, a thermal noise level N of -105.66 dBm and a noise rise target η of 6 dB. Hence the total received power budget is equal to $B = \eta N$. A self-interference of 10% is considered, i.e. $\omega = 0.9$. The assumed path loss is given by $L(d_i) = 123.2 + 35.2 \log_{10}(d_i)$ (in dB) with d_i the distance of user in zone i to the base station in kilometre.

Both cells are split in $K = 10$ zones[2] and the neighbour cell in $S = 6$ sectors[3] (see Figure 4.3(b)). Given a cell radius of 1.65 km we applied straightforward link budget calculations to determine the zone radii, see Table 3.1, the path losses and the maximum received powers at the reference cell, originating from users in both the reference and neighbour cell. In the calculation of the instantaneous rate we have used an E_b/N_0 target of 5 dB for EUL transmissions and

[2]Extensive numerical experiments showed that this granularity is sufficient for our purposes. Note that the division is relevant only for the evaluation of the internal cell behaviour and hence does not depend on the number of evaluated neighbour cells.

[3]Our calculations show that working with six sectors delivers acceptable detail of differentiation between positions and still keeps the complexity of the model reasonably low.

$P_{\max}^{tx} = 0.125$ Watt. The applied mean file size is $F = 1000$ kbit and the flow arrival rate is set to $\lambda = 0.7$ flows/sec for both cells.

4.5.2 Inter-cell interference distribution

For each of the scheduling schemes, the induced inter-cell interference in a particular system state can be characterised by a unique cumulative distribution function (CDF), see [56], Chapter 2. Deconditioning the state-dependent CDF with respect to the different system states we obtain a scheduler-specific overall CDF (for a given arrival rate), as depicted in Figure 4.6. Observe that due to the limited set of user locations and therefore possible inter-cell interference values, these (overall) CDFs have a discrete step-like form. This is best observed for the OBO scheduler, as its single-user policy leads to a CDF with a very limited set of possible values.

An important difference between all schemes is the maximum inter-cell interference level that is observed. As expected the OBO scheme has the lowest maximum, due to the single-user policy. For the same reason its curve is shifted to the left indicating low interference values exhibit more often. The PP-w scheme has the highest maximum, because it jointly schedules users that are close to the reference cell in the same TTI. Furthermore, the CDF of the FP scheme is the farthest to the right implying that high interference values are more prone to occur. This could be explained by the longer service time particularly when applied to users at highly influential locations (close to the reference cell edge).

We use the CDF curves as a parameter to compare the different scheduling schemes in terms of the inter-cell interference process that they define. Later, when we evaluate the flow-level performance, we use a track record of the average inter-cell interference level generated during a cycle for each particular state in which the NC has passed.

4.5.3 Budget excess probability

As described in Section 4.4.3, several approaches are possible to reserve part of the total interference budget for inter-cell interference. The particular choice results in different budget excess probabilities, see Section 4.4.4. These probabilities are an end result defined also by the specific arrival rate for which they were computed.

On the one hand, a high budget excess probability is undesired, since it may lead to packet loss, which would require data retransmission and hence cause

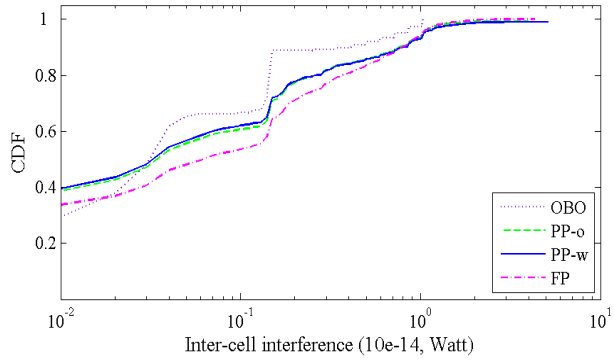


Figure 4.6: CDF of the inter-cell interference levels for each of the different schedulers.

Table 4.2: Budget excess probability.

reservation level	OBO	FP	PP-w	PP-o
no reservation	5.00%	33.00%	12.40%	11.94%
20% percentile	2.29%	0.00%	1.28%	1.29%
40% percentile	1.69%	0.00%	1.16%	1.15%
60% percentile	0.81%	0.00%	0.14%	0.13%
80% percentile	0.19%	0.00%	0.00%	0.00%
100% percentile	0.00%	0.00%	0.00%	0.00%

increase in the transfer times. On the other hand, if a low excess probability is achieved by an excessive reservation level, the available budget for EUL data transmissions is unnecessarily small, causing low transfer rates and long flow transfer times.

Our findings are presented in Table 4.2. Observe that, as expected, the budget excess probability is decreasing in the reservation level. In the extreme case that the maximum reservation level is applied, the budget excess probability is zero. If no reservation is applied, the OBO scheme induces the lowest budget excess probability. This is due to its single-user scheduling policy and the fact that a user can mostly not fill up the budget on its own, which causes fewer excess events compared to PP and FP, whose goal is maximum utilisation.

When using percentile-based reservations, these percentiles are generally ex-

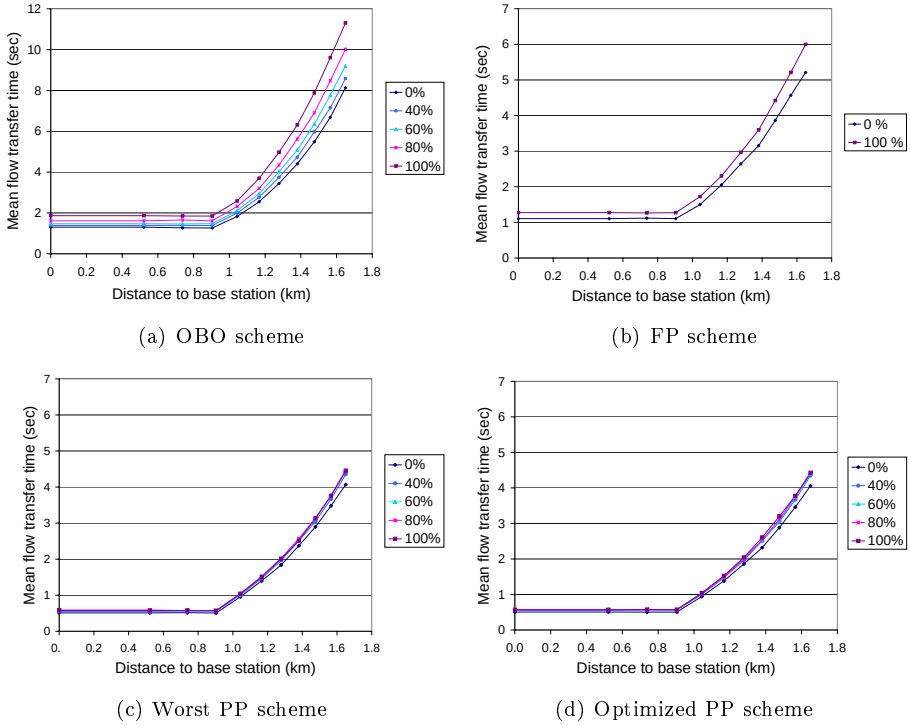
tracted from a discrete value-based CDF, which is non-trivial. In our implementation the next-higher percentile is selected as a reservation level. In particular, in the case of FP scheduling, the inter-cell interference level in a given system state, and hence the reservation in that state, is inherently deterministic. Although Table 4.2 indicates that for a given reservation percentile the two PP schemes lead to somewhat lower budget excess probabilities, the small differences must be regarded with care in light of the above-mentioned discretisation effect. For instance, considering a target reservation level of 20% and a system state with six users in the reference cell, the OBO scheduler establishes a cycle length of six TTIs and thus effectively bases the reservation on the 33% percentile, while the PP scheduler may have a cycle length of e.g. two TTIs and thus effectively apply a reservation based on the 50% percentile. This higher relative reservation, combined with the fact that PP generally establishes a higher inter-cell interference due to parallel transmissions, results in even higher absolute reservation and consequently low budget excess probability.

4.5.4 Mean flow transfer times

We now assess the impact of the different scheduling schemes on the mean flow transfer times in the reference cell, conditional on the distance from the mobile station to the base station. It is stressed that aside from the intra-cell impact of the scheduler, viz. the different policies according to which the budget is shared by the active users, the schedulers also affect the level of this shared budget via the reservation that is applied to cope with the scheduler-specific inter-cell interference effects. Although it is readily understood that the mean flow transfer times will increase in the reservation level, we are particularly interested in the performance impact of the scheduling scheme.

As Figure 4.7 shows, indeed, using no reservation results in the lowest mean flow transfer times, but this comes at a cost of high budget excess probability as shown in Table 4.2. This has negative impact on transfer times due to data retransmissions. Although we have chosen not to include in the model the system reaction to exceeding the budget, the effects of that should be well understood.

Applying a percentile-based reservation for inter-cell interference reduces the budget excess probability but leads to an increase in the mean flow transfer times, since a smaller part of the total budget B is available to the EUL users in the reference cell. The impact of the percentile value on the flow transfer times is negligible for the PP scheme, but is significant for the OBO scheme. The longer cycle length of OBO, i.e. the larger set of inter-cell interference values, allows



(a) OBO scheme

(b) FP scheme

(c) Worst PP scheme

(d) Optimized PP scheme

Figure 4.7: Mean flow transfer times realised by the three scheduling schemes.

choosing different percentiles to result in different reservation values, while in PP the different percentiles might result in the same value.

Compare now the difference in the achieved mean flow transfer times under no and maximum reservation levels. It is observed that the relative impact is largest under OBO scheduling. This seems to be somewhat counter-intuitive given the relatively low inter-cell interference levels established under OBO scheduling (see Figure 4.6). Note however that under OBO scheduling, the highest inter-cell interference levels are caused by MSs close to the reference cell and hence relatively far from their serving base station. Such MSs have low transfer rates, long transfer times and hence have a long-lasting negative effect on the budget availability in the reference cell. In contrast, the maximum inter-cell interference levels of the PP and FP schemes, even if higher in absolute value,

have a smaller effect on the budget availability in the reference cell, because the schemes serve the flows faster.

If we concentrate on the distinct implementation of the PP scheduler, worst and optimized, we observe that the performance difference in terms of the location-dependent mean transfer times is negligible. Apparently, the fact that PP-w requires higher absolute reservation levels due to its larger inter-cell interference variability (see Figure 4.6) hardly affects the performance at flow level. This is most likely because the neighbour cell states inducing higher reservation levels do not occur very often and therefore their contribution to the mean flow transfer times remains small. Hence the exact ordering policy of a PP scheduler is of importance for the inter-cell interference pattern in individual system states, but hardly affects the performance at flow level.

We can conclude that applying the maximum reservation level seems to be an acceptable option for the FP and PP schedulers, while for the OBO scheme lower reservation percentiles are more suitable. Regardless of what reservation level is applied, the PP scheme is noted to outperform the other schemes, while for the current scenarios situation the FP scheme performs second best and the OBO scheme yields the worst performance.

4.6 Numerical results scenario II

In this section we evaluate performance for the more complex Scenario II. We have simulated 600 000 state changes in one of the NCs (recall that the abstract NCs generate constant inter-cell interference) and half a million state changes in the reference cell. The more complex form of the model requires longer simulation times, e.g. 20 to 40 minutes, compared to Chapter 3.

First, we lead a discussion on the difference in cell coverage depending on which parameter - instantaneous data rate or flow throughput - is used. In the former case we introduce *idealistic coverage*, denoted C_{id}, based on instantaneous rates and without considering inter-cell interference. In the latter case a *realistic coverage*, denoted C_r, is evaluated that is based on flow dynamics and accounts for inter-cell interference. Next, we compare the realised mean flow transfer times for three schemes - OBO, FP and PP (optimal implementation). The parameter setting is identical to the one for scenario I, see Section 4.5.1.

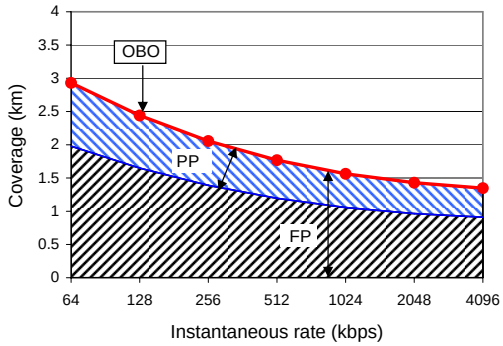


Figure 4.8: Idealistic system coverage C_{id} based on instantaneous rates, for OBO, PP and FP scheduler.

4.6.1 System coverage

Estimating the coverage of a single cell is crucial for the proper planning of the whole cellular network. As coverage we define the maximum distance at which a particular data rate can be realised (for a given arrival rate). Hence, this particular data rate is realised by a user located at the cell edge, i.e. in the last cell zone.

Calculating coverage purely based on user's characteristics, e.g. transmit power, is straightforward to do but does not reflect environmental factors such as interference and could thus lead to service starvation. For example, interference effectively shrinks coverage and by this leaving users at the cell edge deprived from (appropriate) service. Hence, in order to guarantee service to all its customers a network operator needs to consider all factors that could significantly affect coverage.

We begin with a discussion on the idealistic coverage C_{id}. Although C_{id} disregards flow behaviour and inter-cell interference, it allows us to isolate the impact of the scheduling scheme on the coverage. However, to provide a more practical estimation we later calculate a realistic coverage C_r that accounts for these factors.

The C_{id} graphs are presented in Figure 4.8. Their understanding requires careful analysis incorporating knowledge of the schedulers. A one-by-one sche-

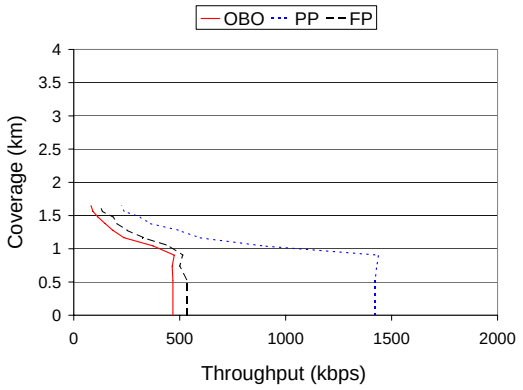


Figure 4.9: Realistic system coverage C_r accounting for flow-level dynamics and inter-cell interference for OBO, PP and FP scheduler; arrival rate of 0.6 flows/sec over the whole cell.

duler has an instantaneous rate (of a particular user), which is not influenced by interference from other EUL users in the cell. Therefore, from Equation (4.1) a straightforward relation between rate and coverage can be established. Hence, to each possible data rate value corresponds a single coverage distance and the coverage for a range of rates can be represented by a single line graph. On the contrary, a scheduler that allows parallel transmissions inherently suffers from interference from other EUL users. The coverage is interference dependent and can be represented by an area of possible realisations. The area is characterised by a lower and upper bound. The upper bound is determined by a single-user state in which case PP and FP behave as a OBO scheduler. Worst case interference, i.e. full budget utilisation, sets the lower bound. As in PP scheduled users transmit at maximum powers, we can easily calculate the lower bound on coverage, see Figure 4.8. Such positive minimum does not exist for full parallel because by definition the scheme does not set a lower limit to power decrease. Theoretically very large number of active users can lead to fairly small transmit powers and coverage will drop towards zero, see Figure 4.8.

In order to provide more realistic values and practically useful observations we calculate a realistic coverage C_r, which takes into account the effect of the flow-level dynamics and the inter-cell interference. Figure 4.9 shows that the realised throughput, for a given network load, are considerably lower than the

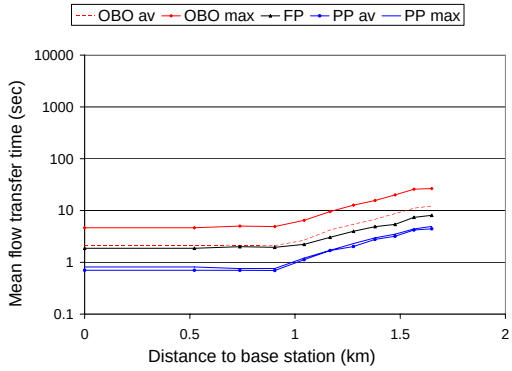


Figure 4.10: Mean flow transfer time for a load of 0.6 flows/sec for each zone in the RC.

upper bounds from Figure 4.8, which is a result of the negative impact of inter-cell interference and flow-level dynamics on the effective throughput. Note that the difference between C_{id} and C_r varies depending on the applied system load - higher load leads to higher difference. However, in all cases the realistic coverage stays below the idealistic upper bound.

Particularly interesting is the difference in coverage C_r between the schedulers. OBO shows to have the worst coverage even if it had the best in Figure 4.8. The advantages of a single user policy in terms of low intra-cell interference cannot compensate for the poor efficiency of budget utilisation and the long transmission cycle. PP outperforms the other two schemes because of its optimal combination of parallel transmissions at maximum transmit power.

4.6.2 Mean flow transfer time

As Figure 4.10 shows, the mean flow transfer time depends on the user's location in the cell, i.e. the distance to the base station. Long transmission path, i.e. for far located user, means high path loss and low received power. During the evaluation of the OBO and PP schemes we have used both the average and maximum reservation B_{oc} to model inter-cell interference at the RC. Therefore OBO and PP are represented in Figures 4.10 and Figure 4.11 by two graphs. It seems that PP can effectively combine the advantages of the other two sche-

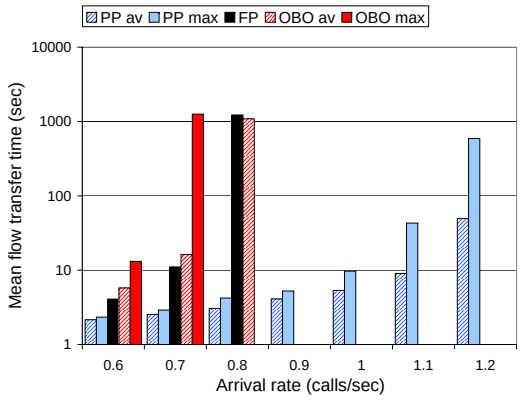


Figure 4.11: Mean flow transfer time aggregated for all zones and for a range of arrival rates.

mes and by this it outperforms them. On the one hand, FP, although fast to reassign the budget, suffers from decreased received powers and increased intra-cell interference. On the other hand, the performance of OBO, which keeps the maximum received powers, suffers from the long cycle, consequence of the single-user policy.

Figure 4.11 shows the same results but spanned over a range of arrival rates. An advantage of working with range of arrival rates is that we can identify tendencies. For example, the impact of the B_{oc} reservation strategy, i.e. average or maximum. The difference between the two strategies depends on the cycle length. A longer cycle decreases the contribution of the maximum value on the average and thus causes significant differences between the maximum and average reservation and thus in the registered mean flow transfer times. This explains why the results for OBO, which has long cycle, lay far from each other and the ones for PP are close. We can conclude that for a PP scheme it is irrelevant which B_{oc} reservation is chosen. However, in the case of a OBO scheduler using a high percentile value, e.g. 90%, is more beneficial.

A sudden, significant increase in the mean flow transfer time indicates that the system reaches its stability limit. For example, a OBO scheme with maximum reservation reaches instability around an arrival rate of 0.7 flows/sec. It appears that a system with partial parallel scheduler has twice as big capacity as a system with OBO or FP scheme. The less efficient the scheduler and the

longer the service time the faster instability is reached. This explains why in Figure 4.11 OBO with maximum reservation reaches saturation first while a PP scheduler with average reservation - last. Bear in mind that with average reservation there is a possibility that the available budget is exceeded. As we explained in Section 4.5.3 this could negatively influence mean flow transfer times. Therefore our results should be taken as a reference point and not an absolute performance metric.

4.7 Conclusions

In this chapter we discussed the mutual effects inter-cell interference and packet scheduling have on each other. Three scheduling schemes with distinctive approaches towards parallel transmissions - none, few or all users - were presented termed one-by-one (OBO), partial parallel (PP) and full parallel (FP) respectively. We were particularly interested in the impact of the scheduling schemes on the pattern of the inter-cell interference. The specific (flow-level) performance response of the scheduler to the interference has been investigated as well.

In order to be able to examine different aspects of the mutual influence we have developed two modelling scenarios. In the first, more simplified approach, we observe the behaviour of a system of two cells. Such a basic model keeps the complexity level low and allows us to make important observations on the details of the inter-cell interference process and its impact of performance. The second more complex scenario, considers a system of a central cell and its six neighbouring cells. By keeping abstraction level high we were able to evaluate the behaviour of packet schedulers in realistic terms. For the evaluation of both models we modified the hybrid analysis approach developed in Section 3.4 for the basic single cell scenario.

We were able to provide several interesting results and insights on deployment. First, we quantified the impact reference and neighbour cells have on each other's applied power levels (and thus interference levels), which leads to iterative power adaptation. We have shown that this mutual influence can be largely captured by a single iteration. A beneficial consequence of this is that performance computations can be considerably sped up. In addition, we have proposed and evaluated several approaches to set up an adaptive reservation of the total interference budget in order to account for inter-cell interference. For each examined scheduling scheme we were able to make recommendations on the proper choice of reservation approach.

The resulting performance has been evaluated in terms of measures such

as mean flow transfer times, coverage and probability to exceed the available interference budget at the base station. Our most prominent findings are: (i) the extent to which the schedulers' performance is affected by the chosen reservation approach varies greatly; in particular, the OBO scheduler is very sensitive to how the reservation is chosen; (ii) the particular implementation of a partial parallel scheme affects the inter-cell interference process but hardly influences data traffic performance at flow level; and (iii) a PP schemes is the best in performance followed by FP and OBO.

Part II

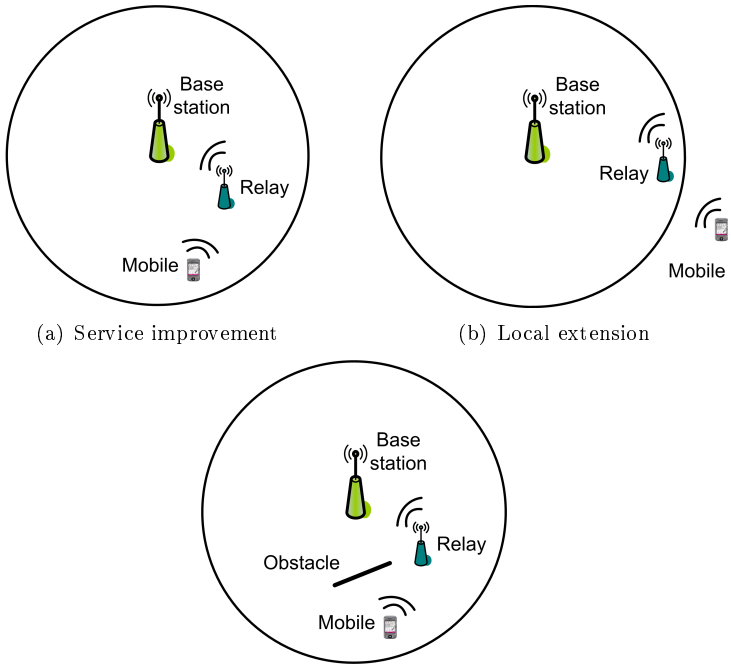
Relaying

5 Flow-level analysis of relay-enabled uplink scheduling

5.1 Introduction

The results in the previous chapters among others showed and quantified, for different scheduling schemes, the impact of the location of the users on the throughput they receive. In particular, users near a cell edge suffer considerably from the large distance to the serving base station. Improving the performance of remote users has received a lot of attention. One possibility to improve their performance is to give them higher priority in the scheduling process, but from the insights gained in the previous chapters it is clear that this will go at the cost of the performance of other users, see e.g. also [13, 41, 42]. Another, more beneficial solution might be relaying, which was briefly introduced in Chapter 1. In this chapter we intend to investigate the various effects that relaying has on the performance, accounting also for the specifics of the scheduling scheme.

Further in this section we will first, in Section 5.1.1, briefly introduce the concept of relaying. Next, Section 5.1.2 discusses available studies in the area of relaying for mobile networks. The novelty of the research presented in this chapter and the main contributions are presented in Section 5.1.3. Finally, the structure of this chapter is given in Section 5.1.4.



(a) Service improvement (b) Local extension

(c) Coverage extension

Figure 5.1: Relaying applied for mobile networks.

5.1.1 The relaying concept

Relaying is a concept that can be implemented at the physical layer (referred to as amplify-and-forward) or at the link layer (referred to as decode-and-forward), see [50]. Placing a relay station can positively influence data rates by breaking a long communication path (mobile station - base station) into two shorter paths (mobile station - relay and relay - base station). This is illustrated in Figure 5.1. The performance gain that may then be obtained (depending among others on the position of the mobile station with respect to the relay station and the base station) is essentially due to the earlier mentioned non-linear decrease of signal strength as function of the path length, see Section 2.3.3. Additional benefits of relaying resulting from this property may be achieved in terms of lower required transmit powers of the mobile stations, which is beneficial for battery lifetime,

and decreased inter-cell interference, which is beneficial for the network capacity.

Relaying is primarily considered as a mean for user performance improvement (e.g. increasing data rates), see [54, 65, 67]. This is the application case in Figure 5.1(a). Still, there are many studies where the focus is on the use of relaying for coverage extension. In some studies, e.g. [58, 59, 68], coverage is extended towards small areas that 'happen' to fall just outside the coverage provided by the base station as illustrated in Figure 5.1(b). In other studies, see [72], the signal penetration is obstructed by natural or artificial obstacles, see Figure 5.1(c), which is e.g. a common problem for indoor signal propagation. Mostly, implementation or deployment issues are then addressed, such as what functionality a relay should have or what is an optimal relay planning to cover a target area, see [59, 68].

The idea of relaying is attractive, however it also poses new implementation and resource management choices, e.g. how to appropriately align scheduling and relaying. It is our primary goal in this chapter to investigate the potential performance benefits of relaying in the uplink, hereby focusing on the role of scheduling.

In the following section we will briefly discuss related work on relaying available in the literature.

5.1.2 Literature - application of relaying

A good introductory research to the topic of relaying, which covers relay deployment concepts in wireless networks, is presented in [50]. The authors discuss topics such as cooperative and multi-hop relaying, routing and radio resource management. Other studies, e.g. [37] and [70], are more focused on non-technical issues such as cost assessment of relay deployment in the context of beyond-3G systems. The potential cost reduction is evaluated related to factors such as number of deployed relays or transmitted powers.

The type of mobile network technology has strong impact on the strategy towards introducing relays in the network, and there is a variety of studies addressing relaying in specific mobile technologies. In the scope of UMTS downlink [64] discusses the deployment of relays and shows that in particular in the presence of strong interference, for a single MS, sending via the relay is beneficial in terms of packet errors and delays. Other studies which analyse the introduction of a relay to a UMTS downlink upgraded with HSDPA are [25] and [31]. The potential of relaying for OFDM-based technologies has also received a lot of attention. Several interesting studies in the specific context of LTE networks can be found in [8, 54, 58] and [69]. In particular, the authors of [18] propose

several deployment scenarios with detailed discussion on the spectrum utilisation, but do not consider the impact of the particular scheduling scheme and other system characteristics. A few examples of studies that concentrate on the use of relaying in WiMAX networks are [7, 65] and [78].

Unfortunately most studies on relaying concentrate on the downlink and only few consider uplink scenarios. Furthermore, to our knowledge there are no papers specifically focusing on the role of packet scheduling in scenarios with relaying, in particular how different 'relay-enabled' scheduling strategies influence the performance of mobile stations. Moreover, similar as in the context of 'plain' scheduling (see the previous chapters), in most studies on relaying simple traffic scenarios with fixed numbers of active users are considered, i.e. the influence of flow-level dynamics on the user performance is mostly ignored.

In the present chapter we will investigate the application of relaying for improving the user's performance, where we pay particular attention to these two aspects (scheduling and flow-level dynamics) that are hardly considered in the existing literature.

5.1.3 Contributions

Our goal in this chapter is to provide an initial explorative study on the potential gains of relaying if applied to an uplink of a mobile network. Our study is set in the context of UMTS/EUL. However, the proposed analysis approach has the potential to extend towards other mobile systems, e.g. LTE.

We identify two main issues which accompany the introduction of a relay: identifying the various aspects in which relaying can affect performance and determining how relay transmissions can be accommodated in the scheduling scheme. As possible benefits of relaying we expect better service, i.e. higher data rates, and decreased interference, due to lower required power to reach the receiver. At the same time sending via a relay introduces an additional packet transmission delay as well as consumes two times more radio resource.

In order to determine whether relaying is beneficial and to quantify its impact on performance we have compared the performance of the proposed relay-enabled scheduling schemes to a reference scheme with no relay. Performance is observed in terms of data rates, flow throughputs and energy consumption as well as generated inter-cell interference. Based on the promising performance trends, registered in this chapter, we continue in Chapter 6 to investigate specific relay scenarios in more detail and try to provide more insights of the relaying aspects.

We can summarise the major contributions of the study presented in the current chapter as follows:

- Determining the various performance gains, e.g. increased data rates, that relaying can offer if appropriately introduced in the uplink of a mobile network. We have investigated the impact of relaying on aspects such as performance, energy consumption and inter-cell interference.
- Evaluating the performance on both packet level, in terms of data rates, and on flow level, in terms of flow transfer times and throughputs, for several relay-enabled scheduling schemes. The schemes, which we define, can be grouped into two categories - theoretical proposals that help us to establish upper bounds on performance, and practically feasible proposals that provide us with more realistic evaluation figures. More importantly, we show that by performing flow-level analysis we were able to determine that relaying benefits even users that do not use the relay.
- Developing a flexible and computationally-efficient analysis methodology and showing its importance for performance evaluation. The methodology is a modified version of the one proposed in Chapter 3 and hence combines analysis on packet and flow level.

5.1.4 Outline

This chapter is organised in six sections. Section 5.2 discusses the relaying concept and introduces the scheduling schemes of interest. The considered two modelling scenarios and any assumptions made during the research are presented in Section 5.3. Subsequently, the methodology applied to evaluate performance is discussed in Section 5.4. Section 5.5 presents our approach to examine the inter-cell interference. The performance results are discussed in two consecutive sections, namely in Section 5.6 and Section 5.7, corresponding to the two evaluation scenarios. Eventually, in Section 5.8, we summarise our work.

5.2 Relay-enabled scheduling

A relay station (RS) is another participant on the communication path whose transmissions need to be accommodated in the radio resource. In Chapter 2 we introduced scheduling as the mechanism that manages the distribution of the radio resource over the active users. Recall also that in UMTS/EUL the key

channel resource is the shared budget B, which is the total received interference at the base station. Since the channel access is organised in Transmission Time Intervals (TTIs) of 2 ms, scheduling decisions are taken at the beginning of each TTI. In these decisions a relay-enabled scheduler needs to dedicate certain TTIs to the service of relay transmissions. In this section we first elaborate on the effects relaying has on MSs' performance and on transmission ranges. Second, the scheduling schemes proposed for evaluation are presented.

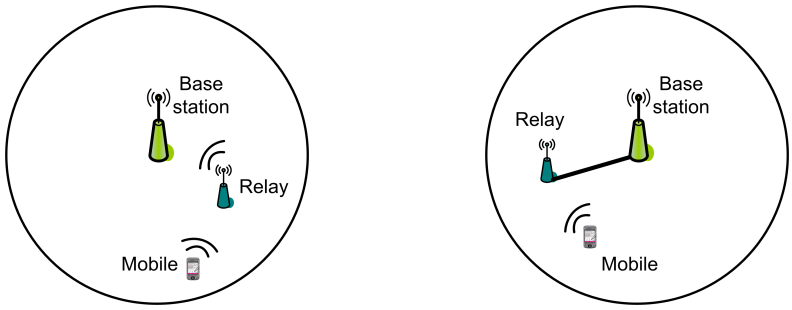
5.2.1 Relay deployment

In the case of uplink the relay station forwards data from the MSs to the BS and by this effectively changes (shortens) the transmission ranges. Advantages of shorter communication path are smaller path loss and lower transmitted power required to reach the receiver. As a result, we can expect improved performance, i.e. higher data rates, and decreased interference towards neighbour cells. However, relaying has the drawback of increased transmission time and lower maximum achievable data rate due to data forwarding. Hence, whether relaying can be beneficial is not a trivial question with a simple answer but depends on (i) the position of the particular MS relative to the relay and base station; (ii) the scheduling scheme; and (iii) the characteristics of the relay station.

Typically the communication between MS and BS is referred to as a *direct link* or *direct path*, see [8]. In contrast, the path MS-RS-BS we term *indirect path*. The indirect path consists of two sub-paths - *mobile sub-path* MS-RS and *relay sub-path* RS-BS. Each (sub-)path is characterized by a set of transmission parameters:

- distance between the communicating devices d_{zz},
- path loss L_{zz},
- transmitted power P_{zz}^{tx},
- duration of a transmission opportunity τ_{zz} and
- realised data rate r_{zz} during a transmission opportunity.

The index zz refers to the specific (sub-)path, i.e. ms for the direct path MS-BS, mr for the mobile sub-path MS-RS, and rs for the relay sub-path RS-BS. The transmission times τ_{mr} and τ_{rs} as well as the realised data rates on the indirect path (r_{mr} and r_{rs}) are scheduler specific and are further discussed in



(a) Radio interface

(b) Fixed infrastructure interface

Figure 5.2: Deployment scenarios for relay-enabled scheduling: (a) with using only radio interface connections and (b) with a fixed-line connection between the base and relay station.

Section 5.4. We will now continue to introduce the specific scheduling schemes considered in our study.

Scheduling decisions are typically taken by the base station. However, it is also possible to place more 'intelligence' in the relay and make it responsible for the scheduling decisions. In order to be able to do that a relay needs information on the experienced channel conditions on the links towards the mobile and base station. Such relay devices are relatively more expensive than the more simple ones, which receive scheduling instructions from the base station. We have chosen to work with base station-based scheduling. Given our choice, several approaches towards incorporating relay transmissions into the scheduling process are possible. In the upcoming Section 5.2.2 we are going to discuss few schemes that we consider interesting to study.

5.2.2 Scheduling schemes

All scheduling schemes that we consider in this study are of the type Round Robin (RR) where mobile stations are given equal opportunity to reach the channel and are served in one-by-one (OBO) fashion. We have chosen a OBO scheme, despite some of its disadvantages, based on the results of previous studies. Some studies, e.g. [53] and [55], show that a OBO strategy is inefficient in resource utilisation when serving users with limited transmit power (a feature associated with battery-fed mobile stations). This inefficiency is attributed to

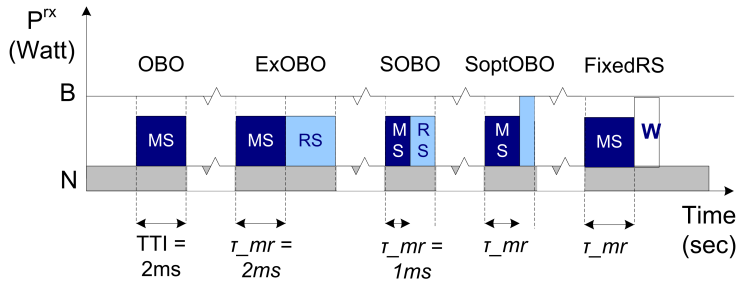


Figure 5.3: Scheduling schemes - OBO, ExOBO, SOBO, SoptOBO and FixedRS. Wireline transmission denoted by W does not happen over the radio interface.

the inability of such users to fully use the available budget B, a situation particularly applicable for remote users. However, in the previous Chapters 3 and 4 we showed that when the condition for full budget utilisation is met a OBO scheme seems to be advantageous. Since a relay is closer to the MSs (lower path loss), it is more likely that a single MS is capable to fill the budget B. In combination with the fact that OBO is also more tractable for evaluation we believe that an OBO scheme is an appropriate choice.

In order to determine what is an optimal manner to schedule relay transmissions, we propose five implementations of a relay-aware OBO scheduler. For all of these schemes the communication between MS and RS and between RS and BS uses the same radio interface, see Figure 5.2(a). In addition, we also discuss a relay-enabled scheme in which the relay station is connected to the base station via an (existing) fixed line, see Figure 5.2(b). As a reference scheduler, which allows us to evaluate the benefits of relaying, plain *OBO* is considered, i.e. without relaying.

All schemes are presented in Figure 5.3 where the x-axis shows the time progress, measured in TTIs, while the y-axis gives the received power at the destination. Note that the maximum received power is determined by the total budget B. Below each of the schemes is presented in more detail concerning the specific power assignment and transmission time.

In OBO a MS always transmits directly to the base station independently of its location in the cell. With τ we denote the duration of a single transmission opportunity, which in OBO equals one TTI, i.e. $\tau_{ms} = 2$ ms. The received power at the base station depends on its location in the cell and can be at most

equal to the data budget B_D.

The *Extended OBO* (ExOBO) scheme assigns (in the case that indirect path is chosen) a single TTI to the transmissions on both sub-paths, i.e. $\tau_{mr} = \tau_{rs} =$ 2 ms. Therefore, for the service of a particular mobile station in a single round of the RR scheme the base station sets aside two consecutive TTIs, meaning $\tau = 4$ ms. Transmissions on a direct path are assigned one TTI.

In a *Shared OBO* (SOBO) scheme transmissions on both direct and indirect paths are assigned one TTI, i.e. $\tau = 2$ ms. In the case of an indirect path, the single TTI is divided into two equal intervals of 1 ms and MS and RS both receive one interval to transmit, i.e. $\tau_{mr} = \tau_{rs} = 1$ ms.

Both schemes ExOBO and SOBO assigning fixed and equal-length transmission times to the two relay sub-paths. As result the received power that characterises the indirect path is determined by the smaller potential received power on one of the sub-paths. This is partly inefficient since power resources are left unused, something that SoptOBO (see below) tries to avoid. Hence, static division may ease the implementation but it could be inefficient when the achievable rates on the two sub-paths differ. Furthermore, the received power can be lower (example in Figure 5.3) or equal to the data budget B_D depending on the user's distance to the base station.

The *optimized SOBO* (SoptOBO) scheduler also uses one TTI to serve a single MS, both for direct and indirect paths, i.e. $\tau = 2$ ms. However, for indirect paths, it selects the transmission times on the sub-paths MS-RS and RS-BS such that the same amount of data is transferred, i.e. $\tau_{mr} r_{mr} = \tau_{rs} r_{rs}$. In such a way each device gets to make the best use of its available transmit power. Although SoptOBO maximises the resource utilisation it is rather challenging for implementation. The selection of the specific transmission times for the sub-paths requires more complex functionality in the base station most notably decoding and recoding of the received signal. In terms of received powers the mobile and the relay station may differ; in the example in Figure 5.3 the relay has higher power capacity than the mobile and can fully use the data budget.

In order to establish a best case performance when the relay station does not form a bottleneck we designed scheduling schemes for which the following two assumptions apply: (i) the interference budget at the relay station and (ii) the transmitted power of the relay station are unlimited. The former assumption implies that mobile stations, given a maximum transmitted power, can realise their maximum achievable data rate towards the relay station. The latter assumption has the consequence that the relay can, independently of its location, fully utilise the budget at the base station. Given that the relay is eliminated as potential bottleneck, we can observe purely the limitations of the scheduling

scheme and mobile station on performance. As result of the made assumptions these 'ideal' schemes are closer to a theoretical design than to a practical solution but they can provide us with some notion of best-case performance.

Such idealised schemes were designed for SOBO and SoptOBO, termed *ideal SOBO* (SOBOid) and *ideal optimized SOBO* (SoptOBOid) respectively. In any other aspects, e.g. TTI assignment, the schemes operate as their counterparts.

The last relay-enabled scheme in our analysis, namely *FixedRS*, makes use of a fixed-line connection between the relay and the base station. Using a fixed-line connection has the advantage that transmissions from relay to base station do not take away radio resources that can be used by the mobile station. Of course, there is an additional cost of connecting the relay station to the base station. This connection could be realised using a wireline connection (e.g. fiber), or using a different (fixed) wireless technology such as a microwave connection. Such line will generally have a limited capacity W, which is typically sufficiently high to allow optimal utilisation of the more costly radio interface. Since the relay sub-path does not form a bottleneck a mobile station served via the relay is able to realise its maximum achievable data rate. Furthermore, the mobile can send a new data packet and does not need to wait for the previous one to be resent by the relay. Hence, the only time penalty of relaying in the FixedRS scheme is an additional TTI necessary for the transmission of the last packet from the RS to the BS. However, compared to the flow transmission times the TTI duration is negligibly small.

5.3 Model

One expected effect of relaying is influencing (improving) the performance of MSs, e.g. their data rates. For this purpose a single cell model suffices. Another expected effect is a decrease in the inter-cell interference I_{oc} due to the lower power a mobile station needs to reach the relay. Modelling a single cell in this case does not allow us to observe the consequences of relaying for the inter-cell interference.

To be able to evaluate the various effects of relaying we designed a model of two cells - a *reference cell* (RC), where performance is examined, and a *neighbour cell* (NC), which is the origin of the inter-cell interference. Depending on the particular behaviour (inside or outside the cell) that we want to observe we can focus on the RC only or consider the system of two cells.

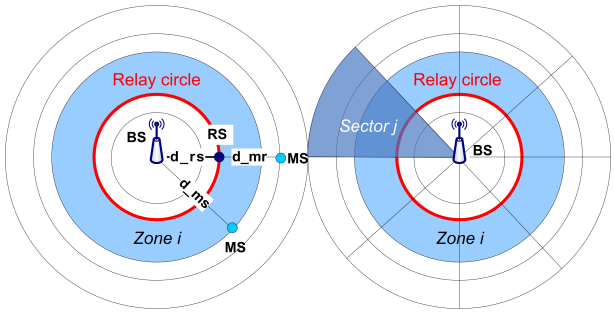


Figure 5.4: Single cell model with cell division in zones. The distance between MS-BS, MS-RS and RS-BS are denoted as d_{ms}, d_{mr} and d_{rs}, respectively.

5.3.1 Evaluation scenario

For tractability, we have defined two evaluation scenarios - Single-cell and Multi-cell. In the *Single-cell* scenario we observe only the performance within the reference cell without taking into account inter-cell interference. Hence, the neighbour cell is entirely ignored. In the *Multi-cell* scenario we determine (i) the inter-cell interference I_{oc}, which a NC generates, and (ii) the performance of MSs in the RC, accounting for I_{oc}. In this second scenario the assumption is made that the NC behaves independently from the RC. Most other modelling assumptions are the same for both scenarios as any differences are explicitly commented on.

5.3.2 Cell model

The model of the reference cell is similar to the single cell model, described in Section 3.3. Hence, the reference cell is divided in K *zones* with equal area as each zone is characterized by its distance to the base station d_i. This distance is taken from the outer edge of the zone and has a corresponding path loss $L(d_i)$, $i = 1, \cdots, K$. The neighbour cell, just as in Section 4.3, is divided in K zones and S sectors in order to allow correct representation of inter-cell interference, see Figure 5.4. The intersection of zones and sectors determines the *segments*. Each segment (ij) is characterised by a distance d_{ij} to the base station of the reference cell; $i = 1, ..., K$, $j = 1, ..., S$. Both distances are measured from the outer edge of the segment (i, j) and can be determined by applying simple trigonometry rules, given an inter-site distance and the distance d_i (towards the

base station of the neighbour cell).

The distribution of (active) MSs over the zones in the RC is represented by the system state vector $\underline{n} \equiv (n_1, n_2, \cdots, n_i, \cdots, n_K)$, where n_i is the number of currently active MSs in zone i. The NC can be described by the more detailed state $\underline{n} \equiv (n_{11}, n_{12}, \cdots, n_{ij}, \cdots, n_{KS})$, where n_{ij} is the number of currently active MSs in a segment (i, j), $i = 1, ..., K$ and $j = 1, ..., S$.

Figure 5.4 introduces some additional notation regarding the reference cell: the distance between a mobile and the base station is denoted d_{ms}, the distance between the relay and the base station is d_{rs} and the distance between a mobile and relay station is denoted d_{mr}. While mobile stations can have different distances to the base and relay station, the location of the relay, in our study, is fixed.

5.3.3 Relay station

In our study a relay is a fixed device with one basic functionality, namely resending data from MS to BS. In most cases the relay uses the same radio interface as the mobile stations to reach the base station. In the special case of FixedRS scheduler the independent link connecting the relay and base station has by default very high capacity such that it never forms the bottleneck on the indirect path. We consider all technological issues related to the fixed line to be transparent for our analysis.

We assume that a mobile station always has the shortest path towards the base station when transmitting via the relay, i.e. mobile, relay and base station lay on a straight line[1]. Although this assumption of shortest path can be optimistic, it simplifies analysis and allows for a more generic evaluation, which provides us with an upper bound on the expected performance gains. A fixed path selection policy based on distance is adopted according to which all MSs located beyond the relay, i.e. $d_{ms} > d_{rs}$, transmit via the indirect path and in all other cases, i.e. $d_{ms} \leq d_{rs}$, direct transmission is chosen. An instantaneous switching process between transmitting and receiving at the RS is assumed where the same frequency band is used.

[1] In fact, this assumption requires unlimited number of relays on a concentric circle but it can be shown that this ideal situation can be very well approximated by eight or ten RS. In principle, the analysis in this chapter can easily be extended to any other assumption concerning the number and position of the RSs. However, for the purpose of the present chapter, this specific choice suffices.

5.3.4 Channel resource

The interference budget B available at the base station is the same also at the relay station and can be derived from an operator-specific noise rise target η, e.g. $B = \eta N$, where N is the thermal noise at the base station. SOBOid and SoptOBOid schemes are an exception since the RS budget is considered to be unlimited. Part of the budget is lost to interference signals. In addition to thermal noise, self-interference of reflected paths of the own signals and inter-cell interference are considered. A variable reservation B_{oc} is introduced at the BS and the RS in the RC to cope with anticipated inter-cell interference. After we account for all interference, the budget available for the service of EUL data users, i.e. *available data budget* B_D, becomes $B_D = B - (B_{oc} + N)$. In the Single-cell scenario, since we abstract from inter-cell interference, $I_{oc} = 0$ and hence $B_D = B - N$. In the Multi-cell scenario B_D varies since B_{oc} is dynamically updated to follow the changes in the actual inter-cell interference I_{oc} generated by the NC. We assume these updates to be instantaneous.

5.3.5 Mobile stations

Mobile stations become active, at uniformly distributed positions in the cell, according to a Poisson process with rate λ, and have to transfer a file with size sampled from an exponential distribution with mean F. Given the assumption for equal size of the zones, the arrival rate per zone in the RC λ_i equals λ/K and the arrival rate per segment in the NC λ_{ij} equals $\lambda/(KS)$. Active MSs keep their position during the file transfer. All users have the same maximum transmit power P^{tx}_{max} but different maximum received power at the base station $P^{rx}_{i,max}$ due to the zone-dependent path loss. Depending on their location either the maximum transmitted power or a lower power is used, i.e. $P^{tx} \leq P^{tx}_{max}$, such that a MS maximises its utilisation of the available budget B_D. A mobile station selects the direct or indirect path purely based on its position - if located between the base station and the relay the direct path is chosen; if located beyond the relay the indirect path is selected.

5.4 Analysis - single-cell scenario

In the following Sections 5.4.1 to 5.4.6 we will analyse the various scheduling schemes in the Single-cell scenario. The same combined packet and flow-level analysis, as described in Section 3.4, is applied. However, the introduction of a relay station requires certain modifications to be made. These modifications

mainly affect the calculation of the received powers and data rates and the form of the Markov chains, which model the flow dynamics in the cell.

5.4.1 Received powers

Already in Section 3.4.1 we derived the expression for the received power as a function of the transmit power and the user specific path loss. In order to ease the rest of the discussion we repeat Equation (3.1):

$$P_i^{rx} = \frac{P^{tx}}{L(d_i)}, \tag{5.1}$$

where the transmit power is given by Equation (3.2). In principle, unless it will not overload the data budget at BS or RS depending on the path, a user always uses its maximum transmit power. The path loss is according to the Okumura-Hata path loss model, namely $L(d_i) = 123.2 + 10a\log_{10}(d_i)$ (in dB); a is the path loss exponent and d_i is the distance between the communicating devices in kilometres. The index i stands for the zone number of a user.

In a relay-enabled mobile network we should differentiate between received power at the base station and received power at the relay. In the case of direct transmission, independently of the scheduling scheme, the received power at the base station is the power realised over the direct path MS-BS, i.e. $P^{rx} = P^{rx}_{ms}$, see Table 5.1. Note that for OBO this is also the only received power that we can identify in the cell since OBO does not use a relay, regardless of the MS's location.

When the relay path is chosen the received power at the base station is the power realised over the relay sub-path RS-BS. The received power at the RS is the power applied at the mobile sub-path MS-RS. The ExOBO, SOBO and SOBOid schemes adopt the same power allocation policy on the indirect path. In particular, the received power on both sub-paths MS-RS and RS-BS is equal and is determined by the lower of the two powers P^{rx}_{mr} and P^{rx}_{rs}, that could be realised over the mobile and relay sub-path respectively. Hence, the received power at the base station (and the relay) is given by $P^{rx} = min(P^{rx}_{mr}, P^{rx}_{rs})$, see Table 5.1.

The received power at the base station for SoptOBO and SoptOBOid differs from the received power at the relay, see Table 5.1. This is a consequence of the different power allocation on mobile and relay sub-path in order to satisfy the condition $\tau_{mr} r_{mr} = \tau_{rs} r_{rs}$. Note that the two schemes differ in the maximum transmit power of the relay station.

Table 5.1: Received power at BS and RS for the various scheduling schemes.

	OBO	ExOBO	SOBO SOBOid	SoptOBO SoptOBOid	FixedRS
at BS(direct path)	P_{ms}^{rx}				
at BS(relay path)	-	$min(P_{mr}^{rx}, P_{rs}^{rx})$		P_{rs}^{rx}	-
at RS(mobile path)	-	$min(P_{mr}^{rx}, P_{rs}^{rx})$		P_{mr}^{rx}	P_{mr}^{rx}

Finally, the FixedRS scheduler registers only received power at the relay station P_{mr}^{rx} since the connection between relay and base station is via a fixed line, see Table 5.1.

5.4.2 Instantaneous rate

If we adapt the definition from Section 3.4.2 for the case of a relay-enabled scheduler, we can define the *instantaneous rate* r_{zz} as the data rate achieved on a (sub-) path zz considering only the transmission channel conditions, i.e. received power and interference. Hence, this is the rate realised during the period τ_{mr}. The instantaneous rate of a user in zone i is given by:

$$r_{zz,i} = \frac{r_{chip}}{E_b/N_0} \cdot \frac{P_i^{rx}}{N + I_{oc} + (1-\omega)P_i^{rx}}. \tag{5.2}$$

In the above equation r_{chip} is the system chip rate and E_b/N_0 is the energy-per-bit to noise ratio. The index $zz = \{ms, mr, rs\}$ refers to the (sub-)path. The inter-cell interference I_{oc} depends on the chosen modelling scenario and the type of cell:

$$I_{oc} = \begin{cases} 0 & \text{for Single-cell scenario,} \\ 0 & \text{for Multi-cell scenario, NC,} \\ \neq 0 & \text{for Multi-cell scenario, RC.} \end{cases} \tag{5.3}$$

The maximum possible data rate that a MS from zone i can realise on a given path is determined by the condition that the user can fill up the budget B_D on its own, i.e. $P_i^{rx} = B_D$.

Table 5.2: Effective rate for the various scheduling schemes.

	OBO	ExOBO	SOBO SOBOid	SoptOBO SoptOBOid	FixedRS
direct path	r_{ms}				
indirect path	-	$\frac{1}{2}r_{mr}$		$\frac{r_{mr}\tau_{mr}}{\tau}$	r_{mr}

5.4.3 Effective rate

The effective rate r_{eff} accounts for the effects of relaying and is the rate realised by a MS for the whole period τ dedicated by the base station to the service of that particular user, including its relay transmission. On the direct path the effective rate is the same as the instantaneous, i.e. $r_{eff,i} = r_{ms,i}$ for a user in zone i, because the entire τ is used by the MS. This is also the rate that a user achieves in the OBO scheme.

On the indirect path however, due to data forwarding, the effective rate is lower than the instantaneous and depends on what part of τ is used by the mobile, i.e. on τ_{mr}. Given the scheduler-specific time assignment policy, for the effective rate realised on the indirect path MS-RS-BS is given in Table 5.2:

SOBO, SOBOid and ExOBO assign only half of the total transmission opportunity τ to a mobile station, i.e. $\tau_{mr} = \tau/2$, which explains the factor $1/2$ in Table (5.2). In SoptOBO and SoptOBOid $0 < \tau_{mr} < TTI$ holds depending on the MS's location such that $r_{mr}\tau_{mr} = r_{rs}\tau_{rs}$ and $\tau_{mr} + \tau_{rs} = TTI$. Finally FixedRS gives the total period τ to the mobile and hence effectively behaves as a OBO scheme, i.e. $r_{eff,i} = r_{mr,i}$.

5.4.4 State-dependent throughput

Both the instantaneous and the effective rate do not take into account the number n of active MSs in the cell. As we showed in the previous Chapters 3 and 4 however, how many users actively transmit can significantly affect the performance of an OBO type of scheme and should be accounted for. In particular, after being served a MS may have to wait several TTIs before receiving service again, which causes a decrease in its actual data rate. To reflect the impact of n we introduce the parameter *state-dependent throughput* $R(\underline{n})$, which is given by:

$$R_i(\underline{n}) = \frac{r_{eff,i}}{n}, \tag{5.4}$$

where $i = 1, ..., K$ indicates to which zone a user belongs. Equation (5.4)

holds for all schemes but ExOBO. The difference for ExOBO originates from the different number of TTIs assigned to MSs using the direct and the relay path - one and two TTIs respectively. Hence, for the state-dependent throughput in this case we can write:

$$R_i(\underline{n}) = \begin{cases} \frac{r_{eff,i}}{n + n_{relay}} & \text{for the direct path,} \\ \\ \frac{2r_{eff,i}}{n + n_{relay}} & \text{for the relay path,} \end{cases} \tag{5.5}$$

where n_{relay} is the number of MSs which make use of the relay station. Note that $r_{eff,i}$ is different for each path according to Equation (5.2).

5.4.5 Flow throughput

At several occasions, e.g. Section 4.5.4, we commented on the importance of including flow dynamics, i.e. the changing number of active users, in the performance analysis of packet scheduler. In this chapter we continue to conform to this argument and introduce performance measures such as mean flow transfer time and *flow throughput*. Flow throughput is the average data rate seen by a flow for the duration of its transfer, during which period the number of active users has changed in a random manner. These changes can be very well modelled by a continuous time Markov chain, see Section 3.4.4. The assumptions made in Section 5.3, namely a Poisson flow arrival process and exponentially distributed flow size, permit such mapping to be made. A state in the Markov chain corresponds to a system state $\underline{n}$ and the transmission rates of the Markov chain can be mapped to the arrival rate of the Poisson process and the state-dependent throughputs, derived in Section 5.4.4.

For the performance within a cell only the distance of the user to the base station is important, which is why we divide the cell - both reference and neighbour - in K zones. The Markov model of such system is K-dimensional and, based on expressions (3.6), for the transitions in the Markov model we can write:

$$\begin{aligned} &(n_1, \cdots, n_i, \cdots, n_K) \to (n_1, \cdots, n_i + 1, \cdots, n_K) \qquad \text{at rate } \lambda_i, \\ &(n_1, \cdots, n_i, \cdots, n_K) \to (n_1, \cdots, n_i - 1, \cdots, n_K) \quad \text{at rate } \frac{n_i R_i(\underline{n})}{F}, \end{aligned} \tag{5.6}$$

where $i = 1, ..., K$ indicates the zone number. Note that the Markov model of the reference cell is indeed K-dimensional while in the model of the neighbour cell is more complex due to the additional division in segments. We will comment on that in more detail in Section 5.5.1.

The Markov models of the RC for all scheduling schemes discussed in Section 5.2.2 but the ExOBO scheme appear to be a multi-class M/M/1 processor sharing (PS) model. The basics of this model were given in Section 3.5.1. Since the steady-state distribution of a multi-class M/M/1 PS model is readily available, e.g. [26], we are able to provide closed-form expressions for performance parameters such as mean flow transfer times, see Equation (3.9). In our case of a relay-enabled scheduling scheme we can rewrite Equation (3.9) as:

$$T_i = \frac{F}{r_{eff,i}} \cdot \frac{1}{1-\rho}, \tag{5.7}$$

where $\rho = \sum \rho_i$ is the system load, $\rho_i = \lambda_i F / r_{eff,i}$ is the load in zone i, $i = 1, \cdots, K$, and F the mean flow size.

Knowing the mean flow transfer time T_i and the file size F we can derive the average flow throughput D_i, defined as F/T_i, as:

$$D_i = r_{eff,i} \cdot (1-\rho). \tag{5.8}$$

For the ExOBO scheduler the Markov model has a more complex form due to the different state-dependent throughput $R_i(\underline{n})$ for direct and relay path. In this case we have chosen to simulate the Markov chain in order to derive the flow-level performance parameters. The model is simulated according to the approach described in Section 3.6.3.

5.4.6 Energy consumption

Generally there are various factors which determine the energy consumption, e.g. energy cost of reception. In mobile networks however the dominant factor is the transmit power. Hence, the energy consumption for a single data transmission (from mobile to base station independently of the selected path) from zone i is denoted E_i and is defined by the applied transmit power and the duration of the transmission opportunity τ. Hence, when the direct path is chosen we can calculate $E_i = P_i^{tx} \tau_{ms}$ where $\tau_{ms} = 2$ ms (one TTI). Since using a relay requires retransmission of data we can expect higher energy consumption E_i compared to direct transmission. In the case of a relay path we need to take into account the transmission of the mobile station and the transmission of the relay, resulting in $E_i = P_i^{tx} \tau_{mr} + P_{rs,i}^{tx} \tau_{rs}$. The transmit power of the MS (P_i^{tx}) is typically different from the transmit power of the relay ($P_{rs,i}^{tx}$) due to different path length and thus path loss.

Although useful E_i is an absolute measure and does not account for the fact that different schemes might transfer different amounts of data within the transmission opportunity τ. To compensate we use the energy-per-bit E_i^{bit} measure, which incorporates the details on data traffic and is given by:

$$E_i^{bit} = \frac{E_i}{r_{mr,i}\tau_{mr}} = \frac{E_i}{r_{rs,i}\tau_{rs}}. \tag{5.9}$$

The product $\tau_{mr}r_{mr,i}$ gives the amount of data transferred during the period τ_{mr} on the relay path, which is a scheduler-specific value.

Evaluation of the energy consumption in a relay-enabled network is interesting since one of the expected effects of relaying is increased energy transfer.

5.5 Analysis - multi-cell scenario

We continue with the Multi-cell scenario, which allows us to investigate the inter-cell interference in the case of a relay-enabled scheduler. More specifically, we propose approaches towards examining the inter-cell interference I_{oc} generated by the NC, see Section 5.5.1, and how it can be accounted for at the RC, see Section 5.5.2. Additionally, we comment on how the performance of mobile stations in the presence of inter-cell interference can be evaluated, see Section 5.5.3.

The analysis of inter-cell interference with a relay-enabled scheduler is quite complex. On the one hand, we need to remember that each relay transmission is characterised by two distinct inter-cell interference levels - one is generated by the mobile when sending to its corresponding relay and another by the relay when sending to the base station. On the other hand, the inter-cell interference experienced by the base station and the relay station in the reference cell differs as well due to different distances to the source of the interference. The analysis of the inter-cell interference process at the BS and the RS is conceptually the same and therefore we focus on describing the analysis for the base station. However, both processes are considered in this chapter.

5.5.1 Inter-cell interference process

In order to characterise the probability distribution of the inter-cell interference process we need information on the possible interference values I_{oc} and with what probability they occur. Note that the sequence in which the particular values occur is not considered here. The range of possible I_{oc} values depends

on the modelling granularity, i.e. the number of segments, and on the station generating it. The inter-cell interference from a user in segment (i, j) in the NC, denoted $I^{ms}_{oc,ij}$, can be calculated by modifying Equation (5.1) as follows:

$$I^{ms}_{oc,ij} = \frac{P^{tx}}{L(d_{ij})}, \tag{5.10}$$

where P^{tx} is given by Equation (3.2).

The inter-cell interference from a relay station can be determined in a similar manner:

$$I^{rs}_{oc,ij} = \frac{P^{tx}}{L(d^{rs}_{ij})}, \tag{5.11}$$

where d^{rs}_{ij} denotes the distance of the relay in the NC to the base station in the RC.

The probability distribution of the interference processes at the BS can be analytically computed under the assumption that the neighbour cell behaves independently of the reference cell. In fact, we abstract from the inter-cell interference that the RC generates toward the NC[2]. In such interference-free environment the performance in the neighbour cell, e.g. effective rate r_{eff}, depends only on known parameters such as received powers, transmission opportunities τ and number of active users n, see Table 5.2. In such case the probability of a particular inter-cell interference level $I_{oc,ij}$ by a user in segment (i, j) is determined by the fraction of time that the user actually transmits, i.e. τ_{mr}, and can be derived as:

$$\rho^{mr}_{ij} = \frac{\lambda_{ij} F}{r_{mr,ij}}, \tag{5.12}$$

where $r_{mr,ij}$ is the instantaneous rate of a mobile in segment (i, j) and is given by Equation (5.2).

Accordingly, the probability with which an interference level $I^{rs}_{oc,ij}$ exhibits depends on the transmission opportunity τ_{rs} during which the relay transmits toward the base station. These probability can be then expressed as:

$$\rho^{rs}_{ij} = \frac{\lambda_{ij} F}{r_{rs,ij}}, \tag{5.13}$$

where $r_{rs,ij}$ is the instantaneous rate realised by the relay station, depending on the particular TTI division applied by the studied scheduling scheme. Note

[2]In practice this assumption does not hold but based on out observations on the mutual impact of neighbour cells from Section 4.4.5 we believe that this abstraction is acceptable.

that since we assume a circle of relays in fact $\rho_{ij}^{rs} \neq 0$ only for the segments belonging to the zone where the relay are placed.

Finally, the probability of zero interference equals the probability of the system, i.e. NC, being empty. Hence, for the analytical approach we can write:

$$I_{oc} = \begin{cases} 0 & \text{with probability } 1 - \sum_i \sum_j (\rho_{ij}^{mr} + \rho_{ij}^{rs}), \\ I_{oc,ij}^{ms} & \text{with probability } \rho_{ij}^{mr}, \\ I_{oc,ij}^{rs} & \text{with probability } \rho_{ij}^{rs}. \end{cases} \tag{5.14}$$

Another possible approach to determine the probability distribution of the inter-cell interference is by simulating the behaviour of the neighbour cell. In fact, we set up a Markov model of the NC and simulate the actual state transitions. For each state we register the exhibited interference values I_{oc} and the time spent in that state. Accumulating the state times for each unique I_{oc} value and dividing by the total simulation time we can derive the probability of a specific I_{oc} value.

5.5.2 Interference modelling in reference cell

The performance in the reference cell depends among others also on the inter-cell interference, see Equation (5.2). In order to take into account inter-cell interference we introduce the reservation B_{oc} at the receiver. Note that B_{oc} has to be introduced at both BS and RS and that the inter-cell interference at each of them is different. Hence, the available data budget B_D also differs at the BS and the RS. Since the interference varies over time the reservation B_{oc} has to be dynamically adapted in order to capture these variations.

The selection of the reservation B_{oc} is an important issue. Hypothetically, B_{oc} should be adapted at each change in the I_{oc}, i.e. $B_{oc} = I_{oc}$, leading to recalculation of the data rates, see Equation (5.2). However, given a system state $\underline{n}$, the interference changes at a frequency of one TTI, i.e. 2 ms, and such fast changes are impossible to account for in a real system. It is more feasible to adapt B_{oc} only at state changes in the NC while for the duration of a state $\underline{n}$ the reservation stays at a fixed value. We propose two strategies towards the selection of the B_{oc} value. In the first strategy the maximum inter-cell interference $I_{oc,max}$ that is generated during a state $\underline{n}$ in the NC is used yielding $B_{oc} = I_{oc,max}$. In the second strategy we take the average interference $I_{oc,av}$ appropriately weighted over all values generated during a state $\underline{n}$, i.e. $B_{oc} = I_{oc,av}$.

Table 5.3: Relation between zone numbers and distances d_{ms}.

Zone number	1	2	3	4	5	6	7	8	9	10
Distance, km	0.63	0.89	1.09	1.26	1.41	1.55	1.67	1.79	10.9	2.00

5.5.3 Performance in reference cell

For the performance evaluation of mobile stations in the RC the same analysis methodology as for the Single-cell scenario can be applied with a minor modification - in the calculation of the instantaneous rates, see Equation (5.2), the inter-cell interference has a non-zero value and can be represented by B_{oc}. All other calculations continue to be applicable.

Subsequently, the combined analysis of reference and neighbour cell results in a Markov model of the two-cells system with $(K + K * S)$ dimensions describing the behaviour of the system at flow level. The first K dimensions correspond to the division of the RC into K zones, while the second term represents the NC, i.e. $K * S$ segments. Due to the dependability of the backward transition rates on the inter-cell interference I_{oc}, see Equation (5.6) and (5.4), the form of the Markov model is quite complex. Therefore, we have chosen a simulation approach to determine its steady-state distribution. After we derive the steady-state distribution of the Markov chain we are able to determine the mean flow transfer time for a user in any zone of the RC, see Section 5.4.

5.6 Numerical results - single cell scenario

In this section we present a quantitative evaluation of the performance at both packet and flow level for the discussed scheduling schemes in Section 5.2.2. We are particularly interested in performance parameters such as received power, energy consumption, effective rate and mean flow transfer time. Special attention is given to the impact of flow-level dynamics. Most results are generated by mathematical analysis with only the flow-level performance of ExOBO being determined by applying simulation of the corresponding Markov model.

5.6.1 Parameter settings

In the numerical experiments we apply a system chip rate r_{chip} of 3840 kchips/s, thermal noise level N of -105.66 dBm and noise rise target η of 6 dB, see [29]. For SOBOid and SoptOBOid the available budget B at the RS is unlimited. Self-interference of 10% is considered, i.e. $\omega = 0.9$.

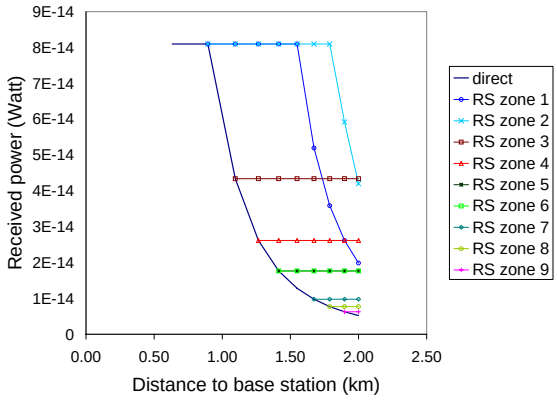


Figure 5.5: Received powers at the base station for a scheduling scheme without relay, e.g. OBO, and a relay-enabled scheme, e.g. SOBO and ExOBO.

A single cell with radius of 2 km is divided in $K = 10$ zones with equal area, see Figure 5.2. The relay path is selected by all MSs that are located farther than the relay. We will later comment on how appropriate this selection criterion is. Given the equal zone area we can easily calculate the zone radii, see Table 5.3. In the rest of the chapter for simplicity we talk about zone numbers and each zone is indicated by a marker at the graphs.

By default relay stations are located at 1.09 km (zone 3) from the BS and we explicitly state when this changes. Both MS and RS have maximum transmit power $P_{\max}^{tx} = 0.125$ Watt. EUL flows are taken with an E_b/N_0 target of 5 dB, see [29], and mean file size is $F = 1000$ kbit. The flow arrival rate is set to $\lambda = 0.4$ flows/sec.

5.6.2 Received power levels

One of the parameters affected by the relay is the received power due to the shorter communication path to and from the relay (and the consequently lower path loss). Evaluating the improvement in received power provides us with an indication of the benefits of relaying abstracting from the impact of the scheduling scheme. In Figure 5.5 we present a comparison between the performance of the OBO and SOBO (and ExOBO) schemes. The received power at the base

station is plotted as a function of the distance and, in the case of the relay-enabled schemes, of the position of the relay. Note that the SOBO and ExOBO schemes have the same received power P^{rx} characterising the relay path MS-RS-BS, since both schemes use the power of the bottleneck sub-path (relay and mobile), see Table 5.1.

Figure 5.5 confirms that, as expected, relaying increases received powers but the position of the RS has strong impact on the gain. Moving the RS towards the edge of the cell lowers the gain - an effect caused by increased distance d_{rs} between RS and BS and fewer MSs using the relay.

Observe the flat part of the graphs in Figure 5.5. Recall that transmission on the relay path can be limited by both relay or mobile station. When the mobile is the limitation (due to its constrained transmit power) the received power decreases with distance as we can see from the graph of direct transmission, i.e. OBO. However, when the RS forms the bottleneck, all MSs served by the relay are characterised by the same received powers at the BS which explains the constant performance after the relay is switched. When the limit changes from relay to mobile again depends on the position of the relay relative to the base station.

In our observation for a relay in zone 3 we register an improvement for the largest number of zones. Therefore, in the rest of the numerical evaluations we have chosen to work with a relay station in zone 3.

5.6.3 Effective data rates

The discussion on effective rates is split into two parts. First, we compare the OBO, ExOBO, SOBO, SoptOBO and FixedRS schemes since they all apply the same assumption of finite budget at the base and relay station. We refer to these schedulers as *practical set*. Second, the effects of eliminating the relay as a bottleneck are evaluated by comparing the SOBO, SOBOid, SoptOBO and SoptOBOid. These set of schedulers is referred to as *ideal set*.

Figure 5.6 presents a comparison of the effective rates for the practical set. The relay station is fixed at zone 3 and fixed path selection policy is applied, i.e. all mobile stations from zone 4 to 10 make use of relaying. In all cases when direct transmission is used all schemes show the same performance. However, beyond zone 3 (after the relay is used) the relay-enabled schemes outperform OBO, meeting our expectations. Only for SOBO and ExOBO the rates realised by users in zone 4 are actually lower when compared to OBO. This result illustrates very well the negative impact of relaying on transmission time due to data forwarding. Hence, we can conclude that relaying is generally effective

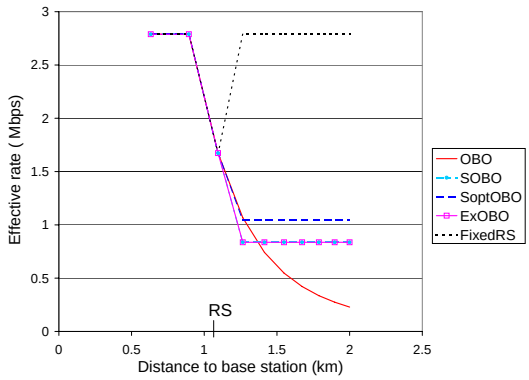


Figure 5.6: Effective data rates reazlied by a non-relay OBO scheduler and by relay-enabled schemes: SOBO, SoptOBO, ExOBO and FixedRS.

but should be selected only when the indirect path delivers higher effective, and not instantaneous, rate than the direct one.

The graphs of SOBO and ExOBO coincide, since both schemes apply the same received powers at BS and MSs use only half the transmission time τ, see Equation (5.2). Furthermore, complying with our expectations, the two schemes are outperformed by SoptOBO. The flexible adaptation of the transmission times τ_{mr} and τ_{rs} supports the best possible utilisation of the available budget. The FixedRS scheme offers further improvement since it in fact dedicates the whole TTI to the user and the capacity of the RS-BS link is much higher than the radio interface link. Actually, the higher effective rates for FixedRS, after the relay switches in, imply that even mobile stations in front of the relay might benefit from its use. We explain that with the fact that users from zone 3 lack capacity to fully utilise the budget at the BS but can fill up the budget at the RS. Since the fixed link does not form a bottleneck these users, if sending via the relay, are also able to fill up the budget at the BS.

Furthermore, in OBO the rate degrades with distance due to the lower received power and consequently lower data rate, see Equation (5.2). Alternatively, given a relay in zone 3, all transmissions on the relay path in SOBO, SoptOBO, ExOBO and FixedRS are limited by the relay, which explains the flat graphs.

The effective rates for the ideal set of schedulers are presented in Figure 5.7. As we can expect, the ideal version of each scheme (SOBOid and SoptOBOid)

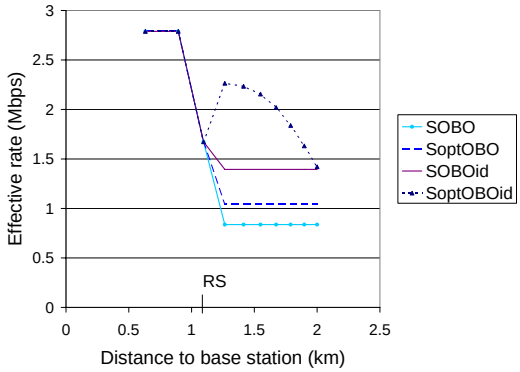


Figure 5.7: Comparison between a realistic and idealistic scheduling realisation, i.e. SOBO vs SOBOid and SoptOBO vs SoptOBOid.

outperforms its practical implementation (SOBO and SoptOBO respectively). By not limiting the transmit power of the relay we ensure that it is capable to always fully utilise the budget at the base station. The consequence for the SOBO scheme is that all stations, which use the relay, can realise the maximum possible effective rate. Note that due to the data forwarding this maximum possible effective rate is twice smaller than the maximum possible instantaneous rate, which a user on the direct path realises.

The performance of the SOBOid scheme is still limited by the inefficient TTI division - an issue that is avoided by the SoptOBOid scheme. This more efficient time scheduling, combined with the unrestricted budget at the relay, suggests that a mobile station is limited in performance only by its transmit power and distance to the base station. In this case users from different zones will experience different path loss, which explains the decreasing in distance graph of the SoptOBOid scheme. Also note that SoptOBOid in fact registers better performance for the users using the relay than the users in zone 3, which select the direct path. It seems that an efficient TTI division among the mobile and relay transmission can extend the benefit of relaying also to users located between the relay and the base station, which once more confirms our conclusion for the practical set.

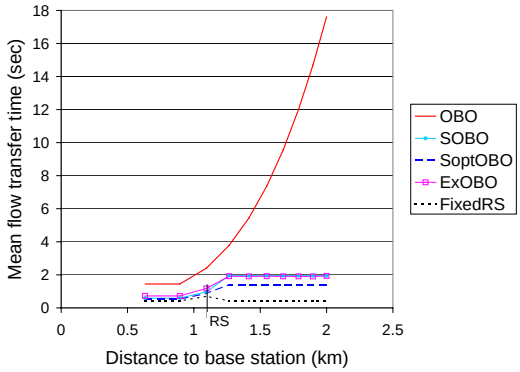


Figure 5.8: Mean flow transfer times reazlied by a non-relay OBO scheduler and by relay-enabled schemes: SOBO, SoptOBO, ExOBO and FixedRS.

5.6.4 Mean flow transfer times

Figure 5.8 presents the dependency of the mean flow transfer times on the distance to the base station for the practical set (SOBO, SoptOBO, ExOBO and FixedRS and the reference OBO scheduler). The observations made for the effective rates in Section 5.6.3 still hold: (i) OBO is the worst performing scheme; (ii) flexible TTI division, e.g. SoptOBO, provides benefits over fixed division, e.g. SOBO and ExOBO; and (iii) the higher capacity of the fixed line in FixedRS puts it as the best performing scheme.

New trends in the performance exhibit as well. In particular, Figure 5.8 shows that the gain achieved by using RSs is even more emphasised when we consider the mean flow transfer times as a performance measure. An interesting new observation is that even the performance of close-by MSs, which use direct transmissions, is improved significantly as result of relaying. Higher data rates for remote MSs translate to lower system load (number of active MSs n), which is beneficial for all active MSs, see Equation (5.4), independently whether they use a relay or not.

Again a comparison with idealistic transmission conditions at the RS is presented, see Figure 5.9. The same trends as in the one for the effective rates can be made. The combined effect of flexible TTI division and distance on performance continues to be observed in the case of SoptOBOid. Also here the

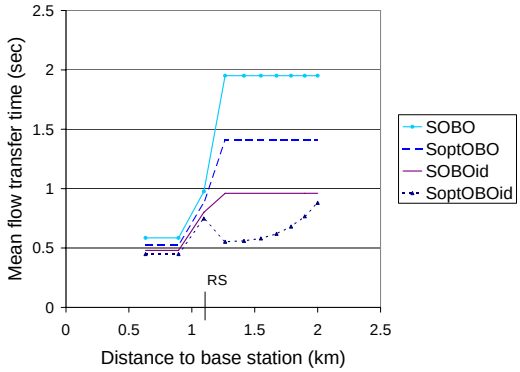


Figure 5.9: Comparison between a realistic and idealistic scheduling realization, i.e. SOBO vs SOBOid and SoptOBO vs SoptOBOid.

difference in performance is not only for MSs which use the relay but also for such stations which transmit via the direct path.

5.6.5 Energy consumption

Figure 5.10 shows the energy-per-bit consumption as a function of distance. All relay-enabled schemes record lower E_i^{bit} than OBO due to their lower applied powers and higher achievable data rates. SoptOBO has the best performance since its optimal TTI division maximises data rates. Energy consumption increases with distance as a result of increasing transmit powers of the MSs. The effect of distance is largest for SoptOBO due to the flexible transmission times, which imply that stations with high transmit powers send for a long period, see Section 5.2.2. On the contrary, in SOBO and ExOBO the transmission times are fixed and the increase in energy consumption smoother.

5.6.6 Concluding remarks

In this section we showed that relaying with appropriate scheduling increases performance on both packet and flow level. Purely based on the numerical results we could conclude that FixedRS, SOBOid and SoptOBOid are the best performing schemes. Such conclusion is however misleading and we need to

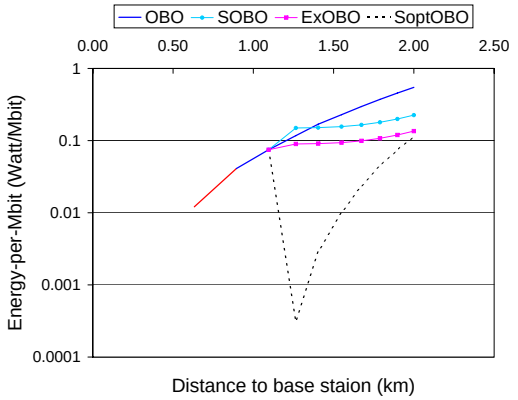


Figure 5.10: Energy per bit consumption.

account for other aspects related to the schemes such as practical feasibility. The idealised schemes SOBOid and SoptOBOid are, as we discussed already in Section 5.2.2, analysed in order to show the theoretical upper bound on the performance benefits provided by relaying. In practice the two assumptions of unlimited transmit power and available budget at the relay are hard to achieve. Although we could try to 'fake' the former by considerably increasing the transmit power of a relay above the power of a mobile, the latter restriction (of limited budget) can hardly be overcome.

The FixedRS scheme is practically feasible and it indeed shows the best performance. However, it also rises many other implementation related questions, which need to be studied before a general conclusion on performance and applicability can be made. Revising our conclusion on the best performing relay-enabled scheme we see that in fact SoptOBO is the best scheduling strategy.

Given the feasibility issues arising for FixedRS, SOBOid and SoptOBOid, in the rest of this chapter, we focus on the SOBO and SoptOBO. ExOBO is not discussed further because it showed to behave almost identically to SOBO.

5.7 Numerical results - multi-cell scenario

The previous Section 5.6 was dedicated to the impact of relaying in a single cell scenario. However, relaying can also influence inter-cell interference. The

investigation on the relation between relay-enabled scheduler and inter-cell interference is organised into two parts. First, for each scheduler, the inter-cell interference process, generated by a neighbour cell, is investigated. Second, we present an evaluation of the performance in the reference cell at both packet and flow level in terms of effective data rates and flow throughputs respectively. The schemes compared in this section are OBO, SOBO and SoptOBO.

5.7.1 Parameters settings

The majority of parameter settings are the same as indicated in Section 5.6.1. Some modifications are necessary mainly to the modelling of inter-cell interference. In Section 5.5.1 we introduced two possible approaches to characterise the inter-cell interference - analytical and by simulation. For the analytical approach we divide the neighbour cell into $K = 200$ zones and $S = 360$ sectors. For the simulation approach less fine division is used, i.e. $K = 20$ zones and $S = 36$ sectors, in order to keep evaluation time feasible. Recall that cell radius is 2 km and the relay is at 1.09 km from the base station, i.e. $d_{rs} = 1.09$ km.

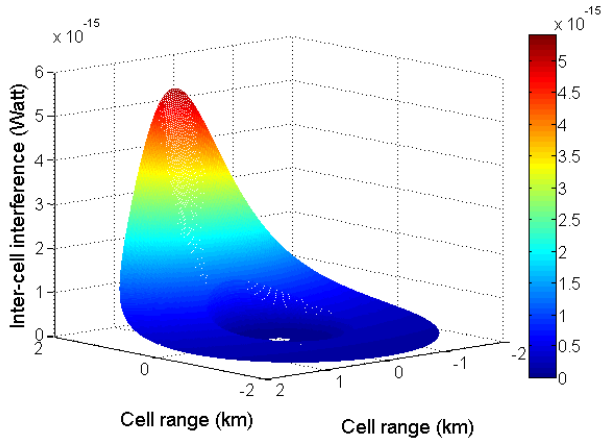
Furthermore, self interference is not considered, i.e. $\omega = 1$ and the flow arrival rate λ is set to 0.5 flows/sec for both reference and neighbour cells, which leads to cell load of $\rho = 0.82$ in the case of OBO scheduler. Note that the actual load for a relay-enabled scheduler is lower due to different (higher) effective data rates, see Section 5.4.5.

5.7.2 Inter-cell interference process

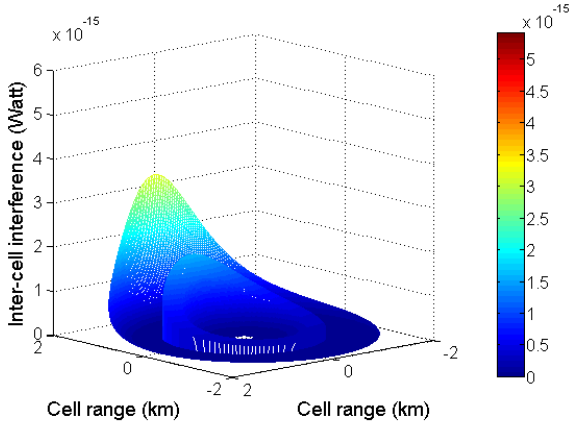
The investigation on inter-cell interference starts with discussion on the cumulative distribution functions (CDFs) and continues next with comparison of the two different approaches to characterise the interference process.

Inter-cell interference levels

The contribution of a particular MS (or RS) in the NC to the I_{oc} at the BS of the RC for OBO and SOBO is presented in Figures 5.11(a) and 5.11(b) respectively. The increase in the graphs corresponds to the area in the NC that faces the RC. The figure shows that an SOBO scheme generates lower maximum inter-cell interference, i.e. $I_{oc} = 3.2 * 10^{-15}$ Watt, than a OBO scheme, i.e. $I_{oc} = 5.2 * 10^{-15}$ Watt. The maximum interference level is generated by a MS at the cell edge of the NC closest to the RC. Such MSs are relatively far from



(a) OBO scheduler



(b) SOBO scheduler

Figure 5.11: Spatial representation for (a) OBO and (b) SOBO of the inter-cell interference at the BS of the RC that a MS (or a RS) generates depending on its location in the NC.

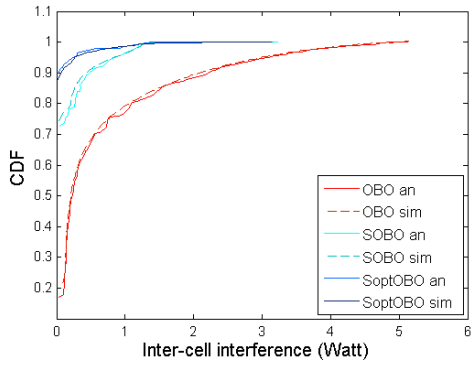


Figure 5.12: CDFs of the inter-cell interference experienced at the BS for the various scheduling schemes accounting for the interference from MSs and RSs in the NC.

their serving base station in the NC and need to transmit at higher powers than users close to the cell centre.

In Figure 5.11(b) when moving away from the centre of the cell a sudden drop in the inter-cell interference is observed. At that moment relaying becomes beneficial and a MS needs lower transmit power for successful reception thus leading to lower inter-cell interference. Subsequently as the distance between MS and RS increases the inter-cell interference is again on the rise.

The unique CDF graphs of the inter-cell interference process for each of the scheduling schemes is presented in Figure 5.12 with solid lines. The graphs are generated as result of analytical calculations. The maximum observed I_{oc} for the SOBO and SoptOBO schemes is lower than the value for OBO because MSs located at the cell edge transmit through the relay and use lower transmit power. For the same reason, in the relay-enabled schemes, low interference values are more probable to occur. We have taken into account also the empty system state in which no interference is generated. Observe that the probability for zero interference for the OBO scheduler is more than three times lower compared to SOBO and SoptOBO. This is due to the fact that MSs under relay-enabled schedulers have higher transfer rates and lower transmission times. Further, the difference in the CDFs of SOBO and SoptOBO is hardly observable.

The specific impact of the relay stations on the inter-cell interference can be also observed in Figure 5.12. The CDF for the SOBO and SoptOBO sche-

dulers does not change smoothly. The part of the graph enclosed by the inflex points reflects the impact of the relay station. On the one hand, for each relay transmission actually two interference levels exhibit - the interference from the mobile and the interference from the relay. On the other hand, each relay serves several mobile stations and therefore its interference level $I_{oc,rs}$ appears more often.

Process modelling

We will now compare the inter-cell interference patterns generated by two independent approaches - analytical and simulation. For the ease of notation we will refer to the CDF generated via analytical computation as *CDF-an* and to the CDF generated during simulation as *CDF-sim*.

The CDFs for both approaches and for each of the three schedulers are presented in Figure 5.12; CDF-an is given by solid lines and CDF-sim by dashed. The general impression is that the graphs of the CDF-an and CDF-sim, taken per scheduler, lay close to each other; the differences are explained with the lower modelling granularity of the simulation approach. The maximum values of the inter-cell interference registered by the CDF-an and CDF-sim coincide as well. The lower granularity of the simulation approach also explains the discrete step-like form of the 'simulation' curves in Figure 5.12, which effect is best observed for the OBO scheduler.

We could conclude then that the choice of generation approach is irrelevant. The analytical approach however exhibits several advantages. First, the analytical approach is faster since it requires only simple computations. This is an interesting observation since in the previous chapters 3 and 4 simulation was the more time-efficient approach. Second, because it is faster, the analytical approach allows modelling of the NC in finer granularity, e.g. division in larger number of segments. For example, an analytical computation of the probability distribution was about ten times faster than the simulation approach, given the same hardware settings were used, while it allowed hundred times finer modelling granularity, i.e. $200 * 360 = 72000$ segments for the analytical approach versus $20 * 36 = 720$ segments for the simulation.

5.7.3 Performance of mobile stations

While in Section 5.6 we focused on the benefits of relaying for performance in a single cell scenario, in this section we concentrate on how relaying affects inter-cell interference and on the consequences of that. We expect the trends

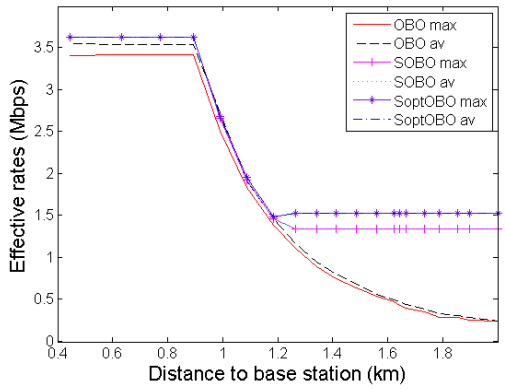


Figure 5.13: Performance evaluation on the packet level in terms of effective rates.

as observed in Section 5.6 to persist, but the performance to be worsened since inter-cell interference occupies part of the available budget. First, effective data rates are considered followed by a discussion on flow throughputs.

Effective rate

Figure 5.13 presents the results for the effective rate of the three schedulers OBO, SOBO and SoptOBO for the two B_{oc} reservation strategies - at the maximum, i.e. $B_{oc} = I_{oc,max}$, and average, i.e. $B_{oc} = I_{oc,av}$, values of the inter-cell interference during a transmission cycle. As we expected relaying increases realised data rates. More importantly, relay-enabled schedulers register better performance even for MSs using the direct path, i.e. distance ranges up to 1 km in Figure 5.13. This is a consequence of the lower I_{oc} that a relay-enabled cell generates, see Section 5.7.2 and Equation (5.2). Furthermore, SoptOBO outperforms the SOBO scheme due to its better flexibility in channel assignment as a function of MS's transmit capacity.

Several other general observations can be made. The first (higher) flat section of the graphs in Figure 5.13 corresponds to MSs close enough to the BS such that they can fill up the budget B_D on their own, i.e. $P_{ms}^{rx} = B_D$. With the chosen parameter settings, the maximum distance for which $P_{ms}^{rx} = B_D$ holds is 0.9 km, as Figure 5.13 indicates as well. The second (lower) flat section for

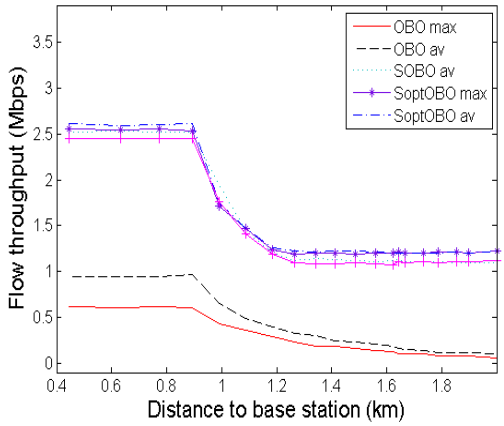


Figure 5.14: Performance evaluation on the flow level in terms of flow throughput.

the graphs of SOBO and SoptOBO is determined by the fact that for all MSs using the indirect path the relay forms the bottleneck.

Flow throughput

Figure 5.14 shows a comparison between the achieved flow throughput D_i for OBO, SOBO and SoptOBO scheme. Again SoptOBO shows the best performance, followed by SOBO and OBO with the worst performance. All qualitative observations made for the effective data rates hold for the flow throughputs as well.

However, on the flow level, the big performance difference between OBO and the relay-enabled schemes becomes even more prominent, especially for MSs which do not use the relay. The faster service offered by the relay to remote MSs translates to lower average number of active flows from which all MSs, independently of location, benefit, see Equation (5.4). These results strongly support our believe that evaluation at flow level is crucial for understanding the complex effects of relaying.

Impact of B_{oc} reservation

We now focus on the impact of the B_{oc} reservation choice, shown in Figure 5.13 and 5.14. For the OBO scheme performance differs significantly - 3.55 Mbps for the average value $I_{oc,av}$ against the lower 3.41 Mbps with reservation at $I_{oc,max}$. On the contrary, SOBO and SoptOBO show negligible change in performance.

Note that the choice of reservation is more expressed for mobile stations that use the direct path. These stations have higher received power capacity compared to remote users due to their closer position to the base station. Hence, they are more capable to use any extra budget that might be available. At the same time a reservation at the average of the inter-cell interference leaves more budget to EUL data users than a reservation at the maximum level does.

The results on both effective rates and flow throughput show that, contrary to an OBO scheme, the SOBO and SoptOBO schemes are not very sensitive to the B_{oc} reservation strategy. This lack of sensitivity makes relay-enabled schemes attractive also from practical aspects, e.g. determining a maximum inter-cell interference value requires constant monitoring of the environment and is generally difficult to achieve; it is much easier to calculate an average interference value over a periodically activated measurement period.

An OBO scheme however, in consistence with our previous observations from Section 4.5.4, shows big differences in performance depending on the applied reservation strategy. Therefore, selecting a percentile of the I_{oc} distribution to serve as B_{oc} might prove better, as we also showed to be the case in Section 4.5.4. Unfortunately such approach is more challenging to implement since it requires detailed knowledge of the inter-cell interference in real time.

5.8 Conclusions

This chapter presented an initial evaluation of the uplink performance in a mobile network with relaying. Our interest was concentrated mainly on: (i) relay-enabled packet scheduling; (ii) effects of relaying on mobile station performance; and (iii) effects of relaying on inter-cell interference. Since this study was meant to be more of an exploration than an exhaustive evaluation, we made several modelling assumptions which simplify analysis but are expected to keep the main system characteristics.

We proposed several possible 'relay-enabled' scheduling schemes which incorporate relaying out of which we could recommend few as having potential for practical deployment. The performance of mobile stations was examined for

situations without and with inter-cell interference based on performance measures such as achieved data rates, flow throughout and energy consumption. Additionally, we provided a comprehensive methodology to model the inter-cell interference process.

As expected, the results indicated that relaying successfully increases both achievable data rates and flow throughputs. Interestingly, users who do not use the relay may also considerably gain from relaying, which trend becomes clearly expressed when flow throughputs are compared. Therefore we argue that analysis on the flow level is crucial for proper evaluation and should be performed before conclusions are drawn. Our other expectation that inter-cell interference decreases was also confirmed, which only adds to the attractiveness of relaying for practical deployment.

6 Impact of relay characteristics on uplink performance

6.1 Introduction

In Chapter 5 we presented an exploration study on the role of relaying for improving uplink performance in a cellular network. The main purpose of the study was to establish what the potential performance gains from relaying are and what the impact of the scheduling scheme is. Our investigations were based on the 'best-case' assumption that, independent of their precise location, mobile stations near the cell edge can always use a relay station that lies on a straight line between the mobile station and the base station. This assumption eases the analysis, but its practical implication is a large number of relays, which is financially not profitable and challenging for deployment. Therefore, we decided to undertake a new, more practically oriented approach towards the modelling of a relay-enabled mobile network.

In the current chapter we concentrate on the situation with a single relay station at an apriori determined position in the cell, which uses the same radio interface for transmissions as mobile stations. We are especially interested in the combined effects of relay deployment and packet scheduling on the performance observed by mobile stations. In particular, we investigate the impact of certain relay characteristics such as transmit power and position and determine the area over which relaying offers performance gains.

6.1.1 Literature

In the literature studies can be found that focus on a specific deployment situations by including many deployment and technical details. Although insightful and providing useful feedback for development, such studies are very tightly related to the particular practical situation that is discussed. Therefore their results are difficult to interpreted for a broader evaluation range. One such study is [61], which gives concrete values for performance gains of relaying for several deployment scenarios. The authors of [59] take the research on relaying one step further by evaluating the performance of a relay-enabled network of a small city. Using topology information and retracing [59] evaluates capacity and coverage gains. Also practically oriented but for indoor situations is [72], in which, in the context of OFDMA networks, no relay, single relay and multiple relay scenarios are compared in performance.

Investigations on the service area of a relay station are more difficult to find and tend to focus more often on downlink. One such research, see [31], takes the introduction of a relay to HSDPA but unfortunately the authors limit the research to link level analysis concentrating on throughput gains. It is interesting that, even if performed for downlink, the study reveals a shape of the service area, which is similar to the one that our results will later show.

The work we present in the current chapter is related to [54] which discusses similar radio-interface relay configurations for LTE downlink. In particular, the two main research questions in [54] are where to position relays and how many to deploy. Similar to us the authors consider a scheduler whose parameters adapt according to the user specific channel conditions. The paper however keeps the number of users fixed where each user constantly has date to transmit, i.e. persistent flow are modelled. Although insightful, [54] is purely based on simulation results and does not provide any analytical expressions to characterise performance.

6.1.2 Contributions

This chapter presents our findings on several design, deployment and performance issues which in our opinion are missing or not well studied in the existing literature. These issues arise during the incorporation of the relaying concept in a realistic setting. In particular, we focus on the role of packet scheduling and evaluate two relay-enabled scheduling schemes by comparing them to a conventional scheme that does not use a relay. We distinguish ourselves from other studies on the topic by: (i) including flow dynamics in the analysis; (ii) investi-

gating the combined effects of packet scheduling and relaying on performance; and (iii) evaluating the impact of several characteristics of a relay station on the mobile station's performance.

As main contributions of our study we identify:

- Determining the service area of a relay station. The *service area* (SA) of a relay station can be defined as the aggregate of all locations within a cell which register better performance, i.e. higher data rates, when transmitting via the relay. The shape and the size of the service area have crucial importance for deployment decisions such as how many relay stations need to be set up, at what locations, etc. It can be expected that the service area is scheduler-specific since each scheme has a different approach towards organising relay transmissions. We made it our goal, on the one side, to provide means to derive the SA of a relay and, on the other side, to compare the SAs of different schedulers.

- Evaluating the performance impact of certain characteristics of a relay station such as location and transmitted power. Having a good understanding of these effects is important for the identification of an optimal deployment scenario. Each characteristic has a different influence on performance, which requires its independent evaluation. Eventually, we can identify the relay characteristics relevant for performance and determine what their optimal settings are.

- Quantifying the benefits of relaying in terms of realised data rates, which give direct gains, but also flow throughputs, which reflect the more long-term gains. Incorporating flow dynamics, i.e. varying number of ongoing flows, into the analysis allows us to detect certain performance trends which could be otherwise left unobserved. One example is the improved service for users that actually do not directly use the relay.

6.1.3 Outline

The structure of this chapter is as follows. In Section 6.2, we briefly describe the scheduling schemes which we evaluate and which were already presented in detail in Chapter 5. Next in Section 6.3 the network scenario and accompanying assumptions are presented, which form the base for the analysis presented in Section 6.4. The analytical derivation of a service area of a relay station is described in Section 6.5. Numerical results of our analysis are given in Section 6.6. Finally, a summary of the work is presented in Section 6.7.

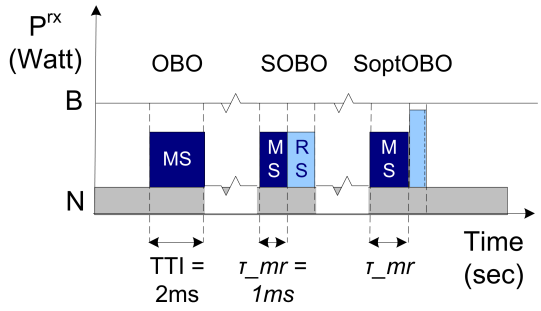


Figure 6.1: Scheduling schemes under study.

6.2 Considered relay-enabled schedulers

Already in Chapter 5 we introduced the modifications that a relay stations brings to a mobile network's infrastructure and the communication path. Here we recall that we distinguish between a direct path - between mobile and base station - and an indirect path - the path mobile - relay - base station. Furthermore, the indirect path consists of two sub-paths - mobile sub-path MS-RS and relay sub-path RS-BS. Each path is denoted by a corresponding index, i.e. *ms* for the direct path, *mr* for the mobile sub-path and *rs* for the relay sub-path. Each path is also characterized by a set of transmission parameters, which we presented in Section 5.2.1.

Of all proposed relay-enabled scheduling schemes in the previous Chapter 5 we now focus only on the shared one-by-one (SOBO) and shared optimal one-by-one (SoptOBO). These schemes showed the highest potential for practical deployment, see Section 5.6.6. Their performance is compared to a reference OBO scheme that does not make use of a relay station. All schemes are of the Round Robin (RR) family in which mobile stations are served in a predefined order, irrespective of their channel conditions. The scheduling schemes are presented in Figure 6.1. Below the essence of each scheme is presented and for more detailed description we refer to Section 5.2.2.

In OBO a MS always transmits directly to the base station independently of its location in the cell. Hence, the user can use the whole single transmission opportunity τ for its transmission, i.e. $\tau_{ms} = \tau$. The transmission opportunity τ is the amount of time that the BS station dedicates to the service of a mobile transmission, independently whether it uses a relay or not. Note that in all

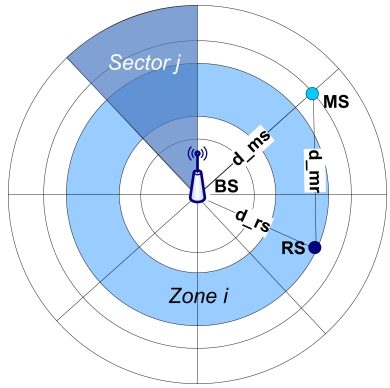


Figure 6.2: Single cell model with cell division in K zones and S sectors.

three schemes, discussed in this chapter the transmission opportunity equals a single TTI, i.e. $\tau = 2$ ms. On the indirect path, the SOBO scheme divides the TTI into two equal intervals of 1 ms and MS and RS both receive one such interval to transmit, i.e. $\tau_{mr} = \tau_{rs} = 1$ ms. The communication on the relay path is limited by the slower sub-path, which situation depends on the user's relative location to relay and base station. As a result the SOBO scheme lacks flexibility to adapt to the user's power capacity and is less than optimal. In the SoptOBO the channel utilisation for the indirect path is optimized by adapting the transmission opportunities τ on the sub-paths such that both mobile and relay sub-paths can transfer the same amount of data, i.e. $\tau_{mr} r_{mr} = \tau_{rs} r_{rs}$.

6.3 Model

Most of the assumptions made in Section 5.3 apply also for the research presented in this chapter. Since we choose to work with a single relay station however the cell model needs to be slightly adapted. We consider a single cell with a single relay station located at distance d_{rs} from the base station. In order to take into account the different distances that a mobile station can have to the relay the cell is divided into S equal-sized *sectors* (characterized by angle) and K equal-sized *zones* (characterized by distance), see Figure 6.2. The resulting $K \times S$ *segments* are in fact all positions at which a mobile can be located. Segments belonging to the same zone are characterised by the same distance d_{ms}

to the base station but have each different distance d_{mr} to the relay station, both measured from the outer edge, $i = 1 \cdots K, j = 1 \cdots S$.

Given the chosen segmentation of the cell, the number of active MSs and their distribution over the segments define the system state $\underline{n} \equiv (n_{11} \cdots n_{KS})$. With K and S sufficiently big we can assume that all MSs belonging to the same segment behave identically. Since we consider the behaviour of mobile users and not individual data packets the arrival process can be modelled correctly by a Poisson process with rate λ. Taking into account the equal area of the segments and a uniform distribution of the users, the arrival rate per segment can be derived as $\lambda_{ij} = \lambda/(KS)$. Flow size is exponentially distributed with mean F.

The service area is the aggregation of all locations (segments) within the cell, from which a MS receives better service if using the relay. Alike in Chapter 5 where a relay transmission is chosen purely based on location, in this chapter a mobile selects the path which provides higher effective data rates, see Section 6.4.2. An instantaneous switching mechanism between transmitting and receiving at the RS is assumed as the same frequency band is used. In our study a relay is a device with one basic functionality, namely resending data from MS to BS. Such simple design keeps the cost down and still provides improved performance.

Furthermore, the assumptions of Sections 5.3.4 and 5.3.5 apply with some slight modifications. The available data budget B_D is derived from the total budget B after subtracting the thermal noise at the receiver (base station or relay station), i.e. $B_D = B - N$. In this study we have intentionally disregarded important factors such as inter- and intra-cell interference in order to focus on the relative effects that a relay has on performance. However, as we showed in Chapter 5, we understand that such factors may have significant effect on the absolute performance gains. The assumptions applicable for the mobile stations such as maximum transmit power P_{max}^{tx} are the same as in Section 5.3.5.

6.4 Analysis

The same generic analysis methodology presented in Chapter 5 is applied meaning we first calculate received power levels from which data rates and later flow throughputs can be derived. Some modifications, mainly in the applicability of the calculations on the indirect path, are necessary in order to reflect the presence of a single relay station.

6.4.1 Received powers

On the direct path MSs located in the same zone but in different sectors experience the same path loss and thus have the same received power $P^{rx}_{ms,i}$ at the base station, which can be calculated according to Equations (5.1). On the indirect path however MSs, even if belonging to the same zone, from different sectors have different distance d_{ij} to the relay and hence differ in the received powers $P^{rx}_{mr,ij}$. These powers can be calculated according to a modified version of Equation (5.1), namely:

$$P^{rx}_{mr,ij} = \frac{P^{tx}}{L(d_{ij})}. \tag{6.1}$$

The path loss model we apply is the Cost 231 Hata propagation model, see Section 5.4.1.

6.4.2 Data rates

In Chapter 5 we introduced the three types of data rates - instantaneous, effective and state-dependent throughput. The *instantaneous data rate* is the rate, which a MS realises when it is transmitting data and hence it is the rate realised for the duration τ_{mr}. The *effective rate* $r_{eff,ij}$ is the rate at which data is transferred from a mobile station in segment ij, via the relay, towards the base station. It is derived from the instantaneous rate and is calculated over the transmission opportunity τ. Note that on the indirect path, as a result of data forwarding, the effective rate is lower than the instantaneous rate. The effective rate is scheduler-dependent. Finally, the *state-dependent throughput* $R_i(\underline{n})$ captures the impact of the number of ongoing flows n on the performance of a single flow. The corresponding expressions for each of the defined data rates are given by Equation (5.2), Table 5.2 and Equation (5.4) respectively. For more detailed explanation on the data rates we refer back to Section 5.4.

Note that in contrast to the analysis in Chapter 5, in the current study, the rates realised by MSs belonging to the same zone but different sectors and sending via the relay differ; this holds for any of the three rates - instantaneous, effective and state-dependent throughput.

6.4.3 Flow dynamics

As we have discussed before, the changing number of ongoing flows, which is a result of flow initiations and completions, needs to be accounted for and we do that by using continuous time Markov chains (CTMC). Given our applied cell

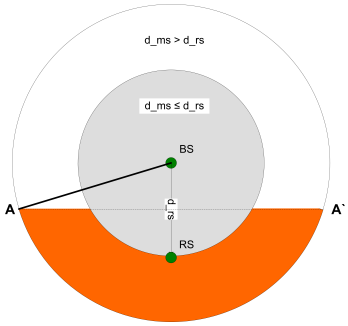


Figure 6.3: Service area (SA) of a single relay station based on received power levels.

model division in $K \times S$ segments, the resulting CTMC is $K \times S$ dimensional. In Section 3.4.4 we discussed in detail how a CTMC can be mapped to model the flow dynamics within a cell while in Section 5.4.5 we presented how the Markov chain modelling can be applied to a relay-enabled cell. In this chapter we only want to recall that the Markov chain of each of the three schedulers - OBO, SOBO and SoptOBO - belongs to a multi-class M/M/1 processor sharing (PS) model. The particular form of the Markov chain is scheduler-specific due to the different form of the state-dependent throughputs. In Section 5.4.5 we presented an expression for the calculation of the mean flow transfer time, see Equation (5.7), which when modified to fit the modelling in segments takes the form:

$$T_{ij} = \frac{F}{r_{eff,ij}(1-\rho)},$$

where ρ is the system load being the sum of the loads in each segment, i.e. $\rho = \sum_{i=1}^{K} \sum_{j=1}^{S} \rho_{ij}$, where $\rho_{ij} = \lambda_{ij} F / r_{eff,ij}$ is the load in a segment ij; $i = 1, \cdots, K$ and $j = 1, \cdots, S$. F is the mean file size.

Subsequently, from the mean flow transfer time T_{ij} we can derive the average flow throughput D_{ij}, defined as F/T_{ij}, as:

$$D_{ij} = r_{eff,ij}(1-\rho). \tag{6.2}$$

6.5 Service area of a relay station - analysis

Our goal in this section is to analytically determine the critical distance d_c from the BS at which a MS changes its preference from direct to indirect path and which determines the boundaries of the service area. We will show that the shape and size of the service area depends on whether received powers or effective data rates are used in the analysis. As we showed in Section 6.4.2, due to data forwarding the gains in effective rates are lower than what received powers 'promise'. Therefore, it is more appropriate to calculate service area based on effective rates. However, working with received powers provides us with a good reference base to illustrate our claim.

6.5.1 Service area based on received powers

In this scenario, a mobile station selects the transmission path - direct or indirect - that can offer higher received power at the base station. Recall that in SOBO the received power for the indirect path is determined by the sub-path with poorer channel conditions. The condition to select indirect transmission can be written as:

$$P_{ms}^{rx} < min(P_{mr}^{rx}, P_{rs}^{rx}). \tag{6.3}$$

From Equation (6.1), assuming the same transmit power for MS and RS, we can deduce that the received power is only dependent on the distance between the communicating stations. Therefore, the condition from Equation (6.3) can be rewritten as:

$$d_{ms} > max(d_{mr}, d_{rs}). \tag{6.4}$$

Figure 6.3 shows a visualisation of the limitations set by Equation (6.4). A circle with radius d_{rs} around the base station can be drawn, within which $d_{ms} \leq d_{rs}$ holds and a MS always selects the direct path; outside it - the indirect path may be preferred. The single relay station in our model is not sufficient to cover the whole disc for which $d_{ms} > d_{rs}$ is valid. However, by applying basic geometry rules for medians we can find the line AA', see Figure 6.3, all points of which have the same distance to the base station as well as to the relay. All MSs 'below' AA' have $d_{ms} > d_{mr}$ and can benefit from relaying. The intersection of the latter inequality and $d_{ms} \leq d_{rs}$ defines the relay's service area, when based solely on received powers. In Figure 6.4 we show the dependency of the critical distance d_c on two user parameters - distance to the base station and transmit power.

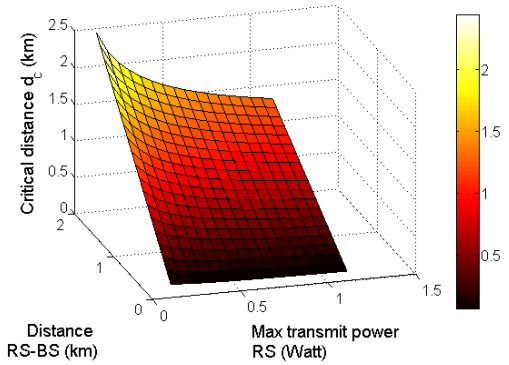


Figure 6.4: Dependency of the critical distance d_c on both distance of relay to the base station and the transmit power of the relay for a SOBO scheme.

6.5.2 Service area based on effective data rates

In order to keep calculations tractable, yet providing useful insights, we consider the scenario when BS, RS and MS lie on a straight line. Both schemes - SOBO and SoptOBO - are discussed as we provide analytical expressions to find the critical distance. We begin with discussion on SOBO.

A MS always selects the path, i.e. direct or indirect, which provides higher effective data rate. The critical distance d_c is then the distance d_{ms} for which direct and indirect paths realise the same effective rate. According to Equation (5.2), the condition can be formally expressed as:

$$r_{ms} = min(r_{mr}, r_{rs})\frac{1}{2}. \tag{6.5}$$

Note that, given a straight line BS-RS-MS, when $d_{ms} > d_{rs}$ holds by default indirect transmission is chosen. Substituting in Equation (6.5) with Equation (5.2) and solving for d_{ms}, we can derive a formal expression for d_c, namely:

$$d_c = d_{rs}\left(2\frac{P^{tx}_{ms,max}}{P^{tx}_{rs,max}}\right)^{1/a}. \tag{6.6}$$

This general expression takes both characteristics of the relay - transmit power P^{tx}_{rs} and position d_{rs} - as parameters; a is the path loss exponent. When a RS

has the same maximum transmit power as a mobile the equation reduces to $d_c = d_{rs}(2)^{1/a}$.

Let's now consider SoptOBO. For SoptOBO, from Equation (5.2) it follows that the critical distance d_c is determined by the equality:

$$r_{ms} = \frac{r_{mr}\tau_{mr}}{\tau}. \tag{6.7}$$

By definition SoptOBO is optimized for the transfer of equal amount of data on both sub-paths, i.e. $r_{mr}\tau_{mr} = r_{rs}\tau_{rs}$. In combination with the limitation set on the transfer times, i.e. $\tau_{mr} + \tau_{rs} = \tau$, we can express τ_{mr} in terms of rates, namely:

$$\tau_{mr} = \tau\frac{r_{rs}}{r_{mr} + r_{rs}}. \tag{6.8}$$

This result when substituted in Equation (6.7) allows us to express the rate of a mobile station towards the base station, i.e. r_{ms} in terms of the rates on the relay path:

$$r_{ms} = \frac{r_{mr}r_{rs}}{r_{mr} + r_{rs}}. \tag{6.9}$$

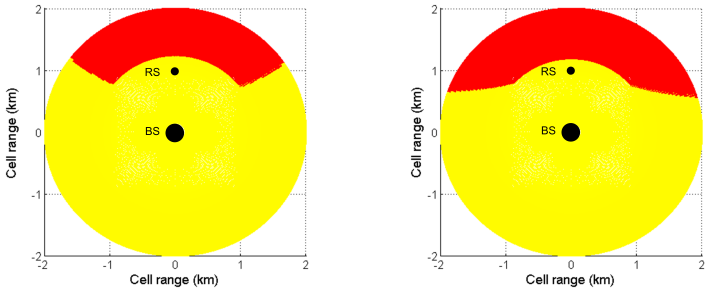
If we solve Equation (6.9) for d_{ms}, after substituting all rates according to Equation (5.2), we can derive an expression for the critical distance d_c. This critical distance can be given as a function of the positions of both the relay and the mobile station. For a solution dependable only on relay characteristics we need to solve the system of equations:

$$\begin{aligned} d_c &= \left(d_{mr}^a + \frac{P_{ms}^{tx}}{P_{rs}^{tx}}d_{rs}^a\right)^{1/a}, \\ d_c &= d_{mr} + d_{rs}. \end{aligned} \tag{6.10}$$

The system of equations (6.10) is in general difficult to solve explicitly and we therefore use numerical approaches. If we assume the relay to have the same transmit power as a mobile station, the first expression of the system (6.10) simplifies to $d_c = (d_{mr}^a + d_{rs}^a)^{1/a}$.

6.6 Numerical results

All results are generated based on the model of Section 6.3 an by consulting Table 5.2 and Equation (6.2) for effective rates and flow throughput respectively for the individual segments. The discussion of the results starts in Sections 6.6.1 with derivation of the service area of a relay station.



(a) Service area SOBO (b) Service area SoptOBO

Figure 6.5: Service area based on effective data rate of (a) SOBO and (b) SoptOBO with a relay station located at 1 km and default transmit power 0.125 Watt.

Subsequently, Section 6.6.2 presents an evaluation of the impact of key relay characteristics such as position and transmit power on effective data rates. Since these parameters are relay and not scheduler related we present the results only for SOBO.

Finally, in Section 6.6.3 the three scheduling schemes - OBO, SOBO and SoptOBO - are compared in terms of effective data rates and flow throughputs. Using spatial graphs we show how relaying can improve performance and what the differences between the relay-enabled scheduling schemes are.

The parameters settings applied in this chapter are identical to the parameter settings of Chapter 5, see Section 5.6.1.

6.6.1 Service area - numerical results

Complementing Section 6.5, which describes the service area of a relay with analytical expressions, this section presents our findings based on numerical calculations. In fact, we represent the studied cell by a set of locations (possible user positions). A single relay station is positioned at 1 km from the base station and has transmit power of 0.125 Watt. Each location is characterised by two distances - one to the base station and another to the relay. Hence, for each location we can define a direct and an indirect path. After we calculate the effective rate for each path we can determine which path delivers the faster effective rate. A location belongs to the service area of the relay when the ef-

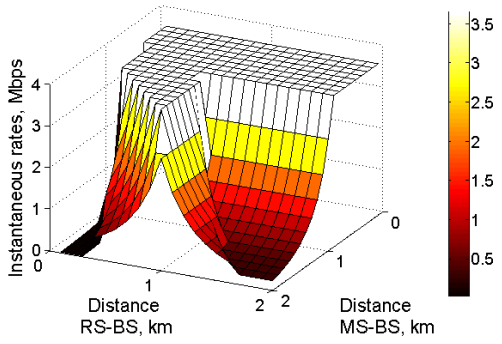


Figure 6.6: Impact of the RS location on MS instantaneous data rates (SOBO scheduler).

fective rate on the indirect path is higher. The results for SOBO and SoptOBO are presented in Figures 6.5(a) and 6.5(b) respectively. The light surface corresponds to the service area, i.e. indirect transmissions, and the dark surface - to all locations from which direct transmissions are preferred.

Comparing Figures 6.5(a) and 6.5(b) we can conclude that using SoptOBO extends the area mainly for locations far from the relay while the changes for close ones are negligible. The bigger the difference in effective rates on the two sub-paths for SOBO, the bigger the improvements SoptOBO can offer. Since this difference is larger for remote self-limiting MSs they also gain the most with SoptOBO. Note that the service area only provides information on which segments can benefit from relaying but not how much the gain in performance, e.g. data rates, is.

It is interesting to observe that the service area does not start right behind the relay which supports our claim in Section 6.5 that the actual service area is smaller than the improvements in received powers may suggest, see Figure 6.3. Solving Equations (6.6) for SOBO on the straight line BS-RS-MS results in critical distance d_c=1.22 km. Concurrently, in Figure 6.5(a) the service area behind the relay starts at the same distance.

6.6.2 Impact of relay station characteristics

In this section we focus on an evaluation of the impact that characteristics of the relay such as relay position and transmit power have on the performance

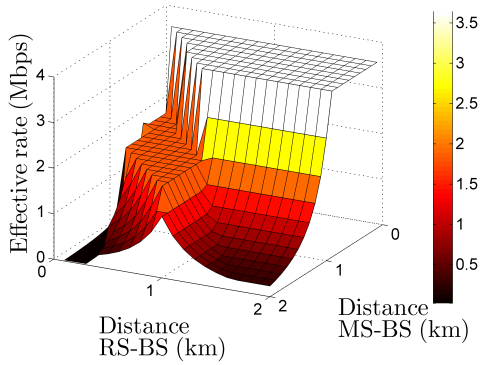


Figure 6.7: Impact of the RS location on MS effective data rates (SOBO scheduler).

of mobile stations. It is expected that the size of the service areas is affected as well. These two parameters are quite important from deployment point of view since they can provide feedback on what optimal deployment choices are to introduce a relay in a mobile network. First, the influence on the relay's position is evaluated and next the relay's transmitted power. Both evaluations are performed assuming best case connectivity scenario, i.e. the relay station on the straight line between mobile and base station.

Position

Figure 6.6 illustrates the impact of the location of both mobile and relay station on the instantaneous data rate τ_{mr}. Note that for the direct and indirect path this rate depends on the received power at base and relay station respectively. It appears that a RS at about 1 km delivers optimal overall performance - largest number of mobile stations with maximised effective rates. Moving the relay closer to the BS causes poor performance for remote MSs due to increased signal degradation on the MS-RS path. Moving the relay further away towards the edge of the cell has a two-fold effect. On the one hand, increased distance to the base station leads to lower RS effective rates. On the other hand, fewer MSs can benefit from relaying and thus more MSs transmit via the direct path potentially limiting their own transmission. For a relay station positioned at 1 km it there is an interesting drop in instantaneous rate for a user at 1 km from

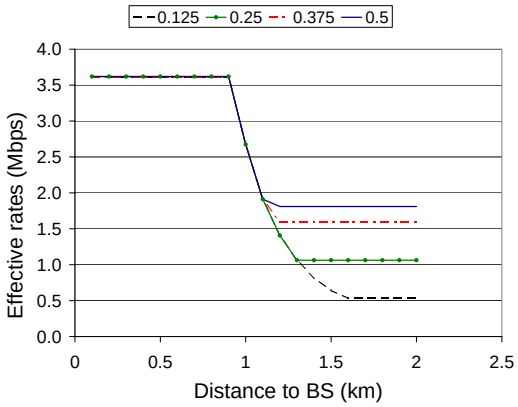


Figure 6.8: Impact of MS's distance to the BS on effective rates for various transmit powers of the RS.

the BS - the reason is that users at 1 km cannot fill up the budget at the base station while users beyond 1 km can fill up the budget at the relay.

The complex effect that relaying has on performance is well illustrated by comparing Figures 6.6 and 6.7. Figure 6.7 presents the dependency of the effective rate on the location of both mobile and relay station. Mobile stations which are capable to fully utilise the total BS resource - on their own or supported by the relay - maximise their instantaneous rate (applies to the flat area in Figure 6.6). In terms of effective rate however such MSs show distinctive behaviour. On the direct path the effective rate is the same as the instantaneous (applies to the higher flat area in Figure 6.7). On the indirect path, due to data forwarding by the relay, it is twice smaller, i.e. lower flat area in Figure 6.7. The performance of MSs who cannot fill up the budget B_D depends on their position and degrades as the distance increases.

Transmitted power

Figure 6.8 presents the dependency of the effective rates on the transmitted power of the relay. The performance on the straight line BS-RS-MS is observed for RS transmitted powers of (0.125, 0.25, 0.375, 0.5) Watt. The RS is located at 1.3 km from the base station. This particular choice is made since it allows us to observe the effects of the transmit power for a larger set of mobile stations

located in front of the relay.

Given a cell radius of 2 km, all indirect transmissions are relay-limited, which explains the lower flat line segment in Figure 6.8. When P_{rs}^{tx} reaches 0.5 Watt the relay can fully utilise the budget B_D. Further increase in the transmitted power will lead to no improvement since the transmissions are now limited by the available resource B_D.

When MSs transmit directly to the base station the graphs coincide as differences start to appear after the relay is used. Interestingly, the higher the RS transmitted power the earlier indirect transmissions become beneficial, i.e. the begin point of the lower flat lines moves to the left. By increasing the transmit power we enable the relay to realise higher BS budget utilisation, i.e. higher data rates, than a MS located in front of the relay can achieve.

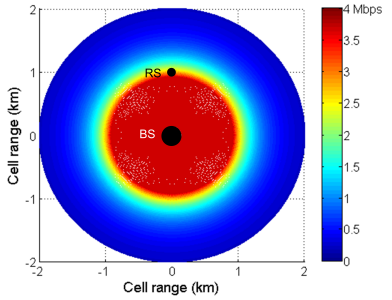
The maximum effective rate $r_{eff,max}$ of a MS is realised when the budget B_D can be fully utilised. Our calculations show that on the direct path $r_{eff,max}$ of 3.6 Mbps can be achieved at distances up to 0.9 km while on the indirect path, due to data forwarding, the $r_{eff,max}$ is twice smaller, i.e. 1.8 Mbps for P_{rs}^{tx}=0.5 Watt.

6.6.3 Performance of mobile stations

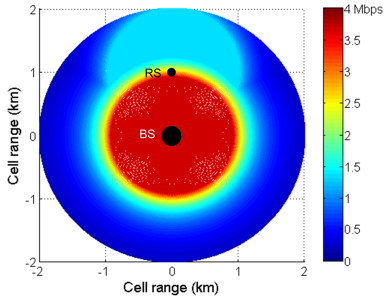
Finally, we perform a quantitative evaluation of the impact a relay station has on the performance of mobile stations. Two performance measures are considered - effective data rates (performance at packet level) and mean flow throughput (performance at flow level).

Effective data rates

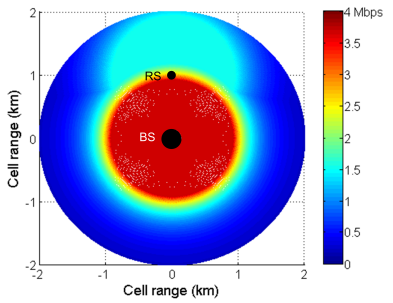
The distribution of the effective rates over the cell area for the three schedulers - OBO, SOBO and SoptOBO - is shown in Figures 6.9(a) to 6.9(c). The brightness changes from high to low as the rates decrease. MSs located in the central white sector around the BS with radius 0.9 km can fully utilise the available budget B. In all scenarios where relaying is used performance visibly improves, e.g. twice or more higher effective rates. Furthermore, a relay configuration with SoptOBO scheme delivers better performance and larger service area than a SOBO scheduler offers. The strategy of SoptOBO to adaptively select the times τ_{mr} according to the MS location gives each mobile the opportunity to optimise its effective rate. The improvement degree depends, as explained, on the difference in effective data rates on the two sub-paths.



(a) Effective rate OBO



(b) Effective rate SOBO



(c) Effective rate SoptOBO

Figure 6.9: Effective data rates in Mbps realised with (a) OBO; (b) SOBO; and (c) SoptOBO given a relay located at 1 km and default transmit power 0.125 Watt.

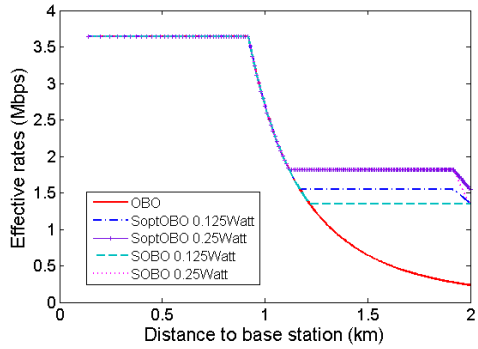


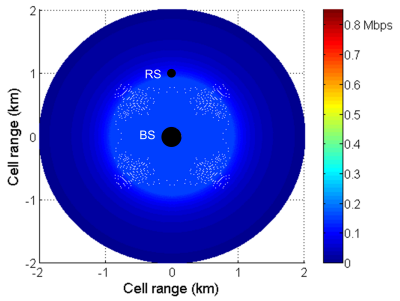
Figure 6.10: Effective data rates for OBO, SOBO and SoptOBO for two different values of the relay transmit power - 0.125 and 0.25 Watt. The line BS-RS-MS is presented.

The effect of doubling the RS transmit power on the effective rates is presented in Figure 6.10. Increasing the RS transmit power is only beneficial for RS-limited transmissions, i.e. MSs close to the relay, and only if the RS cannot fill up the EUL budget, i.e. $P_{rs}^{rx} < B_D$. Therefore the improvements for SOBO are significant. However, in SoptOBO the highly optimized budget utilisation does not leave a lot of room for improvement, see Figure 6.10.

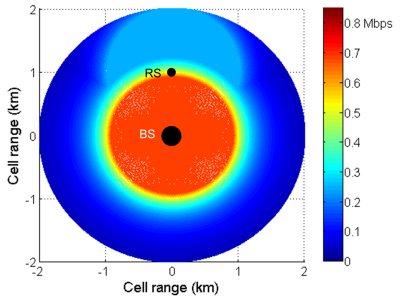
Flow throughput

On the flow level, the benefits of relaying expand towards MSs which use the direct path. As Figure 6.11 illustrates, introducing a single relay station already leads to significant improvements in the mean flow throughput of all MSs irrespective of their location. Note that the color intensity indicates the realised value as the darkest corresponds to the lowest and the brightest to the highest. MSs close to the base station or at the cell edge but far from the relay also register better performance, i.e. brighter colour, even if they do not transmit via the RS. The particular improvements of both throughput and coverage depend on the scheduling scheme - as Figure 6.11 shows the SoptOBO scheme outperforms the SOBO.

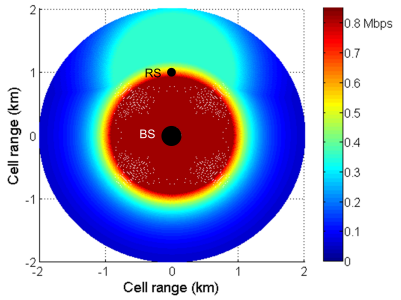
The more detailed Figures 6.12 and 6.13 provide better quantitative insights based on the mean flow throughput realised on the straight line MS-RS-BS (Figure 6.12) and its opposite line, i.e. at 180-degrees (Figure 6.13). Selecting



(a) Flow throughput OBO



(b) Flow throughput SOBO



(c) Flow throughput SoptOBO

Figure 6.11: Mean flow throughput in Mbps realised with (a) OBO; (b) SOBO; and (c) SoptOBO given a relay station at 1 km from the base station and transmit power 0.125 Watt.

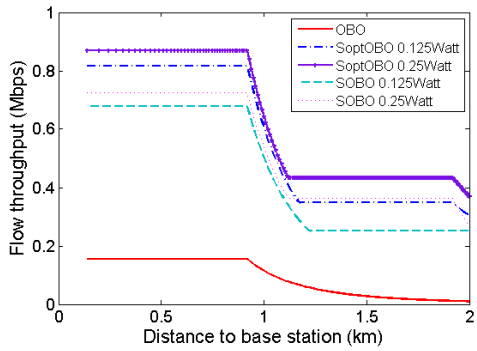


Figure 6.12: Mean flow throughput on the straight line BS-RS-MS with a relay station located at 1 km and transmit power 0.125 Watt.

these two opposite directions allows us to show that the impact of the relay station applies to the whole cell. The graphs suggest that relaying can improve flow performance up to five times for close MSs and up to 20 times for remote MSs which use the relay, see Figure 6.12. Furthermore, as we observed for the effective rates, increasing the transmit power of the relay could lead to further improvements.

The overall improvement at the flow level can be explained by considering the system (cell) occupancy. Increased effective rates for remote users translate into shorter service time and decreased system load, which influence positively all MSs, see Equation (6.2). The poor performance of the OBO scheme is the result of inefficient resource utilisation and thus long service time. As Figure 6.13 shows remote MSs which use a direct path register the smallest improvement due to low effective rates.

We argued in Section 6.1 that a representative performance evaluation should consider flow dynamics. As Figures 6.12 and 6.13 indicate the analysis delivers crucial performance insights and allows us to observe the 'long-term' effects of relaying. We can conclude that accounting for the flow behaviour of MSs is an essential part of the system evaluation.

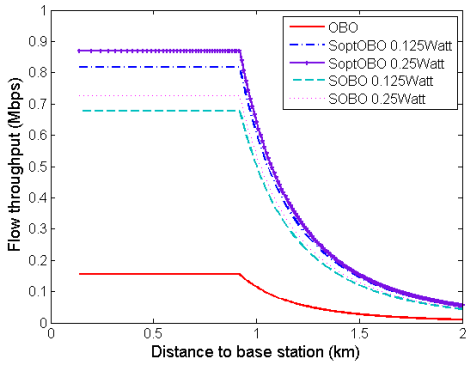


Figure 6.13: Mean flow throughput on the line BS-MS (opposite to the straight line BS-RS-MS) with a relay located at 1 km and transmit power 0.125 Watt.

6.7 Conclusions

This chapter discussed deployment and performance issues for two relay-enabled scheduling schemes in a cell with a single relay station. In order to show the advantages that relaying offers we also considered a reference scenario without a relay. The study was positioned in the context UMTS/EUL.

Our study had three main objectives. The first was to determine the so called service area, i.e. the geographical area where users benefit from transmitting via the relay station. We provided an analytical approach to describe the service area complemented by numerical calculations. This can be helpful during deployment in order to determine how many relay stations suffice to cover a target area. Our results show how the shape and size of the service area depend on the relay characteristics and on the scheduling scheme.

As second objective we evaluated how relay characteristics, such as transmitted power and location, affect the effective data rate of a scheduled user. Specifically, we have determined the optimal position of the RS which maximises the overall performance. Finding an optimal position benefits both the time and financial aspect of introducing relaying to a mobile network.

The knowledge that we gathered while completing these two objectives can help us during the planning of a mobile network in terms of how many relays are needed and at what locations they should be placed. Particularly valuable are the analytical expressions that we derived for the service area but also for the

user's performance. Such explicit expressions can significantly benefit network planning and are easy to interpret.

Finally, the performance of mobile stations was evaluated focusing on the importance of a flow-level analysis. The results showed that relaying has a long-term positive effect on all mobile users independently whether they use the relay or not.

Part III

Uplink scheduling in LTE networks

7 Flow-level analysis of packet scheduling for LTE uplink

7.1 Introduction

The research in the previous Chapters 3 to 6 focused on packet scheduling (and relaying) in UMTS/EUL. In the present chapter, and in Chapter 8, we will consider scheduling in the context of LTE. As pointed out in Chapter 2, LTE is seen as a logical continuation of the current 3G/3.5G networks, which will allow operators to provide higher data rates and to further enhance network capacity.

As previously discussed in Chapter 2.5.2, in OFDMA-based LTE networks the access to the channel resource is organised in two dimensions - time and frequency domain. These two degrees of freedom, together with particular system constraints, make scheduling in LTE a challenging optimization problem, see [30]. For details on the technical aspects of LTE scheduling we refer to Section 2.5.3 and [30].

Most research on LTE scheduling has been treating the downlink scenario, some examples being [36, 71]. Considerably less work has been dedicated to the uplink, where the transmit power constraint of the mobile equipment plays an important role. Furthermore, even fewer studies are concerned with how flow dynamics, i.e. the change in number of active users as result of random flow generation at random cell locations, affects performance for various scheduling schemes.

The research focus of this chapter and Chapter 8 is on investigating the per-

formance of various uplink packet schedulers in an LTE network. In particular, we are interested in the impact of flow-level dynamics. While the present chapter studies in detail the scheduling schemes in a single cell scenario, Chapter 8 has as main objective to evaluate the dependency of inter-cell interference on the scheduler for a multi-cell setting.

In the current chapter we focus on two types of scheduling schemes for LTE uplink with opposing objectives. The first type gives preference to providing fair resource access - all active users are scheduled in a Round Robin fashion and are assigned an equal number of resource blocks to transmit their traffic. The second type aims at optimizing system performance by assigning resource blocks preferably to those users that can make the best use of it. This second scheduling strategy however may lead to starvation of users with poor channel conditions.

In the rest of this section we will further discuss available research on the topic of LTE uplink scheduling (in Section 7.1.1), which is then followed in Section 7.1.2 by the main contributions of the research. The analysis approach applied to evaluate the scheduling schemes is briefly presented in Section 7.1.3 and later in more detail in Section 7.4. Finally the structure of the chapter is given by Section 7.1.4.

7.1.1 Literature

The LTE uplink scheduling problem can in general be formulated as a utility optimization problem, see e.g. [23, 34, 47], where the utility function is usually aggregated throughput maximisation. The complexity of this optimization problem is determined by the complex scheduling freedom in two dimensions, the non-linear relation between allocated radio resources and realised data rate, and, in some studies, the impact of fast fading on scheduling decisions. Still other aspects have shown to influence the complexity of the problem, see [11, 32, 38, 44, 48]. Some examples of such aspects are fairness requirements (e.g. short- or long-term throughput fairness) or specific system characteristics (e.g. regarding fast fading, multiple antennas). As the optimal solutions would mostly be too complex for practical implementation the proposed scheduling algorithms are often based on heuristics yielding reasonable system performance under practical circumstances, see e.g. [5, 76].

Most papers consider the performance (resulting throughputs) of newly proposed scheduling schemes for scenarios with a fixed number of active users in the system (usually split up in different user classes depending on their channel characteristics). For example, the authors of [19] look at a single cell scenario

and define a theoretically optimal scheduling scheme and subsequently propose a suboptimal but practically feasible scheme. The performance of the schemes is evaluated for different numbers of active users; this number, although different per scenario, does not change within the scenario self thus the authors do not really account for flow dynamics. Another interesting study in this context is [77], which proposes three channel-aware scheduling algorithms. The study discusses the realised cell throughput for different number of users but does not specify whether the user population changes dynamically or is pre-set. The manner in which the results are presented suggests that simulations are performed for a fixed number of users but this is not stated in words in the paper. Furthermore, evaluating the aggregated cell throughput does not give much insights on the performance of an individual user.

7.1.2 Contributions

Our investigations on existing literature show that studies that take into account the randomness of user behaviour, leading to a time varying number of ongoing flow transfers, are lacking. Filling this gap, in the present chapter we study the performance of different LTE uplink scheduling schemes for a scenario where initiations of finite-size flows occur at random time instants and locations. We aim at quantifying the resulting performance observed by the users, e.g. flow transfer times, expressing how the performance depends on the user's location in the cell.

We could summarise the main contributions of this chapter as:

- Evaluating various scheduling strategies towards sub-carrier allocation taking into account the influence of flow dynamics. These investigations are particularly valuable since they reveal performance phenomena that are not always visible in the results obtained from studies that assume fixed user population.
- Comparing scheduling strategies with opposing objectives towards performance, i.e. fairness vs. system utilisation.
- Analysing how implementation constrains of the scheduling schemes can affect performance.

7.1.3 Analysis approach

The methodology to study packet scheduling on LTE uplink is based on a similar time-scale decomposition as we used previously in the context of UMTS/EUL,

see Section 3.1.2 and Section 3.4. Certain modifications are necessary due to the additional freedom to schedule in the frequency domain in LTE. Again the approach consists basically of three steps. The first two steps take the details of the scheduler's behaviour into account in a given state of the system, i.e. number of active users and their location in the cell. At these steps the user performance is described in terms of realised data rates by first accounting for the resource block allocation and later for the number of active users. Finally, at the third step for each of the scheduling scheme a corresponding Markov model is built, which describes the system behaviour at flow level. For some special cases of our resource fair scheduling schemes the steady-state distribution of the Markov chain is solved analytically while in other cases simulation is used to derive the steady-state distribution, see Section 3.6.3. From the steady-state distribution of the Markov chain performance measures, such as the mean flow transfer time, can be calculated.

7.1.4 Outline

The rest of the chapter is organised as follows. Section 7.2 provides a brief summary of LTE uplink scheduling and introduces the different scheduling schemes that we will analyse. More details on LTE uplink scheduling were provided in Chapter 2. In Section 7.3 we describe the considered network scenario and state the modelling assumptions. Subsequently, in Section 7.4 the performance evaluation approach is described in detail. Section 7.5 presents and discusses numerical results illustrating the performance of the different scheduling schemes and the impact of the flow level dynamics. Finally, in Section 7.6, conclusions are given.

7.2 Scheduling schemes

In this section we first recall some aspects of scheduling in LTE uplink, necessary for the understanding of the proposed schemes and our modelling choices, and introduce the notation. Subsequently, the proposed scheduling schemes are described.

7.2.1 LTE uplink scheduling features

A key feature of packet scheduling in LTE networks is the possibility to schedule not only in time (as in UMTS/EUL) but also in frequency. The latter option

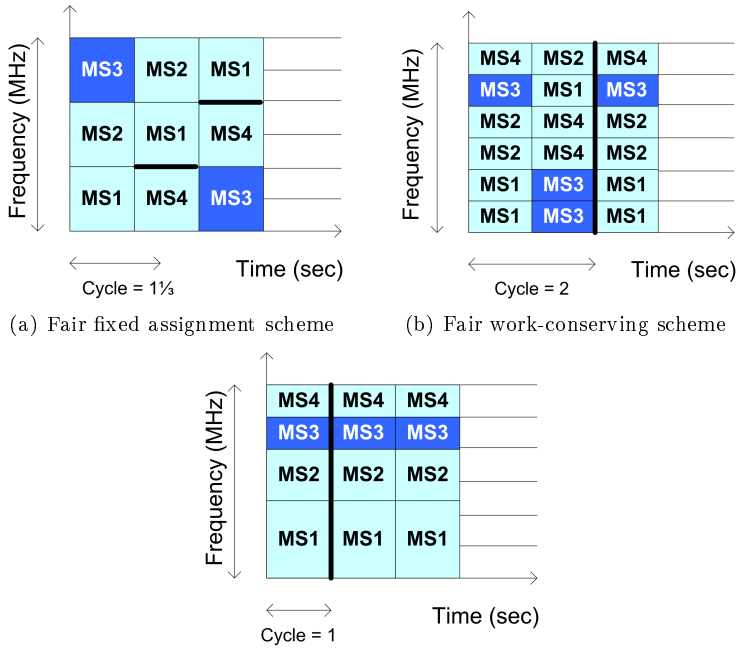
exploits the orthogonality between sub-carriers in SC-FDMA - the radio access technology chosen in LTE uplink. Recall from Chapter 2.5 that the smallest scheduling unit is termed a resource block (RB) and is the intersection between the frequency and time domain scheduling units. In particular, we denote by a resource block a unit of bandwidth of 180 kHz and duration of 1 ms, see Section 2.5.3. If we denote by M the total number of RBs within the available system bandwidth during a single TTI, the scheduler can distribute at most M resource blocks over the active users. The total bandwidth that can be allocated to a single MS depends on the resource availability, the radio link quality and the terminal's transmit power budget.

The transmit power applied by any given MS is equally distributed over its assigned resource blocks, see e.g. [77]. Hence, a higher assigned number of resource blocks implies a lower transmit power per resource block. This has obvious implications for the signal-to-interference-plus-noise ratio (SINR) experienced at the base station, see Section 7.3. Note that the data rate that a user can realise depends on both the number m of assigned resource blocks and SINR experienced per resource block. This issue is discussed in more detail in Section 7.5.2.

The data rate is additionally affected by practical limitations, see [3, 46]. On the one hand, there is a non-zero minimum rate per RB as the SINR is lower bound to a minimum target level, necessary for successful reception at the lowest modulation and coding scheme (MCS). On the other hand, the rate per RB is upper bound by the maximum order MCS that can be technically achieved. In our case we work with 16QAM as the highest modulation order since it should be supported by all terminals but potentially also 64QAM can be used (with limited terminal support). Some studies, e.g. [30] p.203, argue that due to the limited resource available to the Physical Downlink Control CHannel (PDCCH) at most 8 to 10 users can be served in a single TTI. Hence, in practice additional limitation on user's data rate is set by the PDCCH, the impact of which is examined in Section 7.5.4.

7.2.2 Considered scheduling schemes

In our analysis we focus on two (types of) *resource fair* scheduling schemes, which assign equal resource shares to all active users, independently of their respective channel conditions - *fair fixed assignment* (FFA) and *fair work-conserving* (FWC). Furthermore, we consider a greedy scheduling scheme - *maximum added value* (MAV) - as a reference for a strategy that aims at maximum system throughput, given the channel conditions of the active users. Each



(a) Fair fixed assignment scheme

(b) Fair work-conserving scheme

(c) Maximum added value scheme

Figure 7.1: Considered scheduling schemes for LTE uplink; examples with four users.

of the schemes is specified in more detail below, supported by the illustrations in Figure 7.1.

The first scheduler is termed *fair fixed assignment* because it assigns the same, a priori specified, number of resource blocks to each active user (see Figure 7.1(a)). The number of assigned resource blocks per user, denoted m is hence the same for each user, independently from its location, and is an operator-specified parameter. If the number n of active users is such that the total number of requested resource blocks is less than the available number of resource blocks per TTI, i.e. if $n \cdot m < M$, then a number of resource blocks are left idle. Naturally this reflects a certain degree of resource inefficiency in the scheme, especially for situations with low traffic load and hence few active users. When the number of active users is such that $n \cdot m > M$, then not all users can be served in a single TTI and hence it may take several TTIs to serve all

users at least once. We define a *cycle length* c as the number of TTIs necessary to serve all users at least once, as indicated in the Figure 7.1(a). This cycle length can be expressed as $c = \max(1, n/M)$. According to this definition c is not necessarily integral (but at least one) and the start of a given cycle may fall within the same TTI as the end of the previous cycle.

The second scheme, the *fair work-conserving* scheme, aims to avoid the resource inefficiencies of the FFA scheme under low traffic loads, while still preserving the resource fairness property. The scheme's objective is to distribute the available resource blocks evenly over the active users within each individual TTI, see Figure 7.1(b). As result the FWC scheduler is optimal in the class of resource-fair Round Robin schedulers. In principle each of the n users is assigned M/n resource blocks in each TTI. Since M/n needs not be integral, in an implementable version of the FWC scheduler, a scheduling cycle is defined of multiple TTIs during which user-specific resource block assignments appropriately vary between $\lfloor M/n \rfloor$ (low allocation) and $\lceil M/n \rceil$ (high allocation) in order to, on average, achieve the fair assignment of M/n resource blocks. More specifically, the cycle length is equal to the smallest integer c such that $c \cdot M/n$ is integral, which is at most equal to n.

An additional scheme of the type of greedy scheduling is taken as a reference for maximised capacity. The scheme, termed *maximum added value* (MAV), has as main objective to maximise the total data rate realised given the active users present in the system. The scheme distributes the resource blocks to those users that can make best use of it. In particular, scheduling decisions are based on a metric termed *added value*, which, for a particular user, is the gain in data rate that a new resource block can deliver to that user. Recall that the transmit power of a user is equally divided over its assigned RBs and hence giving a new RB to the user leads to a decrease in the realised SINR. If the new resource block will result in an SINR lower than the minimum acceptable, the added value metric is set to zero. Of all active users the one with the highest added value is assigned the resource block. This procedure continues until all resource blocks have been assigned thus resulting in cycle length $c = 1$. In the MAV scheduling it is possible that cell edge users are deprived from service if the system is under high load - since other users can make better use of the available resource blocks cell edge users get none. The operation of the scheme for four active users is illustrated in Figure 7.1(c); the channel conditions are best for MS1 and worse for MS3 and MS4.

7.3 Model

We consider the scenario of a single cell. The cell is divided in K circular zones of equal area in order to differentiate between user's distance to the base station. Each zone is characterized by a distance d_i to the base station, measured from the outer edge of the zone. Mobile users are uniformly distributed over the cell zones and flow arrivals follow a Poisson process with rate λ. Hence the arrival rate per zone λ_i is equal for all zones (due to equal area) can be derived as $\lambda_i = \lambda/K$, where $i = 1, ..., K$. The distribution of the active users over the zones of the cell we term *state* $\underline{n} = (n_1, n_2, ..., n_K)$.

All mobile stations are assumed to have the same maximum transmit power P_{max}^{tx}. Each user distributes this maximum power level equally over the RBs it uses to transmit leading to transmit power per RB $P_i^{tx} = P_{max}^{tx}/m_i$. Note that for the FFA and FWC schemes m_i is the same for all zones but it differs in the case of the MAV scheduler. Due to the different distance d_i each zone is characterized by a distinct path loss $L(d_i)$, where $i = 1, ..., K$. We apply Cost 231 Hata propagation model for the path loss (given in dB), according to which

$$L(d_i) = L_{fix} + 10a \log_{10}(d_i), \tag{7.1}$$

where L_{fix} is a parameter that depends on system characteristics such as antenna height and a is the path loss exponent. In the rest of the analysis linear scale is used for $L(d_i)$. Users belonging to the same zone i have the same distance d_i and hence experience the same path loss. At this stage of the research we consider only thermal noise N from the components at the base station. Neither shadowing nor fast fading have been considered. Note that intra-cell interference can be assumed to be effectively zero due to the orthogonality of the sub-carriers in LTE. As we consider a single cell, inter-cell interference is not taken into account in the current model.

Given a known path loss, the received power (per zone) at the base station P_i^{rx} can be expressed as

$$P_i^{rx} = \frac{P_i^{tx}}{L(d_i)}, \tag{7.2}$$

where P_i^{tx} is the transmit power per resource block which is expressed as $P_i^{tx} = P_{max}^{tx}/m_i$. Eventually, for the signal-to-interference-plus-noise ratio measured at eNodeB from user of zone i we can derive:

$$SINR_i = \frac{P_i^{rx}}{N} = \frac{P_{max}^{tx}/m_i}{L(d_i)N}. \tag{7.3}$$

Recall that it should hold that $SINR_i \geq SINR_{min}$ for each zone.

7.4 Analysis

Our proposed evaluation approach, as discussed earlier, consists of three steps. First we perform packet level analysis, which accounts for scheduler specifics and system characteristics. The so-termed instantaneous rate is defined and is later used in step two to derive a state-dependent throughput that accounts for the effect of the number of MSs in the system and their position, i.e. the system state. Eventually, in step three a Markov model is set up to model the resulting system behaviour at flow level. From the steady state distribution of the model we can derive performance measures such as mean flow transfer times T_i. These steps are explained in more detail below.

7.4.1 Instantaneous data rates

The data rate realised by a user (from zone i) when it is scheduled is what we term *instantaneous rate* r_i. It is determined by the SINR as derived above, the possible modulation and coding schemes and the receiver characteristics related to that MCS (Modulation and Coding Scheme). The instantaneous rate is calculated over all RBs that are allocated to a particular user. In our analysis we use the Shannon formula modified with a parameter σ to represent the limitations of implementation, see Annex A in [3]. Hence, for the instantaneous rate we can write:

$$r_i = (m_i \cdot 180\text{kHz})\sigma \log_2(1 + SINR_i), \tag{7.4}$$

where $m_i \cdot 180$ kHz is the bandwidth allocated to a user in a zone i. Note that both $SINR_i$ and r_i are calculated over the same RB allocation.

In the FFA scheme (with a fixed number of RB allocation per user in a cycle) the instantaneous rate of a particular MS is always the same when the MSs is served. In the case of the FWC and MAV schemes, however, the instantaneous rate depends on the total number of users in the system. Furthermore, for the FWC r_i depends on whether low or high allocation is applicable in the specific TTI, see Section 7.2.2, and hence for the FWC scheme we calculate two instantaneous rates $r_{i,L}$ and $r_{i,H}$ respectively.

7.4.2 Flow-level analysis

The flow-level behaviour can be modelled by a K-dimensional Markov chain with state space $\underline{n} = (n_1, n_2, ..., n_K)$, $n \geq 0$ and $i = 1, ..., K$. The jumps in the Markov chain represent the initiation and completion of flow transfers, where

the corresponding transition rates in a particular state are determined by the (a-priori) given arrival rates λ_i and the long-term flow throughputs in that state. These throughputs are termed *state-dependent throughputs* $R_i(\underline{n})$ and can be derived from the instantaneous rates, which we calculated in Section 7.4.1, and from the cycle length. For the FFA scheduler the state-dependent throughput can be easily expressed as $R_i(\underline{n}) = r_i/c$. The MAV scheduler has by definition a cycle length of a single TTI and thus $R_i(\underline{n}) = r_i$.

For the FWC scheme we need to consider the variation in low resource block allocation ($\lfloor M/n \rfloor$ blocks) and high resource block allocation ($\lceil M/n \rceil$ blocks). Each allocation applies for a fraction a_L and a_H, respectively, of the scheduling cycle as follows:

$$\text{Low allocation}: \quad a_L = \left\lceil \frac{M}{n} \right\rceil - \frac{M}{n}, \tag{7.5}$$

$$\text{High allocation}: \quad a_H = \frac{M}{n} - \left\lfloor \frac{M}{n} \right\rfloor. \tag{7.6}$$

Eventually for the state-dependent throughput we can write for the FWC scheme:

$$R_i(\underline{n}) = a_L r_{i,L} + a_H r_{i,H}. \tag{7.7}$$

Note that the particular form of the Markov chain is scheduler specific via $R_i(\underline{n})$, $i = 1, ..., K$. Determining the steady-state distribution allows us to derive performance measures such as mean flow transfer times.

The steady-state distribution can be found either by simulating the (state transitions of) the Markov chain or, in special cases, by analytical approaches leading to explicit closed-form expressions if the model turns out to belong to a well studied class of Markov chains. In particular, in our study the model of the FFA scheduler appeared to be similar to a M/M/1 processor sharing (PS) queuing model with multiple classes of customers and state-dependent service rates[1]. We will further discuss this below. The Markov chains for the FWC and the MAV scheduler are of more complex form and not trivial to solve, which is why we selected a simulation approach for these cases.

[1] This model is similar to the processor sharing model of an OBO scheme presented in Chapter 3 but in the current model the service rates are state-dependent. This is a consequence of the fact that in the FFA scheme, depending on the ratio between requested by the users resources and the total available system resource, service rates differ.

Explicit solution for the FFA scheme

We argue that the Markov chain of the FFA scheduler is similar to the Markov chain describing the behaviour of a M/M/1 PS queuing model with multiple classes of customers and state-dependent service rates. This queuing model is described and analysed in [10], Section 7. In [10] each 'task', given there are k active tasks, receives a service portion $f(k)$. The Markov chain of the FFA scheduler turns out to be similar to the Markov chain of the M/M/1 PS model. In particular, it is recognised that the cycle length of the FFA scheme, which depends on the number of active users n, actually determines the service portions $f(\cdot)$. Using the expression for c, given in Section 7.2.2, we have for the FFA scheme:

$$f(n) = \begin{cases} 1 & \text{for } n = 1, ..., L, \\ \frac{L}{n} & \text{for } n > L, \end{cases} \tag{7.8}$$

where $L = \lfloor M/m \rfloor$, i.e. the maximum number of MSs that can be served in a TTI. Due to the limited number of RBs per TTI we distinguish in Equation (7.8) two situations - when the number of MSs is smaller than or equal to L, and when the number of MSs is bigger than L.

Given the above relationship between the Markov chains, we can write for the mean flow transfer time T_i, see [10], of zone i:

$$T_i = \tau_i \frac{\sum_{j=0}^{\infty} \frac{\rho^j}{j!} \Phi(j+1)}{\sum_{j=0}^{\infty} \frac{\rho^j}{j!} \Phi(j)}, \tag{7.9}$$

where $\Phi(n) = \left(\prod_{j=1}^{n} f(j)\right)^{-1}$ and $\tau_i = F/r_i$ represents the average service requirement of mobile station in zone i; ρ is the system load defined as $\rho = \sum_{i=1}^{K} \rho_i$ where $\rho_i = \lambda_i F/r_i$ is the load per zone. Substituting $f(n)$ in the expression of $\Phi(n)$ we get:

$$\Phi(n) = \begin{cases} 1 & \text{for } n = 1, ..., L, \\ \frac{n!}{L^{n-L} L!} & \text{for } n > L. \end{cases} \tag{7.10}$$

Eventually, from Equations (7.9) and (7.10) we can write for the mean flow transfer time in zone i:

$$T_i = \frac{\tau_i * A}{\sum_{j=0}^{L} \frac{\rho^j}{j!} + \frac{L^L}{L!} (\rho/L)^{L+1} \frac{1}{1-\rho/L}}, i = 1, ..., K, \tag{7.11}$$

with

$$A = \sum_{j=0}^{L-1} \frac{\rho^j}{j!} + \frac{L^L}{L!\rho} \left((\frac{\rho}{L})^{L+1} \frac{L}{1-\rho/L} + (\frac{\rho}{L})^{L+1} \frac{1}{(1-\rho/L)^2} \right).$$

Note that the impact of the distance of each zone is taken in the specific flow size τ_i, expressed in time.

The condition for stability of the general PS model considered in [10] Section 7, is:

$$\sum_{j=0}^{\infty} \frac{\rho^j}{j!} \Phi(j) < \infty. \tag{7.12}$$

When worked out for our case of FFA the stability condition becomes simply $\rho < L$. We can use this expression to derive the maximum arrival rate that the system can support, namely:

$$\lambda_{max} = \frac{KL}{F} \frac{1}{\sum_{j=1}^{K} \frac{1}{r_j}}. \tag{7.13}$$

The relation between the arrival rate and the RB allocation is further numerically examined in Section 7.5.2.

7.5 Numerical results

In this section we present a quantitative evaluation of the three LTE uplink schedulers introduced in Section 7.2.2. We investigate how flow-level performance is affected by the scheduling policy to assign resource blocks. Beforehand we will present the parameter settings and several preliminary investigations, which allow better understanding of the outcomes of the evaluated scenarios presented in Sections 7.5.3 to 7.5.5.

7.5.1 Parameter settings

A cell with radius of 1 km is divided in ten zones, i.e. $K = 10$. Given an equal zone area, the corresponding distances of the different zones are given in Table 7.1. A system of 10 MHz bandwidth is studied, which, given that a RB has 180 kHz bandwidth (and including control overhead), results in 50 RBs available per TTI, i.e. $M = 50$.

Mobile stations have maximum transmit power $P_{\max}^{tx} = 0.2$ Watt. The lower bound on the SINR is -10 dB while the upper bound on performance is determined by a 16QAM modulation that corresponds to SINR of 15 dB. For the path loss we have used $L_{fix} = 141.6$ dB and path loss exponent of $a = 3.53$, resulting from height of the mobile station 1.5 m, height of the eNodeB antenna 30 m and system frequency 2.6 GHz. The thermal noise per sub-channel (180 kHz) is

Table 7.1: Maximum RB allocation

Zone	1	2	3	4	5	6	7	8	9	10
d_i, km	0.32	0.45	0.55	0.63	0.71	0.77	0.84	0.89	0.95	1
$m_{i,max}$	50	50	50	30	20	15	11	8	7	6

Table 7.2: Maximum flow arrival rate (FFA scheme)

m	1	2	3	4	5	6	7	8	9	10
$L = \lfloor M/m \rfloor$	50	25	16	12	10	8	7	6	5	5
λ_{max}	4.8	2.9	2.0	1.6	1.3	1.0	0.9	0.8	0.7	0.7

-121.45 dBm and with noise figure of 5 dB the effective noise level per resource block is $N = -116.45$ dBm. The attenuation of implementation σ is taken at 0.4, see [3] and Equation (7.4). The average file size F is 1 Mbit and the rate λ at which users become active changes depending on the discussed scenario.

7.5.2 Baseline performance

In this section we discuss three relevant issues: (i) the limitations on performance posed by the minimum required $SINR_{min}$; (ii) the system stability condition; and (iii) the correlation between RB allocation and realised instantaneous rates.

$SINR_{min}$ sets an upper bound $m_{i,max}$ on the number of RBs that can be assigned to a user in a particular zone i. In order for the reception to be successful the signal at the base station needs to reach a certain $SINR_{min}$ value (-10 dB in the current research). SINR by definition depends on the mobile transmit power. As the transmit power of a MS is equally divided over its assigned RBs, the higher the number of assigned resource blocks the lower the transmit power per RB and, hence, the lower the realised SINR. Therefore there is a number of RBs beyond which $SINR_{min}$ cannot be achieved anymore. Naturally this number, differs per zone due to the different distance to the base station, see Table 7.1.

The second issue we discuss applies only to the FFA scheme and concerns the RB allocation policy, i.e. how many RBs are assigned per user. From the stability conditions in Section 7.4.2, i.e. $\rho \leq L$, it follows that more RBs per MS results in lower maximum supported arrival rate λ_{max} by the system. As an example for our chosen settings of file size $F = 1$ Mbit and $M = 50$ RBs, Table 7.2 presents the relation between number of RBs, the maximum possible number of MSs in a TTI, i.e. $L = \lfloor \frac{M}{m} \rfloor$, and the maximum supported arrival

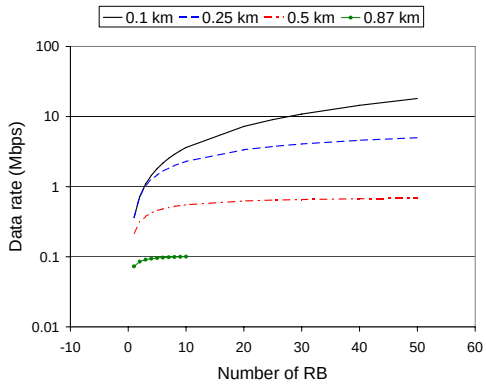


Figure 7.2: Dependency of instantaneous data rates on the number of RBs.

rate λ_{max}. Note that for high loads the large number of active users causes the FWC scheme to behave as FFA with one RB allocation and therefore the maximum arrival rate for the FWC scheme is the same as for FFA with $m = 1$.

Figure 7.2 shows the instantaneous data rate as a function of the number of assigned RBs. Four scenarios corresponding to the distance to the base station (0.1, 0.25, 0.5, 0.87 km) are examined. As Equation (7.4) suggests, increasing the RB allocation leads to an increase in the realised data rates. However, it is interesting to observe that MSs close to the BS benefit more from additionally allocated RBs than remote MSs. This is illustrated by the flattening in the RBs curve for 0.5 km. Recall as well that for remote users $SINR_{min}$ constrains the maximum usable RB allocation hence limiting performance gains. As Figure 7.2 shows for the curve of 0.87 km $SINR_{min}$ is reached for 10 RBs and any allocation bigger than that results in unacceptably low signal quality.

7.5.3 Fair allocation schedulers

In this section we will compare the performance of the two fair allocation strategies FFA and FWC. First, the impact of the scheduling policy in combination with the RB allocation strategy is investigated. Subsequently, we evaluate how the schedulers behave for different arrival rates and are affected by practical limitations on performance.

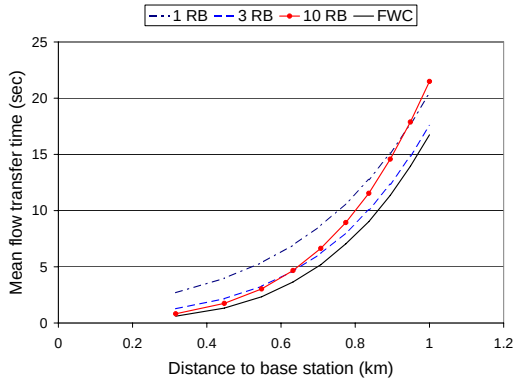


Figure 7.3: Performance evaluation of fair allocation schedulers (FFA and FWC), for an arrival rate λ=0.5 flows/sec.

Impact of scheduling policy

In this subsection we extend the investigation on the impact of RB allocation - both in terms of number of assigned RBs and of allocation policy - towards the flow level. We compare mean flow transfer times for the particular arrival rate of $\lambda = 0.5$ flows/sec.

How the number of assigned RBs affects performance is observed for the FFA scheme. We change the allocated RBs per user from one to three to ten[2], and the results are shown in Figure 7.3. The general trend is that an increase in allocation translates to lower mean flow transfer times, e.g. $m = 1$ vs. $m = 3$. However, for remote MSs high allocation worsens performance, e.g. $m = 3$ vs. $m = 10$. While close-by MSs have sufficient power capacity to exceed $SINR_{min}$ for all of its allocated RBs remote users lack this ability (due to high path loss). They actually use fewer RBs in order to guarantee $SINR_{min}$.

The impact of the scheduling policy is investigated by comparing the results for FFA with $m = 1$ and the FWC scheme, see Figure 7.3. Our choice is motivated by the similar system capacity of the two schemes. The FWC scheme visibly outperforms the FFA with $m = 1$ due to its more efficient distribution of RBs over the active users. Recall that the FWC scheme continues to schedule

[2]These showed to be the most interesting assignments within the range one to ten RBs with a step of one.

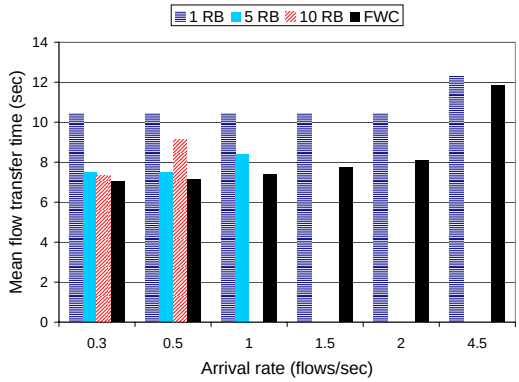


Figure 7.4: Flow level performance as a function of the arrival rate.

users in the same TTI even if all users have been served once. In the FFA scheme however once each user is assigned a resource block in the TTI the scheduling stops thus potentially leaving RBs unused.

Impact of flow arrival rate

Figure 7.4 shows the mean flow transfer time for a range of arrival rates, i.e. (0.3, 0.5, 1, 1.5, 2, 4.5) flows/sec. Again FWC outperforms the FFA scheme. It is more interesting to observe how system capacity changes per scheduling strategy and RB allocation policy. Figure 7.4 confirms the conclusions we already made based on Table 7.2. From the figure it is seen that FFA with $m = 10$ has the lowest system capacity since it is the first to become unstable (already for $\lambda = 1$ flows/sec); hence, no data is plotted. The second worst is FFA with $m = 5$ being unstable for $\lambda \geq 1.5$ flows/sec. FFA with $m = 1$ and FWC both reach their maximum capacity for much higher arrival rates ($\lambda = 4.79$ flows/sec, see Table 7.2). This high λ_{max} explains why initially the performance of FFA with $m = 1$ and FWC does not significantly change. For $\lambda = 4.5$ flows/sec (close to the maximum) an increase in the mean flow transfer times is registered. It is interesting to note that close to their maximum capacity both FFA with $m = 1$ and FWC show very similar performance. This can be explained by the fact that for high load (many active users) FFA with $m = 1$ can, at most times, utilise the total available RBs.

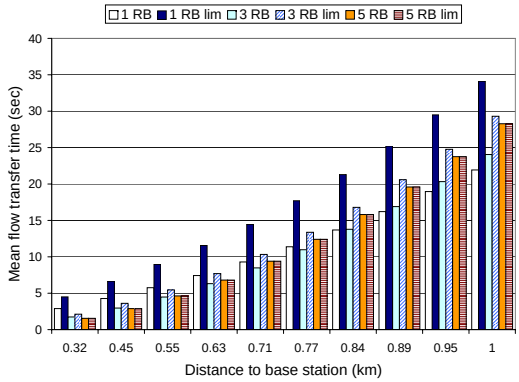
Another valuable observation is that the optimal number of RBs (for the FFA scheme) depends on the arrival rate. Figure 7.4 shows that for high arrival rate it is beneficial to assign few RBs. High arrival rates translate to many simultaneously active users in the system and hence it is very probable to have cycle length bigger than one TTI. By keeping m low we try to diminish the negative impact of a long cycle length. At the same time, independently of its location, a user has the transmit power to fully utilise its assigned bandwidth. On the contrary, an assignment policy of many RBs causes power constrained remote users to leave some RBs unused. This inefficiency, strengthened by the fact that remote users have low service rates, exhibits during flow level performance.

7.5.4 Impact of PDCCH limitations

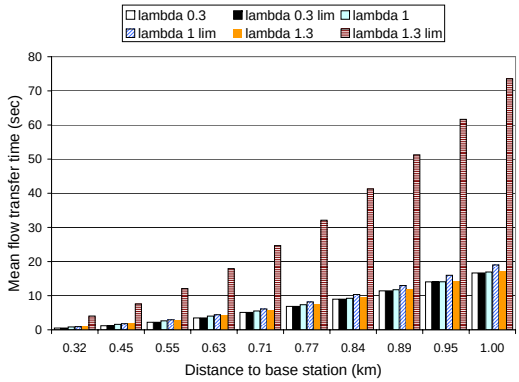
The potential performance gains that the freedom to schedule in the frequency dimension brings could be diminished by practical limitations. One such limitation, according to [30], is the radio resource set up for the PDCCH, which effectively constrains the maximum number of users that can be scheduled in a single TTI. PDCCH carries information from the base station towards a mobile (control packet) about its scheduled uplink transmission. The resource per TTI set aside for these control messages is limited. Therefore, given a fixed size of the control package, only a limited number of uplink transmissions can be served within a TTI. The limit proposed by [30] is eight to ten users; in our evaluation we have chosen for ten users. In the rest of the discussion we will use the abbreviations FFA-lim and FWC-lim to refer to the PDCCH limited versions of the two resource fair schemes.

Each of the access fair schemes, which we discuss, is affected differently by the PDCCH limitation. The FFA scheme will simply serve at the most ten users in a TTI according to the chosen RB allocation policy. Any unused resource blocks will be therefore waisted, which suggests negative impact on performance. It is more interesting however to investigate how strong the effect is and what it depends on. Figure 7.5(a) shows that, independently of the RB allocation policy, i.e. $m = 1, 3$ or 5, the FFA-lim scheme performs worse (or equal) than the original FFA scheme. The difference is biggest for $m = 1$ and non-existent for $m = 5$. The later observation can be explained by the fact that with $m = 5$ or more the maximum number of users that can be served in one TTI is by default ten or less, see Table 7.1.

For the FWC scheme the limitation of up to ten users implies that each user will get assigned at least five resource blocks. Although the mobile stations



(a) FFA scheme



(b) FWC scheme

Figure 7.5: Impact of PDCCH limitation on performance of (a) an FFA scheme and (b) a FWC scheme.

can fully utilise all allocated RBs close by users realise lower data rates than potentially possible causing them to stay longer in the system. This leads to high cycle length and explains the degradation in performance for the FWC-lim case. Hence, at high loads with FWC-lim mobiles use the available RBs much less efficiently than in the original FWC scheme. At low load the two

implementations behave practically the same, which is why also performance is the same. Our results for both scheduling schemes - FFA-lim and FWC-lim allow us to conclude that the effects of the limitation are scheduler specific and generally leads to increased mean flow transfer times.

7.5.5 MAV greedy scheduling

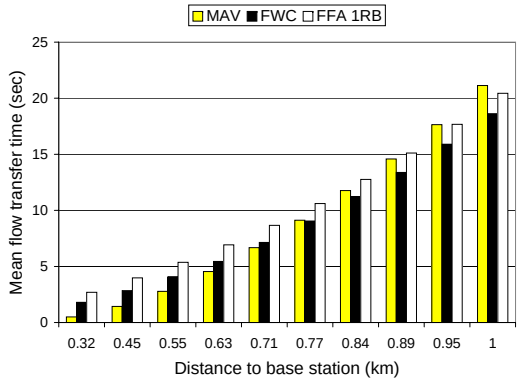
The performance of the two access fair (FFA with $m = 1$ and FWC) schemes and the greedy scheduler (MAV) for an arrival rate λ close to λ_{max} is presented in Figure 7.6(a). The MAV scheme outperforms the other two for close-by users but its performance is worse for remote users. The reason is the preference of MAV to schedule users that can make best use of the available RBs, i.e. close-by users with low path loss, see Figure 7.2.

Figure 7.6(b) presents the average mean flow transfer time over all zones for various arrival rates. Surprisingly, the MAV scheme does not always have the best performance! This is explained as follows. Although MAV tries to maximise the total data rate in each state, such strategy is vulnerable to reaching states where the available resources can not be efficiently used, i.e. states with mainly remote users. Situations as described above are more probable to occur for high load, which corresponds to the results presented in Figure 7.6(b). This hidden inefficiency of MAV also implies that for remote users it is more beneficial to schedule them frequently even if with few RBs per user.

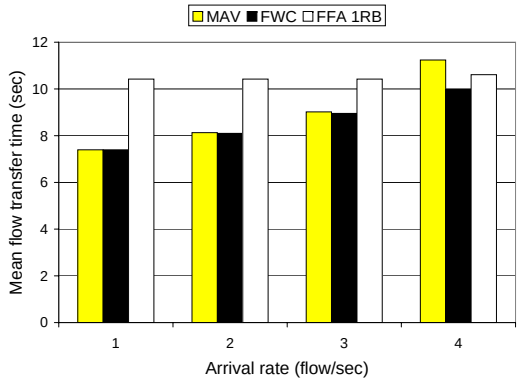
The paradox is that the MAV scheme, which aims at optimising throughput on a per TTI basis, achieves a lower system capacity (overall throughput), when analysed at the flow level.

7.6 Conclusions

LTE relies on an OFDMA-based multiple access technology and as result the scheduling scheme has the freedom to allocate resources both in the frequency and in the time domain. This two-dimensionality turns finding an optimal scheduler into a complex task. In this chapter we presented an investigation of the impact that flow dynamics have on the performance of uplink packet scheduling in LTE. Two low complexity access fair scheduling schemes are examined - both designed to provide equal channel access to the users. Additionally, as a reference base for optimal system performance, a greedy resource allocation scheme is considered that aims at maximising system throughput - this scheme gives preference to users with good channel conditions. All schemes are evalu-



(a) Single arrival rate



(b) Range of arrival rates

Figure 7.6: Comparison of flow level performance for FFA with $m = 1$, FWC and MAV for (a) arrival rate $\lambda = 3$ flows/sec and (b) for a range of arrival rates.

ated by applying a hybrid modelling and analysis approach similar to the one developed in Chapter 3 for UMTS/EUL scheduling. The approach accounts for packet level details such as scheduler's specifics and system details as well as for the dynamic behaviour of users regarding the initiation and completion of flow transfers. Due to its hybrid nature the approach allows for fast evaluation

while considering sufficient details of the investigated model.

Considering flow-level dynamics reveals trends that would be otherwise left unobserved. A particularly interesting example is our finding that the greedy scheduler although maximising system throughput in each system state appears to be less efficient in some cases (and thus leads to longer overall mean flow transfer times) than the access fair schemes in the long term (i.e. when taking into account the evolution of the system) contrary to initial expectations. Hence, before a scheduling scheme is adopted in practice its performance should be carefully evaluated considering the difference aspects related to a single user but also to a full-scale network.

8 Scheduler-dependent inter-cell interference modelling

8.1 Introduction

In the previous chapter we created a basic understanding of how scheduling in an LTE network operates and which important factors need to be considered when designing a scheduling scheme. In order to keep the analysis tractable we abstracted from an important environmental factor - inter-cell interference. However, inter-cell interference could significantly affect performance since it decreases the signal-to-interference-plus-noise ratio realised over a particular communication channel, see Section 2.3.3. In the context of LTE we are not so much interested in the inter-cell interference over the whole system spectrum but in the interference generated over a specific sub-carrier or, in fact, over a specific sub-channel, see Figure 2.5.

Recall that in LTE we have the freedom to schedule in time and frequency which complicates the inter-cell interference process when compared to a shared resource system such as UMTS. Due to the organisation of the spectrum into sub-channels and the independent allocation of these over the users, each sub-channel experiences different inter-cell interference. Hence, we can observe multiple inter-cell interference processes each of which behaves as a stochastic process, just as in a UMTS/HSPA system. Unfortunately, these processes are not mutually independent.

Inter-cell interference for OFDMA-based networks and LTE in particular has

been widely researched with some studies being [9, 52, 75]. Interestingly most studies concentrate on performance evaluation of a resource allocation mechanism with interference mitigation but do not provide insights on the inter-cell interference process self. We argue however that such insights are valuable for the design of an optimal allocation strategy. In particular, we need to understand how the scheduling scheme affects the inter-cell interference process.

The purpose of this chapter is to analyse how the inter-cell interference process depends on the scheduling scheme in the context of an LTE uplink. We intend to show that knowledge on the inter-cell interference can help us to design 'smart' scheduling schemes. Rather than designing an optimal scheduling scheme here, we intend to build an understanding of the mutual interaction between scheduling and interference, especially in the context of flow-level behaviour.

8.1.1 Literature

In the literature we can identify two distinct approaches to deal with inter-cell interference. Some studies, e.g. [45, 74, 79], take a proactive stand and investigate the possibilities of inter-cell interference coordination (ICIC). ICIC techniques show high potential to maximise performance, e.g. overall system throughput, but rely on (extensive) information exchange among base stations and knowledge on radio channel state of the users. Unfortunately these two requirements are not easily met in practice.

Other studies address the topic of inter-cell interference mitigation in a more passive manner without relying on communication between base stations, e.g. [9, 73, 75]. Instead, these studies propose allocation schemes, whose design is based on pre-knowledge of the inter-cell interference. In order to find an optimal solution often the allocation strategy is defined as an optimisation problem in a multi-cell system with common resource. Some examples of this research approach are [73] and [75], both concentrating on OFDMA technology. [73] studies four frequency allocation schemes - reuse-1, reuse-3, soft frequency reuse and priority soft frequency reuse. Unfortunately the authors do not present the inter-cell interference process corresponding to each of these strategies but only evaluate user performance. Furthermore, persistent flows are assumed and the number of active users per cell is changed in a pre-determined fashion. Similar to [73] is [75], which proposes a new Enhanced Fractional Frequency Reuse scheme and compares it to two other schemes, commonly used in the literature.

Interesting studies that are closely related to the research presented in this chapter are [20] and [52]. [20] compares two allocation schemes which resem-

ble some of the schemes we propose in terms of frequency allocation. The one scheme applies a reuse-1 approach while the other is a type of soft frequency reuse scheme. Contrary to other studies, [20] takes into account flow-level dynamics in a study on LTE downlink scheduling. Despite being very insightful and quite extensive the research does not provide information on how the inter-cell interference pattern is affected by the type of allocation, neither does it discuss flow-level performance in terms of throughput (only SINR is evaluated). Furthermore, an LTE downlink is assumed.

The authors of [52] present a basic investigation on the inter-cell interference process of two frequency allocation strategies for LTE - uncoordinated and coordinated - under various traffic models. [52] provides valuable insights into how the ordering of users in the frequency domain can influence inter-cell interference. Although a multi-user scenario is studied the extent to which flow dynamics are considered only an average metrics were examined. The research inspired us to study in more detail the inter-cell interference process in an LTE uplink. More importantly, we are interested in determining the impact of a larger number of scheduler's specifics and the impact of random user behaviour.

8.1.2 Contributions

The novelty of our research lies in two aspects. On the one hand, in the context of LTE uplink, we determine the dependency of the inter-cell interference on the scheduling strategy and, in particular, on the various ways to order users in the frequency and spatial dimension. On the other hand, in a system of two cells, user performance is evaluated at flow level taking into account the continuously changing number of ongoing flows in a system.

We can summarise our main contributions as:

- Providing insights on how various aspects of the scheduling scheme affect the inter-cell interference process in the uplink in LTE. Such insights can prove valuable in the search of an optimal scheduler for inter-cell interference mitigation.
- Evaluating and comparing the flow-level performance of various interference mitigation strategies in combination with scheduling.

8.1.3 Analysis approach

The research of this chapter is organised into two parts. First, we determine how the scheduling scheme in LTE uplink governs the inter-cell interference

process that a cell generates towards its neighbour(s). This is done for various scheduling schemes. Second, we evaluate user performance at flow level (in terms of mean flow transfer times) taking into account the scheduler-specific inter-cell interference.

The applied analysis approach is based on the single-cell analysis described in Section 7.4. Namely, we begin with analysis at packet level, which captures specifics of the scheduling scheme and of the radio environment, e.g. interference. Subsequently, flow-level dynamics, i.e. the random generation of flow transmissions with finite size, are added to the model by modelling the whole system as a continuous-time Markov chain. Eventually from the steady-state distribution of the Markov chain performance measures, e.g. mean flow transfer times, are derived. In contrast to Chapter 7, where some of the schedulers were described by Markov chains whose steady-state distribution could be found by explicit closed-form notations, the Markov models in this chapter are more complex and therefore a simulation approach is chosen.

8.1.4 Outline

The rest of the chapter is organised as follows. Section 8.2 describes the scheduling scheme that we consider and elaborates on certain aspects of the scheduling strategy which may affect the inter-cell interference process. The model and accompanying assumptions are presented in Section 8.3, while Section 8.4 discusses the applied approach to study both inter-cell interference and user performance. The numerical results are split over two sections - Section 8.5, which covers the study of the inter-cell interference process, and Section 8.6, which discusses performance results. Finally, a brief summary of our work is provided in Section 8.7.

8.2 Interference-aware scheduling

In LTE the main factor, external to a transmission, that affects performance is inter-cell interference. Recall that due to the sub-carrier orthogonality intra-cell interference is not an issue. However, transmissions that use the same resource blocks in neighbouring cells disturb each other and thus inter-cell interference needs to be dealt with. As we previously mentioned there are various interference mitigation strategies proposed in the literature whose purpose is to avoid or decrease inter-cell interference. Some of these schemes are very complex and might require communication between base stations, see [45]. We have deci-

ded to take a step back and to first investigate which scheduling features affect the inter-cell interference pattern and in what manner. We believe that having knowledge of the interaction between scheduler and interference can help us to design 'smart' scheduling schemes with favourable interference patterns. The scheduling scheme, which we consider in this chapter, is presented in Section 8.2.1. Subsequently, in Section 8.2.2 we discuss the approaches towards spectrum division that we associate with the selected scheduling scheme.

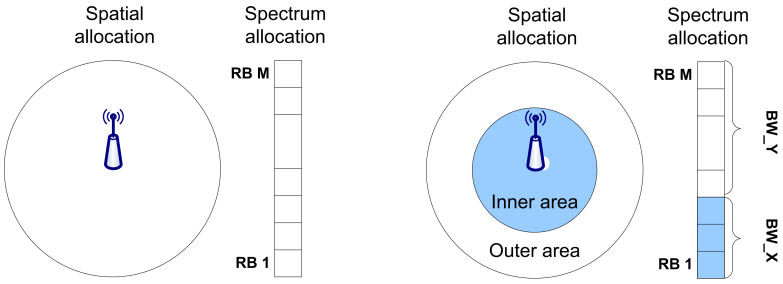
Already in Chapter 4, see Figure 4.5 and the discussion in Section 4.4.2, we showed that the inter-cell interference process is determined not only by the location of the users but also by the specific order in which they are served. Given the additional freedom in LTE to schedule in frequency, we can expect that the order in which the RBs are distributed over the users affects the inter-cell interference pattern as well. In order to investigate this topic, we define five ordering policies in Section 8.2.3.

8.2.1 Scheduling scheme

The scheduling scheme studied in our analysis is in fact a FFA (Fair Fixed Assignment) scheme with $m = 1$, where m indicates the number of resource blocks assigned to a user. We have chosen to start our investigations with a low-complexity scheduler in order to build an understanding on the factors which affect the inter-cell interference pattern. Subsequently we plan to extend the analysis toward more complex schemes such as FWC or MAV, see Chapter 7.

The FFA scheme was presented in Section 7.2.2. Hence, given a total of M RBs in the total LTE uplink bandwidth, there are at most M users that can be served in one TTI. Recall that FFA belongs to the type of access fair scheduling schemes, which indicates that, independently of its channel conditions, each user is given equal opportunity to transmit its data.

Given the single RB per user strategy and the limited number M of RBs in a TTI, two resource utilisation situations are possible. The first situation is when the number of active users n is smaller than the number of available RBs M. In this case one TTI is sufficient to serve all users once. In the second situation - when there are more users than RBs, i.e. $n > M$ - the scheduler needs several TTIs to serve all users once. The total number of TTIs necessary to serve all active users at least once is termed *(transmission) cycle* c and is given by $c = n/M$. In the former case $c = 1$ while in the latter $c > 1$ holds.



(a) SD1 approach (b) SD2 approach

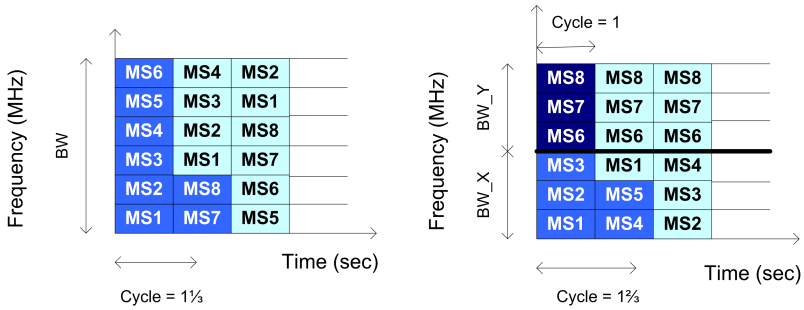
Figure 8.1: Spectrum division approaches taking into account also the geographical division of the cell.

8.2.2 Spectrum division

One aspect of scheduling is how the available LTE uplink spectrum can be divided over the users, taking into account their locations. Spectrum division could provide means to prevent users, located near the cell edge of the 'contact' side of two neighbouring cells to interference with each other. In this chapter we adopt two approaches termed *spectrum division of one* (SD1) and *spectrum division of two* (SD2).

The SD1 approach is characterised by a frequency reuse factor of one, i.e. within a cell the total LTE uplink spectrum can be used to schedule user transmissions, see Figure 8.1(a). Therefore it is possible that users in neighbouring cells are allocated the same resource block and thus their transmissions interfere with each other. As a consequence of the single RB per user strategy there are M inter-cell interference processes - one for each RB (sub-channel). The cycle length for an SD1 approach is determined by the number of active users n. Figure 8.2(a) shows an example with eight users; the cycle length is indicated by a darker shade.

The SD2 approach tries to decrease inter-cell interference by scheduling users over the spectrum depending on their location, see Figure 8.1(b). For this purpose spectrum division of two is applied, which is combined with geographical division of the cell in two as well. Hence, we identify, on the one hand, two spectrum sub-bands BW_X and BW_Y and, on the other hand, inner (close to the base station) and outer (towards cell edge) cell areas. The idea is to have different mappings of sub-bands to cell areas in neighbouring cells. Eventually



(a) SD1 approach (b) SD2 approach

Figure 8.2: Service provisioning for the two spectrum division approaches - SD1 and SD2 - with cycle length bigger than one TTI.

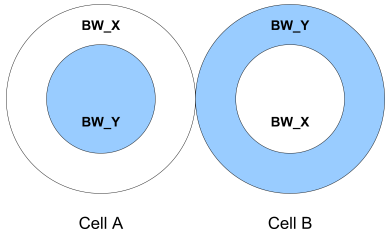
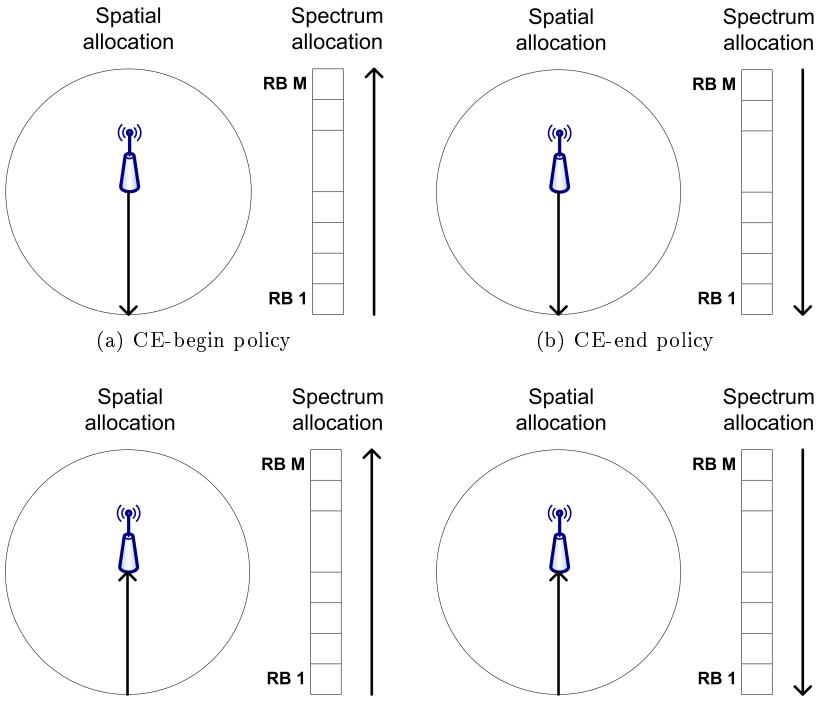


Figure 8.3: Spectrum division of two - an example with two cells.

the goal is to organise transmissions such that users with high received powers at the base station experience high interference from neighbour cells while users with low received powers are only affected by lower interference levels.

The effectiveness of the SD2 approach is observed only if a multi-cell system is considered. In the example of two cells, see Figure 8.2.2, in cell A users from the outer area are scheduled over sub-band BW_X and from the inner area over BW_Y. In cell B the opposite mapping is chosen. Hence, users with low received powers (outer area) in each cell will experience low inter-cell interference (from inner area users) from the neighbour cell.

While SD2 tries to mitigate inter-cell interference it may be less effective in resource utilisation than the SD1 approach. Since the inner and outer area are independently served, it may be that the number of users in each area differs. Hence, the inner and the other areas may have different cycle lengths, a situation



(a) CE-begin policy (b) CE-end policy

(c) EC-begin policy (d) EC-end policy

Figure 8.4: User ordering policies for scheduling in LTE uplink.

which is illustrated in Figure 8.2(b). More importantly, there might be more users than the available RBs in the one area while in the other RBs are left unused.

Note that in a realistic system we would need to apply a spectrum division of at least three. This is necessary in order to limit the spectrum reuse in neighbour cells. We will comment on that in Section 8.6.5. Since we try to gather initial insights on how scheduler's features affect performance a two-cell model is sufficient and hence we apply a spectrum division of two.

8.2.3 Ordering policies

The ordering policy dictates in which sequence users are selected for service and in which order the RBs are allocated to selected users. We base the user selection on the location of the user in the cell. This spatial aspect is important since users at different locations generate different inter-cell interference values at the base station of a neighbour cell, see Figure 4.2. Users close to the base station tend to generate low interference levels. On the contrary, the interference levels from users located towards the cell edge vary significantly depending on their relative position to a neighbour cell. A user close the neighbour cell generates high interference while a user on the opposite end (farthest from the neighbour cell) generates low interference levels.

We have defined five distinct ordering policies accounting for both the spatial (user) and frequency aspect of ordering. The first policy, which is also our reference case, is termed *random* and as the name suggests randomly selects users for service (independently of their location) and allocates RBs to each user in a random fashion as well.

The other four policies have a more 'structured' order of service. The *CE-begin* policy starts allocating RBs from the low end of the used spectrum to users closest to the base station and proceeds with farther located ones, see Figure 8.4(a). Hence, the farthest located users receive RBs towards the high end of the used spectrum. The *CE-end* policy, presented in Figure 8.4(b), adopts the same spatial order but starts assigning first RBs from the high end of the spectrum. The *EC-begin* (see Figure 8.4(c)) and *EC-end* (see Figure 8.4(d)) policies both start serving first the users located farthest from the base station; the former starts assigning RBs from the beginning of the spectrum while the latter starts from the high end of the spectrum. Note that although they may seem similar the CE-begin and EC-end policies in fact differ since not at all times the total spectrum is occupied (some RBs are left unused).

Let us consider again the spectrum division approaches. The SD1 approach can apply any of the described ordering policies in a single cell. Different cells can use different ordering policies. In the SD2 case the inner and outer area can apply different ordering policies. This results in 25 (5x5) possible scheduling combinations for the SD2 scheme. In order to improve efficiency of design we will refrain currently from selecting a specific SD2 scheduling strategy. This is done later in Section 8.6.1 after we have determined the impact of the ordering policy on inter-cell interference.

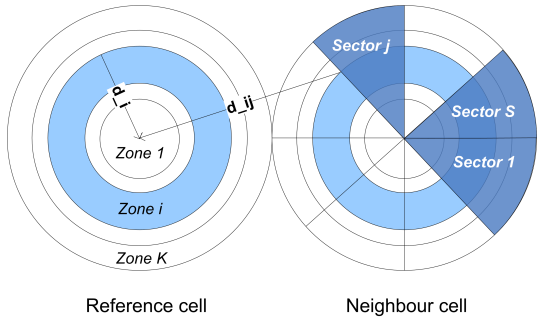


Figure 8.5: System model - reference cell (RC) and neighbour cell (NC).

8.3 Model

The model we use to study inter-cell interference for LTEuplink is presented in Sections 8.3.1. Section 8.3.2 discusses the accompanying assumptions. Most assumptions from Chapter 7 hold here too but some modifications are necessary concerning the study of the inter-cell interference.

8.3.1 System model

Since we are interested in inter-cell interference we need to model at least a two cell system - one cell (neighbour cell NC), where the interference originates, and one cell (reference cell RC), where its effect on user performance is observed. The model is presented in Figure 8.5. For the reference cell the same cell modelling as in Section 7.3 applies, i.e. cell division in K zones of equal area. Each zone is characterised by a distance d_i to the base station, measured from the outer edge of the zone. Additionally the neighbour cell is divided in S sectors also of equal area. The intersection of a zone and a sector is termed a segment, which is characterised by a distance d_i to the base station of the NC and a distance d_{ij} to the base station of the RC derived by applying trigonometry analysis. The number of users in a cell and their distribution is given by the system state $\underline{n}$. Given the chosen cell model the RC is characterised by $\underline{n} \equiv (n_1, n_2 ... n_K)$ and the NC by $\underline{n} \equiv (n_{11}, n_{12}, ..., n_{1S}, n_{21}, n_{22}, ..., n_{KS})$.

8.3.2 Assumptions

The same assumptions about the users apply as in Section 7.3. To briefly recall, users are uniformly distributed over the cell(s) and initiate new flow transfers according to Poisson arrival process with rate λ. Given the equal size of zones and sectors we can define arrival rate per zone (for the RC) $\lambda_i = \lambda/K$ and arrival rate per segment (for the NC) $\lambda_{ij} = \lambda/(KS)$. Alike in Chapter 7, in this chapter we limit the number of active users to the maximum number of resource blocks M that fit in the LTE uplink bandwidth. This is a reasonable assumption, which significantly eases analysis without biasing results. Our observations show that situations with more than M users form 10% of all observations for very high load but are below 1% for medium and low loads. This can be explained by the single RB allocation policy applied by the studied scheduling scheme. Note that the assumption holds only for the scheduling schemes that we study; for other schemes it might not be applicable.

The signal received at the base station is affected by path loss and interferences. The same path loss model as given by the Cost 231 Hata propagation model, see Equation (7.1), is used. As interference sources we identify inter-cell interference and thermal noise.

Furthermore, the neighbour cell is assumed to behave independently from the reference cell meaning it does not experience interference from the RC. As we shown in Chapter 4 this assumption does not significantly affect the qualitative performance.

8.4 Analysis

The analysis of the two-cell model described in Section 8.3 needs to take into account three aspects - (i) the inter-cell interference generated by the neighbour cell, (ii) how the inter-cell interference is accounted for in the reference cell and (iii) user performance at flow level in the reference cell. How our proposed analysis approach tackles these issues is described in Section 8.4.2. Basic component of this 'interference' analysis is the modelling of the user behaviour within a cell (NC or RC) which is achieved by the single-cell analysis of Section 7.4. Certain modifications to this single-cell analysis are required in order to take into account the inter-cell interference. This adapted single-cell analysis is shortly presented next in Section 8.4.1.

8.4.1 Individual cell analysis

The basics of the single-cell model were described in detail in Section 7.4. To suit the purpose of the current research we need to modify it to reflect the specifics of the studied scheduler and to include the factor inter-cell interference. To begin with, due to the single RB per user policy of the considered scheduling scheme, the transmit power per RB applied by any user is the maximum transmit power, i.e. $P_i^{tx} = P_{max}^{tx}$. This holds even if a lower power would be sufficient to achieve the highest possible data rate. Hence, we are looking at a worst-case interference scenario. Furthermore, by adding the inter-cell interference I_{oc} to the signal-to-interference-plus-noise ratio we can rewrite Equation (7.3):

$$SINR_i = \frac{P_i^{rx}}{N + I_{oc}}, \tag{8.1}$$

where i indicates the zone number, $i = 1, ..., K$, and the received power P_i^{rx} is calculated according to Equation (7.2). Note that the inter-cell interference depends on whether the reference or the neighbour cell is modelled.

There are two system aspects that need to be considered in the calculation of the $SINR$. The first one, as explained in Section 7.2.1, is the upper and lower bounds on realised data rate set by the technically feasible modulation and coding schemes. In a specific interference environment these data rates can be translated to corresponding $SINR$ levels. The second aspect is the cycle length c. If the cycle length is larger than one TTI the same user may be served at different sub-channels and will experience different inter-cell interference and hence $SINR$, see Figure 8.2. Given the assumption of maximum M users in the system the SD1 approach always has a cycle of one TTI, i.e. $c = 1$. However, for the SD2 approach, a cycle larger than one is possible since the inner and outer area are served independently.

The implications of a cycle length larger than one TTI concerns the calculation of the *instantaneous rate*. Recall that the instantaneous rate is the rate that a user realises when served and is given by Equation (7.4). If $c > 1$ a user can realise several different rates depending on which RBs is gets assigned. The instantaneous rate in this case is calculated as the average of these realisations.

The other rate introduced in Section 7.4.2 - the *state-dependent throughput* $R_i(\underline{n})$ - is also affected by the cycle length. Recall that it captures the impact of the number of active users and its relation to the instantaneous rate is given by $R_i(\underline{\mathrm{n}}) = r_i/c$. For the SD1 approach holds $c = 1$ and $R_i(\underline{\mathrm{n}}) = r_i$; while for the SD2 approach the general expression holds.

Flow dynamics, i.e. the changing number of active users over time, are

incorporated in our analysis by means of continuous time Markov chains. A state in the model corresponds to the state $\underline{n}$ and as transition rates the Poisson arrival rate λ and the state-dependent throughputs are taken. The form of the Markov chain depends on the cell model. The Markov chain representing the studied scheduling scheme is constructed in a similar manner as the model of a FFA scheduler in Section 7.4.2. The chain has K dimensions corresponding to the K zones of the RC. The Markov chain of the neighbour cell has a more complex form due to the finer granularity in locations, i.e. cell modelling in $K \times S$ segments.

The steady-state distribution of the Markov chains of the scheduler with each of the spectrum division approaches, SD1 and SD2, is challenging to find analytically because of two factors: (i) the dependability of the steady-state throughputs (backward transition rates) on the inter-cell interference; and (ii) the different throughputs realised by the inner and outer area for the SD2 scheme. Therefore, we have chosen to simulate the Markov chain in order to derive performance measures such as mean flow transfer time.

8.4.2 Two-cell system performance

Using the single-cell analysis described in Section 8.4.1 we are able to evaluate the performance of the two-cell model from Section 8.3. The evaluation goes through three phases. In phase one we study the inter-cell interference (per sub-channel) generated by the neighbour cell towards the reference cell as a stochastic process. We are particularly interested in the set of possible interference levels, depending on the users' locations, their probability distribution, reflected by the cumulative distribution function and the entire interference process. In order to achieve that we apply the single-cell analysis for the neighbour cell.

The set of possible I_{oc} values is determined by the modelling of the neighbour cell. The inter-cell interference is in fact the received power at the base station of the reference cell from a user in the neighbour cell. Therefore the number of segments in the NC determines how many distinct I_{oc} levels we can distinguish. Quantitatively, the inter-cell interference levels can be derived based on Equation (5.10) taking into account that maximum transmit powers are used:

$$I_{oc} = \frac{P_{max}^{tx}}{L(d_{ij})}. \tag{8.2}$$

In order to construct the probability distribution we also need information on how often each particular value has exhibited for a specific observation period. Applying straightforward probability calculations we can derive the probability

of each I_{oc} level. Hence, the probability distribution depends on the scheduler and the chosen spectrum division approach since they determine the service time.

The second phase is to take into account the inter-cell interference at the base station of the reference cell. This is easily done for the SD1 approach due to the assumption that there are at most M users in the system. As a result, given a state $\underline{n}$, a user is always assigned the same RB and thus the interference of the RB stays constant for that state. The exact interference values are taken from an earlier generated trace file, which contains information on the inter-cell interference that was generated within each of the M RBs, given state $\underline{n}$ in the NC, and the duration for which these values exhibited, i.e. how long the NC stayed in that state.

A constant interference value within a RB for the duration of state $\underline{n}$ makes the calculation of the $SINR$ and the resulting instantaneous rate straightforward and depending on a single value. The same reasoning holds for the SD2 approach with cycle length smaller than one TTI. However, for cycle lengths bigger than one, in any area (inner or outer), we need to take into account the different inter-cell interference levels that a user experiences. In our analysis the instantaneous rate of a single users is calculated for each of the interference levels that affect the user and then the average realised rate is taken.

The third and last phase of the evaluation is determining the user performance in the reference cell at the flow level. For this again the single-cell analysis is used; this time for the reference cell. The inter-cell interference is accounted for as described in phase two. By simulating the Markov chain of the reference cell we derive its steady-state distribution and mean flow transfer times for each zone.

8.5 Numerical results - inter-cell interference

In this section we present a discussion on the inter-cell interference process that the neighbour cell generates considering the different spectrum division approaches and ordering policies that we described in Section 8.2. In fact, first in Section 8.5.2 we discuss the possible realisations levels of the inter-cell interference, while next in Section 8.5.3 we present the probability distribution functions corresponding to each of the studied ordering policies. Eventually, Section 8.5.4 briefly summarises the main finding of the section. Beforehand, in Section 8.5.1, we state the parameter setting applied during the evaluation.

8.5.1 Parameter Settings

Most of the parameters setting used in this section are identical to those of Section 7.5.1. A system of 10 MHz bandwidth is studied, which, given that a RB has 180 kHz bandwidth results in 50 RBs available per TTI (including control overhead). Cell radius is set to 1 km. The reference cell is divided in ten zones, i.e. $K = 10$; additionally the neighbour cell in divided in ten sectors, i.e. $S = 10$. For the path loss we have used $PL_{fix} = 141.6$ dB considering height of the mobile station 1.5 m, height of the eNodeB antenna 30 m and system frequency 2.6 GHz; and a path loss exponent of $a = 3.53$. The thermal noise per sub-carrier (180 kHz) is -121.45dBm and with a noise figure of 5 dB the effective noise level per resource block is $N = -116.45$ dBm. The attenuation of implementation σ is taken at 0.4 flows/sec, see [3] and Equation (7.4).

Mobile stations have a maximum transmit power $P^{tx}_{\max} = 0.2$ Watt. The lower bound on the $SINR$ is -10 dB while the upper bound on performance is determined by a 16QAM modulation, which corresponds to a $SINR$ of 15 dB. The average file size F is 1 Mbit and the rate λ at which users become active changes depending on the discussed scenario.

The spatial division for the SD2 scheduler is organised into two areas with equal size. This division results at splitting line at 0.7 km. Further, the spectrum is divided into two frequency sub-sets F1 and F2 also of equal size, with 25 RBs each.

8.5.2 Inter-cell interference values

In Figure 8.6 we show the inter-cell interference that a user in the neighbour cell will generate at the base station of the reference cell. The case of SD1 spectrum division is considered where the total spectrum is used by the users. The x and y axes specify the location in the neighbour cell while the z axis indicates the values of the inter-cell interference when a user at that location transmits (with maximum transmit power). The base station of the reference cell in located at coordinates (2,0). As it can be expected from the path loss expression, see Equation (7.1), the closer the user is to the RC, the higher interference it generates. This is visible in the sharply increasing interference values for locations close to the reference cell.

In the case of the SD2 division approach the inter-cell interference levels can be divided into two pools - levels generated by the inner area users in the NC and levels generated by the outer area users. Note that the interference by inner area users is in fact experienced by the outer area users in the reference cell and vice

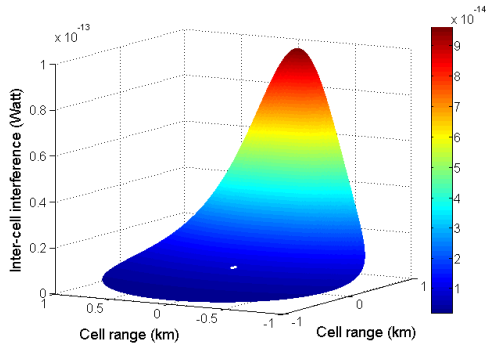


Figure 8.6: Inter-cell interference levels for the SD1 spectrum division approach.

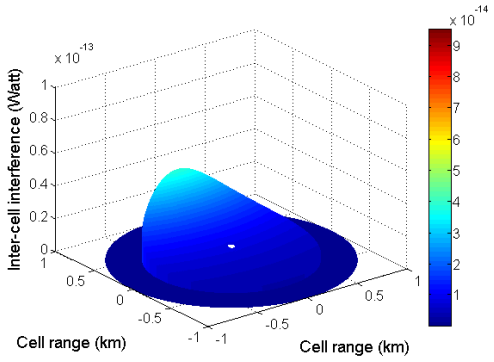
versa. The reason is the mapping between sub-area and frequency sub-set as we explained in Section 8.2. The results are presented in Figure 8.7(a) for the inner area and Figure 8.7(b) for the outer area. The general shape of the graphs is the same as for SD1 where the highest point corresponds to the location closest to the reference cell. The highest value for the outer area is larger than the highest interference level of the inner area due to its closer proximity to the reference cell. The darkest part of the graph represents the area that is not served at the specific frequency sub-set.

The graphs of the inter-cell interference levels, independently of the scheduler, allow us to draw two important conclusions. Though quite obvious these observations are very important for the understanding of the inter-cell interference process:

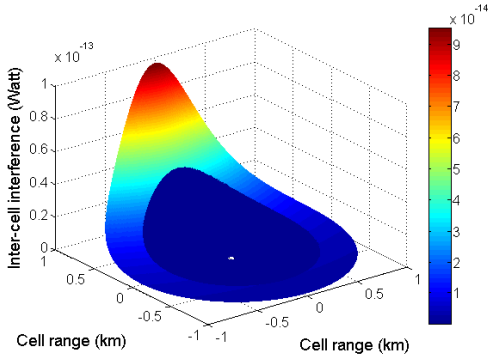
- The interference levels caused by users located close to the base station are close in values, i.e. have low variance.
- The interference levels caused by users located towards the cell edge can differ significantly, i.e. have high variance.

8.5.3 Inter-cell interference process

A cumulative distribution function (CDF), although not providing information on the chronological development of the inter-cell interference, can provide valuable insights in the interference process. The CDF of each of the ordering policies presented in Section 8.2.3 is studied in this section. Note that in fact



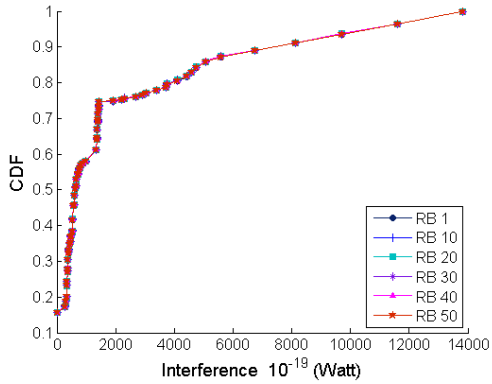
(a) NC: Inner area



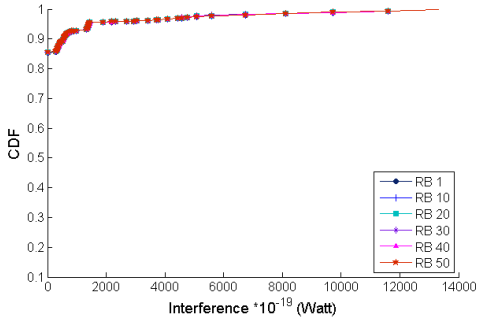
(b) NC: Outer area

Figure 8.7: Inter-cell interference levels for the SD2 spectrum division approach.

there are 50 inter-cell interference processes for each ordering scheme, one for each RB, see Section 8.2. We have chosen to present the CDFs of the inter-cell processes observed at RBs numbers 1, 10, 20, 30, 40 and 50. Hence, in all presented figures the x-axis indicates the observed inter-cell interference levels, while on the y-axes the CDF function for each of the selected RBs is plotted. By default the arrival rate λ in the whole system is chosen at 4 flows/sec such that it is close to the maximum of what the system can support, i.e. $\lambda = 4.8$ flows/sec, see Table 7.2.



(a) High load



(b) Low load

Figure 8.8: CDF of inter-cell interference levels for the *random* ordering policy.

Random ordering policy

Figure 8.8(a) presents the CDF curves observed in the case of a random ordering policy with high load. As we can see the CDF curves coincide, which can be explained by the random manner in which users are selected for service and are assigned RBs. For example while at one time instant on RB1 a user from the centre can be served at the next instant a user from the cell edge can be chosen. Hence, each possible inter-cell interference level has equal chance to exhibit at any RB. The steep initial part of the curves is determined by the many segments

with similar low I_{oc} levels, see Figure 8.6.

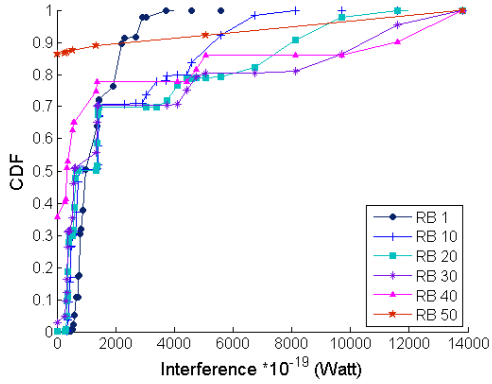
The same general behaviour is observed when low system load is considered with the only difference being the much higher probability of zero inter-cell interference (empty system). This is a logical consequence of the fact that there are only a few users in the neighbour cell and thus inter-cell interference exhibits only for few RBs in the reference cell.

Pre-defined ordering policy

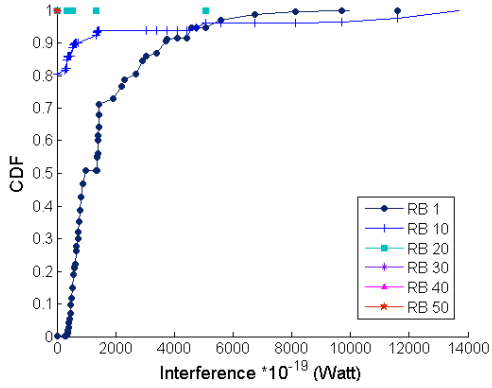
We will now discuss the CDF functions of each of the ordering policies where there is a strictly defined order of serving the users. We begin with CE-begin policy. As Figure 8.9(a) shows, the CDF curves for the CE-begin policy differ significantly. Not only do they not coincide but their form strongly depends on the monitored RB. Observe that, as we move towards the end of the spectrum, initially (from RB1 to RB30) the probability that high inter-cell interference values occur increases. Subsequently however low (including zero) interference values are more often observed, i.e. have higher probability towards the end of the allocated spectrum (e.g. for RB40 and RB50). This behaviour is the combined result of two factors - the spatial ordering policy and the fact that high RBs are often not allocated (if the number of users n is smaller than the available RBs M). On the one hand, with the particular spatial ordering (centre to edge), edge users are served as last and hence receive RBs from the end of the spectrum. These users, when located close to the reference cell, have much higher interference levels than centre users, which explains the initial increase in the probability of high interference levels. On the other hand, towards the high end of the spectrum some RBs are left unused and their corresponding RBs in the RC do not experience inter-cell interference. Because of the specific spatial ordering, these unused RBs are at the end of the spectrum.

The impact of low system load leading to unused RBs exhibits even more if low system load is considered. As Figure 8.9(b) shows the probability of low or zero inter-cell interference is already high for RBs close to the beginning of the spectrum, e.g. RB10.

The same observations hold for the CE-end scheme but in reverse order, see Figure 8.10. RBs from number 50 to 20 experience increasing probabilities for high interference values while RBs ten and one have much higher chance to experience low interference levels. This 'reverse' behaviour is explained by the identical spatial ordering as in the CE-begin scheme but starting RB allocation from the end of the spectrum. Our investigations for low load showed identical observations for CE-end (as well as for EC-begin and EC-end in fact) for which



(a) High load



(b) Low load

Figure 8.9: CDF of inter-cell interference levels for the CE-begin ordering policy.

reason we have not presented the corresponding graphs.

The CDF curves for the EC-begin ordering policy are shown in Figure 8.11. Immediately we notice that the CDF curves of RBs towards the end of the spectrum are characterised with decreasing probability of high inter-cell interference levels and increasing probability of a zero interference. This tendency is explained as follows. Cell edge users (with potentially high I_{oc}) are served first at the beginning of the spectrum, e.g. at RB1. Subsequent resource blocks

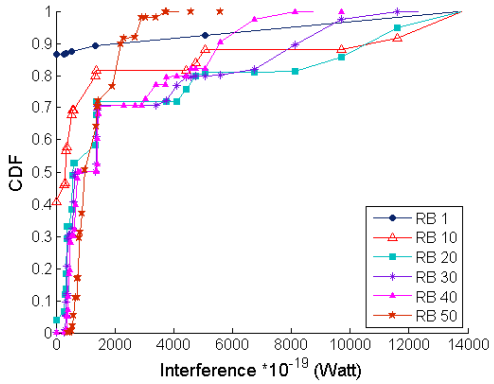


Figure 8.10: CDF of inter-cell interference levels for the CE-end ordering policy.

are assigned to users located steadily closer to the cell centre, which have (on the average) lower induced interference levels. Finally, RBs at the end of the spectrum, e.g. RB40 and RB50, additionally experience the effects of a non-full system.

Another interesting observation is the limited number of inter-cell interference values that exhibit at RB1. Recall that RBs from the beginning of the spectrum (RB1 included) are allocated to cell edge users. These users are farthest away from the serving base station and, due to high path loss, have poor performance, i.e. tend to remain long in the system. As result the variation in inter-cell interference for RB1 is limited to the set of unique I_{oc} levels generated by the cell edge users (which in our case is ten as the number of sectors, see Section 8.5.1).

Figure 8.12 presents the CDF curves for the EC-end ordering policy. It differs from the EC-begin policy only in the directionality of spectrum allocation. Hence, all observations made for the EC-begin policy apply also here as long as we take into account the 'reverse' order of RB allocation.

8.5.4 Concluding remarks

From the results, presented in this section, and their discussion we can draw several conclusions:

- In the case of random ordering the inter-cell interference has the same

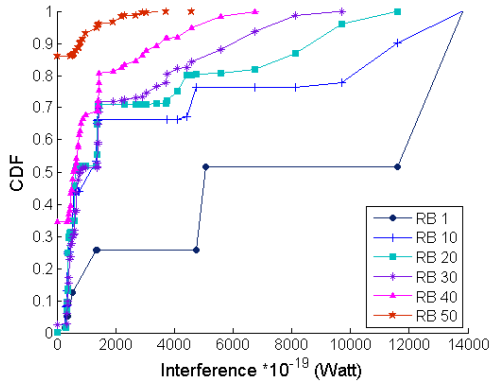


Figure 8.11: CDF of inter-cell interference levels for the EC-begin ordering policy.

probability distribution independently of the part of the spectrum (sub-channel) where it exhibits.

- The ordering of the users over the available spectrum, i.e. from the low or the high end of the spectrum, determines in which part of the spectrum a particular interference level is most probably to occur. Especially, it determines in which part of the spectrum zero interference is likely to occur, caused by unassigned RBs in case of lower system load.

- The ordering of the users depending on their location, i.e. from cell centre to cell edge, determines how interference levels change over the spectrum, e.g. from low to high levels or vice versa. Note that the effects of this aspect of the ordering are independent of the fact from which part of the spectrum RB allocation starts.

- In the case of centre to edge ordering the CDFs of different sub-channels intersect each other, making it difficult to predict the final performance of a multi-cell system.

- In the case of edge to centre ordering the CDFs of different sub-channels lie alongside each other, which could prove advantageous for the coordination of the inter-cell interference of neighbouring cells.

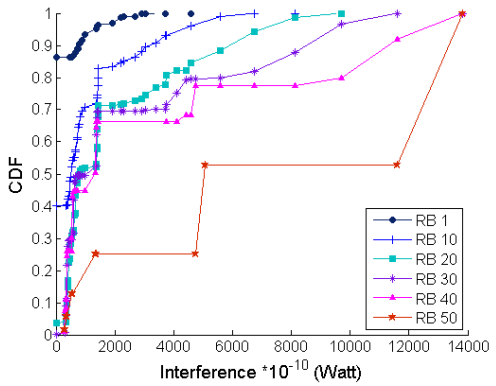


Figure 8.12: CDF of inter-cell interference levels for the EC-end ordering policy.

8.6 Numerical results - flow-level performance

In this second section on numerical results we use the knowledge gathered in Section 8.5 to design evaluation scenarios with highest potential for maximised performance (for both the SD1 and SD2 approaches). These scenarios are described in the following Section 8.6.1. The impact of the ordering policy on flow-level performance is discussed in Section 8.6.2. Section 8.6.3 presents our finding on the interaction between the ordering policies. Finally, in Section 8.6.4 the impact of the spectrum division approach is investigated.

8.6.1 Evaluation scenarios

We designed three groups of evaluation scenarios depending on the effects we want to investigate. The five scenarios comprising the first group, altogether referred to as *test group*, are chosen to allow us to observe the impact of the ordering policy on user performance. The purpose of the scenarios in the second group is to determine how ordering policies in neighbouring cells interact and the implications of that on user performance. Finally, we are interested in the advantages and disadvantages an SD2 approach has to an SD1 approach. Below, the evaluation scenarios are discussed in more detail.

All scenarios of the test group concern the SD1 spectrum division approach. In each scenario the NC adopts one of the five ordering policies, which were described in Section 8.2.3. The RC always adopts same ordering policy, namely

Table 8.1: Evaluation scenarios for SD1 spectrum division approach.

	RC	NC
random CE-begin CE-end EC-begin EC-end	CE-begin	random CE-begin CE-end EC-begin EC-end
random-2	random	random
EC-comb	EC-begin	EC-end
comb-begin	CE-begin	EC-begin

CE-begin, independently of the scenario. This provides us with a common base to compare the effects of the ordering policy in the NC on the performance in the RC. Since each of the test group scenarios is uniquely identified by the ordering policy adopted by the NC in the rest of the discussion we will use the notation test-(CE-begin, CE-end, EC-begin, EC-end or random) scenario.

In order to investigate the interaction between ordering policies (in the case of SD1 approach) we consider two new evaluation scenarios, in addition to the test group scenarios. In the first scenario, denoted *random-2*, both reference and neighbour cell adopt the random ordering policy. This is our reference scenario for user performance at the flow level. In the second scenario, denoted *EC-comb*, the RC applies an EC-begin policy and the NC an EC-end policy. Our choice was inspired by the CDF functions of Figures 8.11 and 8.12, which show that while for the EC-begin policy the probability of high interference decreases in the RB number for the EC-end it increases. All evaluation scenarios for the SD1 approach are summarised in Table 8.1, where the ordering policy for both RC and NC are indicated.

In the case of SD2 spectrum division there is a multitude of possible combinations to evaluation scenarios due to the independent choice of the ordering policy not only in the RC and NC but also in the inner and outer area of each cell. We decided to select one particular combination, which shows high potential for favourable interference environment. The combination itself is as follows - in the RC both the inner and outer areas adopt EC-begin ordering policy; the ordering policy, also for both areas, in the NC is EC-end. In order to explain our choice we use Figure 8.13. The figure shows for each of the ordering policies how users are organised in the radio spectrum depending on their location. Users from the cell centre generate medium inter-cell interference levels; users from the cell edge generate low or high inter-cell interference levels, depending

Figure 8.13: Distribution of users over the RBs for each of the ordering policies - high load example. The interference contribution of a user is denoted by m for medium and l/h for low/high.

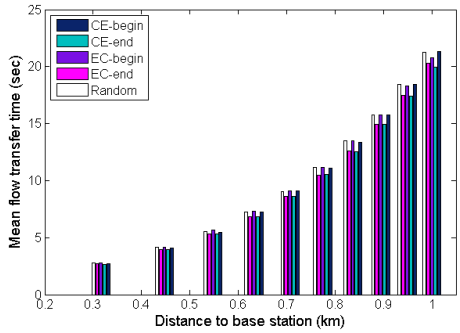
on whether they are on the close or far edge taken to a neighbour cell.

Back to the SD2 scheme, recall that the purpose of SD2 is to organise transmissions such that users at the cell edge to experience as low inter-cell interference as possible. As we can see from Figure 8.13 the most beneficial combination of ordering policies is EC-begin and EC-end. This particular combination has two advantages. On the one hand, cell edge users in the RC are subjected to the moderate interference levels from centre users in the NC. On the other hand, depending on the load in the NC, it may be that some cell edge users in the RC do not experience inter-cell interference (if not all RBs in the NC are allocated). Such situations exhibit more often for low and medium loads and less often for high load. Note that the smallest interference levels are generated by cell edge users at the far end of the RC (best-case interference) but at the same time also cell edge users (close to the RC) generate the highest interference levels (worst-case interference).

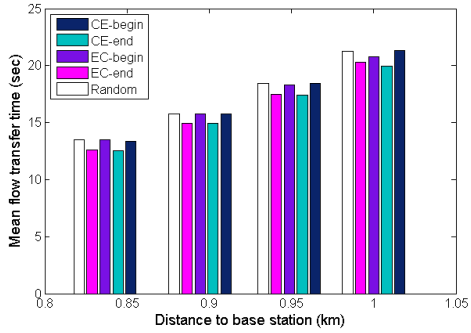
8.6.2 Impact of the ordering policy

In this section we discuss the impact of the ordering policy on the mean flow transfer times of users. For the evaluation we have used with the *test group* scenarios, see Section 8.6.1. First, the behaviour of a system with low load is observed and next a highly loaded system is considered.

The user performance in a lightly loaded system (arrival rate $\lambda = 0.7$ flows/sec) for each of the *test group* scenarios is shown in Figure 8.14(a). In order to ease interpretation of the results we zoom in on the last four zones for which the differences in performance are best observed, see Figure 8.14(b).



(a) Whole cell



(b) Zoomed in view

Figure 8.14: Mean flow transfer times for an SD1 division approach with five ordering policies CE-begin, CE-end, EC-begin, EC-end and random - low arrival rate $\lambda = 0.7$ flows/sec.

The general conclusion from Figure 8.14 is that the type of the ordering schemes does not seem to affect the performance considerably. This will be later observed also for other evaluation scenarios and, as we will explain, is to be explained by the applied scheduling policy of single resource block allocation.

Taking a more detailed look at Figure 8.14(b) applying a pre-determined ordering policy is beneficial for performance when compared to a completely random service order. Furthermore, it seems that ordering from the high end of the spectrum in the NC, e.g. CE-end or EC-end, delivers better results. Please

note that the conclusion is valid in the context of the CE-begin policy in the RC (ordering from the low end of the spectrum). We explain this behaviour as follows. In a system with low load mainly the RBs close to the spectrum end, where allocation starts, are used; RBs from the middle of the spectrum and further are often not allocated. Referring back to Figure 8.13, starting RB allocation in the two cells from the opposite ends of the spectrum then implies that the inter-cell interference for the majority of RBs in the RC is zero.

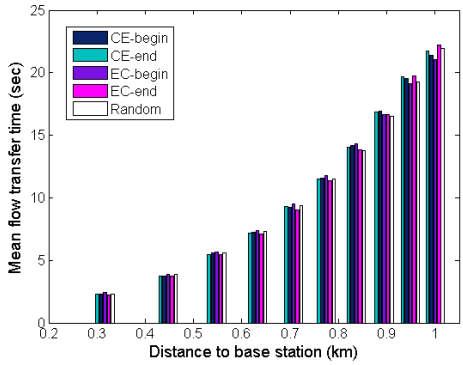
On the contrary, starting RB allocation in the two cells from the same end of the spectrum results in interfering transmissions. Such approach is also inherently inefficient since RBs with zero interference in the RC are left unused. Therefore the CE-begin and EC-begin show slightly decreased performance, i.e. higher mean flow transfer times. Again this observation are valid given the CE-begin policy in the RC.

Based on the observed performance under low load we can conclude that it is beneficial in neighbouring cells to apply ordering policies which start RB allocation from the opposite ends of the spectrum. Further, as Figure 8.13 shows, the spatial ordering of users, e.g. from cell centre to cell edge, has little impact on performance.

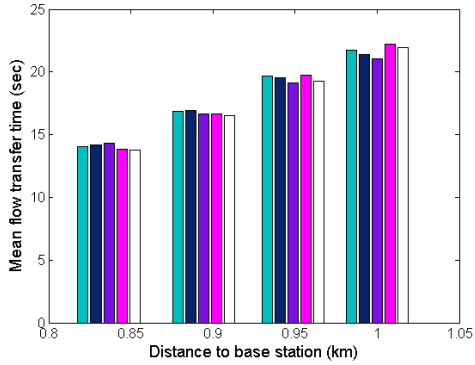
Interestingly when we study the system behaviour with high load (arrival rate $\lambda = 4$ flows/sec) a twist in performance is observed, see Figure 8.15. More specific, the CE-end and EC-end ordering policies are outperformed by the CE-begin and EC-begin policies respectively. Note also that the difference in performance is observed mostly for users towards the cell edge in the RC.

In order to understand the change in performance as the system load increases we refer back to Figure 8.13. If the system is highly loaded the majority of RBs are allocated and inter-cell interference will appear is the biggest part of the spectrum. In such situation, cell edge users in the RC, given the adopted CE-begin ordering, are allocated RBs near the high end of the spectrum. Then, an ordering policy in the NC, which begins RB allocation from the high spectrum end, e.g. CE-end or EC-end, will at all time generate inter-cell interference for the edge users in the RC. On the contrary, an ordering policy in the NC, which starts RB allocation from the low spectrum end, e.g. CE-begin or EC-begin, will generate interference most of the time but not always. Hence, such policy creates the better (lower) interference environment for the cell edge users in the RC.

Note that the mean flow transfer times realised under low and high load do not differ significantly, which is explained by the single RB per user strategy. Under this condition relatively many users (50 with total system bandwidth of 10 MHz) that can be served during one TTI. Hence, increase in the load



(a) Whole cell



(b) Zoomed in view

Figure 8.15: Mean flow transfer times for the SD1 division approach with five ordering policies CE-begin, CE-end, EC-begin, EC-end and random - high arrival rate $\lambda = 4$ flows/sec.

does not quickly cause increase in the cycle length and subsequently decrease in the state-dependent throughputs, see Section 8.4.1. We should also account for the fact that under low load the scheme is inefficient in the resource use, i.e. artificially keeps data rates lower than the potentially possible.

An overall conclusion, independent of the system load, is that in fact user performance does not vary significantly as the ordering policy changes. This may

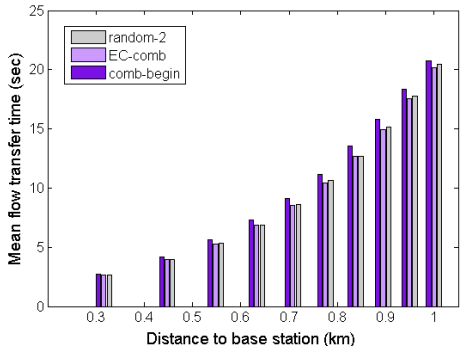
be a little surprising recalling the big differences in CDFs that were observed in Section 8.5.3. It appears that different scheduler aspects affect the inter-cell interference process but they are not crucial for the general performance at flow level.

8.6.3 Interaction between ordering policies

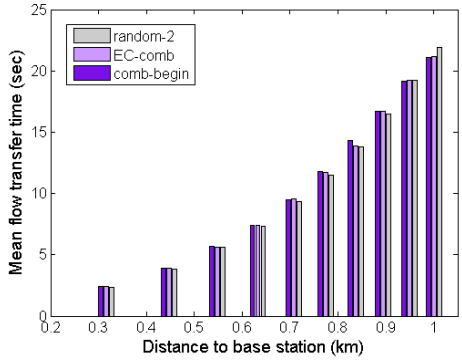
It is interesting to investigate how different ordering policies, when adopted by neighbouring cells interact with each other and how this affects the flow-level performance of users. We have compared three scenarios - random-2 (reference scenario), comb-begin and EC-comb. With *comb-begin* we denote the test scenario for which the RC adopts CE-begin ordering and the NC adopts EC-begin ordering. Among the scenarios compared in Section 8.6.2, this scenario showed the best performance for high loads. All scenarios were explained in Section 8.6.1 and summarised in Table 8.1.

The mean flow transfer times realised by each of the three scenarios under low system load (arrival rate $\lambda = 0.7$ flows/sec) are presented in Figure 8.16(a). The results confirm our initial expectation that the EC-comb scenario performs the best, i.e. it realises the lowest mean flow transfer times. We could expect such result since in EC-comb the RC and NC start allocating RB from the opposite ends of the spectrum. This advantage is especially pronounced for low system load, in which case some RBs are left unused. Given the ordering policies, which EC-comb adopts in RC and NC, empty RBs in the NC (hence causing no interference) correspond in the RC to RBs occupied by transmissions from remote users.

Surprisingly, Figure 8.16(b) shows that for high system load (arrival rate $\lambda = 4$ flows/sec) EC-comb loses its advantage and all schemes seem to have the same performance. Under high load most of the RBs in both NC and RC are used. This brings certain implications that are understood after a careful analysis of how the ordering policies interact. We first discuss the case of EC-comb - RC with EC-begin policy and NC with EC-end policy. From Figure 8.13 we see that with high load cell edge users in the RC are assigned RBs from the low end of the spectrum. In the NC these RBs are allocated to users from the cell centre which generate moderate inter-cell interference levels. If we now look at the comb-begin scenario we will observe a similar pattern. Namely, cell edge users in the RC (CE-begin policy) are served at the high end of the spectrum, where also cell centre users are scheduled in the NC (EC-begin policy). This similarity in allocation pattern explains the little difference in performance between the EC-comb and comb-begin scenarios. The advantage of unallocated



(a) Low load

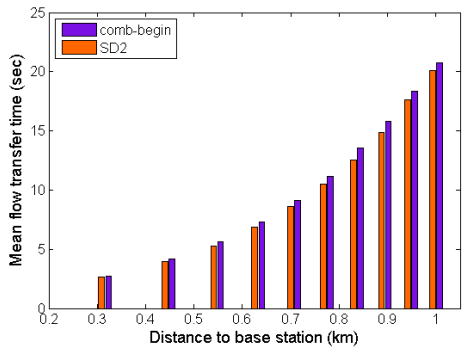


(b) High load

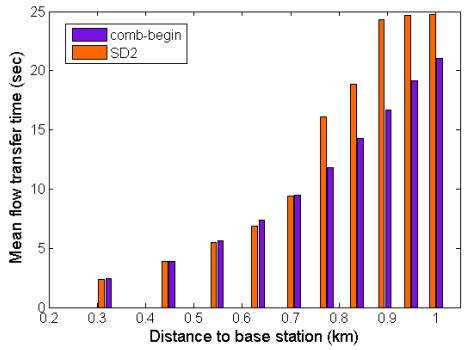
Figure 8.16: Mean flow transfer times for three evaluation scenarios - random-2, comb-begin and EC-comb for (a) low load (arrival rate $\lambda = 0.7$ flows/sec) and (b) high load (arrival rate $\lambda = 4$ flows/sec).

RBs, resulting in zero interference for remote users in the RC disappears at high loads.

We can conclude that the choice of spectrum end, where RB allocation in neighbouring cells starts, affects flow-level performance and can deliver certain gains. However, these gains are mostly observed for lightly loaded system and decrease as the system load increases.



(a) Low load



(b) High load

Figure 8.17: Mean flow transfer times for SD1 and SD2 spectrum division approaches for (a) low load (arrival rate $\lambda = 0.7$ flow/sec) and (b) high load (arrival rate $\lambda = 4$ flows/sec for SD1 and $\lambda = 3$ flows/sec for SD2).

8.6.4 Impact of spectrum division

In this section we investigate the implications of spectrum division for performance. The SD1 approach is compared to the SD2 approach for low and high loads. Again as in Section 8.6.3 for the SD1 approach we have chosen the comb-begin scenario. The comb-begin approach applies a CE-begin ordering in the RC and an EC-begin ordering in the NC. Recall that for the SD2 approach the following scenario is chosen - the RC (in both inner and outer area) adopts an

EC-begin and the NC (also for both areas) adopts EC-end. Hence, the SD2 scenario is equivalent to an SD1 EC-comb scenario but adapted to fit the pre-set spectrum division in two. For both scenarios it holds that cell edge users in the RC experience potential interference from centre users in the NC, which gives a common ground for comparison.

In Figure 8.17 we have plotted the mean flow transfer times for each scheme for two loads - low and high. We start with discussion of performance for low load, i.e. $\lambda = 0.7$ flows/sec, see Figure 8.17(a). We observe that the SD2 approach slightly outperforms the SD1 comb-begin approach. This behaviour is the combined result of the exclusive frequency allocation (characteristic for SD2) and the advantageous choice of ordering policy (characteristic for EC-comb). As a result under low system load users in the reference cell only occasionally experience interference from users in the neighbour cell. On the contrary, in the SD1 comb-begin case since both cells - RC and NC - start RB allocation from the same end of the spectrum, RC transmissions are more often affected by interference from the NC.

Despite the more advantageous ordering policy, the SD2 scenario worsen in performance for high system load, shown in Figure 8.17(b). The arrival rate is chosen such that it is close to the maximum supported load by the system. Already at this point we can detect that the SD2 approach has worse performance - the maximum arrival rate supported by the scheme with SD2 is $\lambda = 3$ flows/sec while for the scheme with SD1 this is $\lambda = 4$ flows/sec. Hence, a scheduler with SD2 spectrum division has lower capacity than a scheduler implementing an SD1 division.

The dissatisfying performance of SD2 is especially well expressed for users in the outer area in the RC, for which a drastic increase in mean flow transfer times is observed. Recall that the outer area starts beyond 0.7 km. The poor performance of the SD2 approach is the result of its inefficient spectrum usage with many active users, inherent to its design as we suggested in Section 8.2.2. More specific, the pre-set division of the spectrum in sub-bands causes resources in one sub-band to be left unused while another sub-band may be able to use them. This can be improved if a dynamic spectrum division is applied where e.g. the size of the spectrum portions depends on the load in the particular cell area.

8.6.5 Concluding remarks

In summary the most prominent conclusions derived in this sections are:

- A pre-set spectrum division, e.g. SD2, is not beneficial in terms of flow-level performance due to the inefficient resource utilisation in the different spectrum parts. This is in turn a result of the different number of users in the cell areas corresponding to these spectrum parts.

- The type of ordering policy does not impact flow-level performance significantly, regardless of the system load.

- Some performance gains can be achieved if ordering policies, which start resource allocation from the opposite ends of the spectrum, are adopted in neighbouring cells. These gains can be further extended, mainly for low system loads, if the ordering policy also uses knowledge on the spectrum range where resource blocks are left unused, e.g. EC-comb scenario.

Note that all concussions made hold given the specifically applied scheduling strategy. They should be treated with care and not generalised. An FFA scheduling scheme which allocates more than one RB per user or a more complex scheduler will lead to other results. We expect that for such schemes the ordering policy will have much bigger impact.

Next, we want to discuss several practical aspects concerning the applicability of the investigated spectrum division approaches. Although resource allocation schemes with frequency division of one are well-known by operators the introduction of a specific ordering policy brings several complications. In fact the base station needs information on the location (represented actually by a received power) of the active users and should adapt the ordering accounting for this information. This needs to be done every scheduling period, i.e. TTI. Given our findings that the choice of user ordering does not considerably affect performance, a random manner of service allocation may be more attractive. Such conclusions should be carefully drawn since we have only investigated a scheduling scheme with a single RB allocated per user. The ordering policy may have other effects on other scheduling strategies, e.g. higher RB allocation per user.

Finally, some remarks on the spectrum division of two approach. An SD2 approach is not entirely appropriate for practical use. In practice at least a spectrum division of three is necessary. However, this has implications for performance - we already showed for the SD2 approach that a pre-set fixed spectrum

division is inefficient in the utilisation of RBs; an effect that we expect to exhibit even more as the spectrum is further divided.

8.7 Conclusions

The research presented in this chapter was dedicated to the analysis of the inter-cell interference process in an LTE uplink and its dependency on the scheduling schemes. Additionally, we also evaluated the effects of the scheduler-specific interference on the flow-level performance. Our objectives were achieved by applying a hybrid analysis approach, which combines analytical evaluation at packet level with simulation at flow level.

A fair access scheduling scheme was considered, which allocates a single scheduling unit per user. In particular, we were interested in two aspects of the scheduler - how it maps the available spectrum to the active users, i.e. spectrum division approach, and in what order users are scheduled for service, i.e. ordering policy. In terms of spectrum division we evaluated two approaches - one, in which the whole spectrum is available to serve the active users, and another, in which the spectrum is divided into sub-bands used to serve users from particular cell areas. In terms of ordering strategy, we discussed five different ways to organise the users for service, accounting for both the users' location in the cell and the spectrum allocation.

The results show that the type of the ordering policy significantly affects the inter-cell interference pattern. However, the flow-level performance of the users is not very sensitive to this policy specific inter-cell interference. The applied spectrum division approach showed to be more important for the performance at flow level. More specific, we showed that exclusively allocating parts of the spectrum to a particular cell area is not beneficial at all; on the contrary it even causes degradation in performance when compared to a non-division approach.

Our findings should be carefully understood and should not be generalised but related to the particular scheduler. We can conclude that the ordering policy affects performance but to what extends depends on the scheduler. This knowledge should be used when trying to design a system with inter-cell interference coordination.

Furthermore, we demonstrated that by including in the analysis the impact of flow dynamics we were able to show that certain factors that exhibit strongly at packet level do not affect the eventual performance on a longer time scale (at flow level). In particular, while the specific user ordering seems to have significant impact on the inter-cell interference patterns it does not lead to big

differences in performance at flow level (in terms of mean flow transfer times).

9 Concluding remarks

The focus of this thesis was on uplink packet scheduling in current and future mobile cellular networks. To be more specific we were particularly interested in how scheduling can contribute to improving the service quality experienced by the users while trying to increase network efficiency. Furthermore, we also considered the topic of relaying. In particular, we investigated the performance benefits that relaying has to offer when combined with appropriate scheduling.

A distinct aspect of our research, which plays an important role throughout the whole thesis, is the analysis of the system performance at flow level taking into account the random behaviour of the mobile users regarding the initiation and completion of data flow transfers. We were able to evaluate this flow-level performance thanks to a novel hybrid analysis approach, which we consider to be an overall contribution of this thesis. The approach is a combination of analytical methods and simulation. Packet-level details, playing on a small time scale, are captured by analytical methods, while the flow-level behaviour, playing on a larger time scale, is simulated. This combination enables fast evaluation and comparison of performance measures such as flow transfer times for different schedulers. Another advantage of the approach is its flexibility - it can be used to model diversity of schedulers and environmental conditions as long as their specifics can be captured by the packet-level analysis.

The thesis was organised into three parts each covering a distinct research topic. The first part was dedicated to uplink scheduling in a UMTS/HSPA network. We determined how the performance of HSPA data users is affected

by other interfering transmissions inside the cell, e.g. higher-priority 'plain' UMTS voice calls, as well as transmissions from outside the cell, i.e. inter-cell interference. The focus of the second part was on analysing the potential benefits that relaying can bring to uplink performance when combined with an appropriate scheduling scheme. Finally, in part three we returned to pure uplink scheduling but this time in the context of an LTE network. In this third part we set up a sound base, providing understanding LTE scheduling specifics as well as their effects on inter-cell interference, which can support a subsequent study on optimising LTE uplink scheduling.

Throughout all three parts we consistently showed that by accounting for flow-level dynamics we were able to observe effects that often remain hidden when the system is studied solely at the packet level (as done in most other studies). One example are the studies dedicated to relaying, presented in Chapters 5 and 6. We have shown that relaying benefits not only users that transmit via a relay (visible at packet level) but also users that do not make use of relaying (visible only at flow level). Our analysis approach also enabled us to quantify these effects. Hence, by considering flow-level analysis the importance of relaying was further stressed. Another example is obtained from the studies on the impact of scheduling on inter-cell interference, see Chapters 4 and 8. Initially, during the analysis at packet level, it appeared that different scheduling schemes generate distinctive inter-cell interference patterns. However, interestingly, these distinctive patterns result in only marginally different performance when the flow level is considered.

Based on the performed research, our general conclusion is that flow dynamics can have significant impact on performance and should therefore be included in the analysis of mobile systems like the ones considered in this thesis. The proposed hybrid analytical/simulation approach capturing system behaviour at both packet and flow level is very flexible and allows changes in factors such as environmental conditions or technological specifications to be easily introduced in the analysis. These features make the approach very appropriate for application in other studies as well. Below we present several related topics for future research, which are in our view not only interesting but also considered to be very timely in the research of mobile networks.

Although our research on inter-cell interference in LTE was very insightful, as we mentioned, it only established a base for the evaluation of more complex scheduling schemes for interference mitigation. Inter-cell interference mitigation is an attractive topic, which has received and continues to receive much attention of the research community in mobile networks. Despite the various available studies on (optimal) scheduling schemes, which are designed to minimise inter-

cell interference, studies that provide performance evaluation on the flow level are not easily found. Hence, it is interesting to set up a research on the topic.

Looking back at the topics considered in this thesis, an investigation on the applicability of relaying for LTE networks for capacity enhancement is a logical choice for further research. Relaying is considered by 3GPP to be included in future releases dedicated to LTE Advanced. In the area of relay deployment for LTE networks two aspects require extra attention - (i) the combined impact of scheduling and relaying on performance and (ii) analysis of performance at flow level. Eventually, special attention needs to be dedicated also to investigating several deployment issues most notably to determining the optimal position of a relay within the cell.

Lastly, it would be interesting to study scheduling in the context of LTE networks with co-existing traffic types of real-time and non real-time nature. As we commented in the introduction of the thesis, future mobile networks should efficiently support a variety of services with distinct traffic characteristics and quality of service requirements.

Bibliography

[1] 3GPP TS 25.308. High Speed Downlink Packet Access (HSDPA); Overall Description, 2004.

[2] 3GPP TS 25.309. FDD Enhanced Uplink; Overall Description, 2006.

[3] 3GPP TS 36.942. LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Frequency (RF) system scenarios, 2010.

[4] B. Abendroth, J. L. van den Berg, and M. Mandjes. A multiple time-scale model for TCP bandwidth sharing under user heterogeneity. In *Proc. of Networking 2005, Waterloo, Canada*, 2005.

[5] M. Al-Rawi, R. Дntti, J. Torsner, and M. Sagfors. On the performance of heuristic opportunistic scheduling in the uplink of 3G LTE networks. In *Proc. of PIMRC 2008, Cannes, France*, 2008.

[6] D. Astely, E. Dahlman, A. Furuskar, Y. Jading, M. Lindstrom, and S. Parkvall. LTE: the evolution of mobile broadband. *IEEE Communications Magazine*, 47(4):44 –51, 2009.

[7] A. Bel, J. Lopez Vicario, and G. Seco-Granados. The benefits of relay selection in WiMAX networks. In *Proc. of ICT-MobileSummit 2008, Stockholm, Sweden*, 2008.

[8] T. Beniero, S. Redana, J. Hamalainen, and B. Raaf. Effect of relaying on coverage in 3GPP LTE-Advanced. In *Proc. of VTC 2009 (Spring), Barcelona, Spain*, 2009.

[9] T. Bonald and N. Hegde. Capacity gains of some frequency reuse schemes in OFDMA networks. In *Proc. of Globecom 2009, Honolulu, Hawaii, USA*, pages 1–6, 2009.

[10] J. W. Cohen. The multiple phase service network with generalized processor sharing. In *Acta Informatica*, volume 12, pages 254–284, 1979.

[11] B. Da and C. Ko. Dynamic subcarrier sharing algorithms for uplink OFDMA resource allocation. In *Proc. of ICICS 2007, Singapore-city, Singapore*, pages 1–5, 2007.

[12] D. C. Dimitrova, H. van den Berg, G. Heijenk, and R. Litjens. Impact of inter-cell interference on flow level performance of scheduling schemes for the UMTS EUL. In *Proc. of WiMob 2008, Avignon, France*, 2008.

[13] D. C. Dimitrova, H. van den Berg, G. Heijenk, and R. Litjens. Flow-level performance comparison of packet scheduling schemes for UMTS EUL. In *Proc. of WWIC 2008, Tampere, Finland*, 2008.

[14] D. C. Dimitrova, H. van den Berg, and G. Heijenk. Analyzing the impact of relay station characteristics on uplink performance in cellular networks. In *Proc. of ERCIM eMobility workshop 2009, Enschede, The Netherlands*, 2009.

[15] D. C. Dimitrova, H. van den Berg, and G. Heijenk. Performance of relay-enabled uplink in cellular networks - a flow level analysis. In *Proc. of ICUMT 2009, St. Petersburg, Russia*, 2009.

[16] D. C. Dimitrova, H. van den Berg, and G. Heijenk. Scheduler dependent modelling of inter-cell interference in UMTS EUL. In *Proc. of NGMAST 2008, Cardiff, UK*, 2009.

[17] D. C. Dimitrova, H. van den Berg, and G. Heijenk. Performance analysis of uplink packet schedulers in cellular networks with relaying. In *Proc. of WMNC 2009, Gdansk, Poland*, 2009.

[18] D. C. Dimitrova, H. van den Berg, G. Heijenk, and R. Litjens. Scheduling strategies for LTE uplink with flow behaviour analysis. In *Proc. of ERCIM eMobility workshop 2010, Lulea, Sweden*, 2010.

[19] A. M. El Hajj, E. Yaacoub, and Z. Dawy. On uplink OFDMA resource allocation with ergodic sum-rate maximization. In *Proc. of PIMRC 2009, Tokyo, Japan*, 2009.

[20] G. Fodor. Performance analysis of a reuse partitioning technique for OFDM based evolved UTRA. In *Proc. of IWQoS 2006, New Haven, USA*, pages 112 –120, 2006.

[21] G. Fodor and M. Telek. Performance analysis of the uplink of a CDMA cell supporting elastic services. In *Proc. of Networking 2005, Waterloo, Canada*, 2005.

[22] A. Furuskar, T. Jonsson, and M. Lundevall. The LTE radio interface - key characteristics and performance. In *Proc. of PIMRC 2008, Cannes, France*, pages 1–5, 2008.

[23] L. Gao and S. Cui. Efficient subcarrier, power and rate allocation with fairness considerations for OFDMA uplink. *IEEE Transactions on Wireless Communications*, 7:1507–1511, 2008.

[24] K. Georgiev and D. C. Dimitrova. Performance analysis of uplink packet schedulers in cellular networks with relaying. In *Proc. of EW 2010, Lucca, Italy*, 2010.

[25] M. J. Hart and S. Vadgama. Dimensioning and system level analysis of an HSDPA network with relaying nodes. In *Proc. of VTC 2005 (Fall), Dallas, USA*, pages 2561–2565, 2005.

[26] B. Haverkort. *Performance of computer communication systems.* John Wiley & Sons Ltd, 1999.

[27] K. W. Helmersson, E. Englund, M. Edvardsson, C. Edholm, S. Parkvall, M. Samuelsson, Y-P. E. Wang, and J-F. Cheng. System performance of WCDMA enhanced uplink. In *Proc. of VTC 2005 (Spring), Stockholm, Sweden*, 2005.

[28] H. Holma and A. Toskala. *WCDMA for UMTS.* John Wiley & Sons Ltd, 2001.

[29] H. Holma and A. Toskala. *HSDPA/HSUPA for UMTS.* John Wiley & Sons Ltd, 2006.

[30] H. Holma and A. Toskala. *LTE for UMTS: OFDMA and SC-FDMA based radio access.* John Wiley & Sons Ltd, 2009.

[31] R. Hu, S. Sfar, G. Charlton, and A. Reznik. Protocols and system capacity of relay-enhanced HSDPA systems. In *Proc. of CISS 2008, Princeton, USA*, pages 179–184, 2008.

[32] J. Huang, V. G. Subramanian, R. Agrawal, and R. Berry. Joint scheduling and resource allocation in uplink OFDM systems for broadband wireless access networks. *IEEE Journal on Selected Areas in Communications*, 27: 226–234, 2009.

[33] R. Jain. *The Art of Computer Systems Performance Analysis: techniques for experimental design, measurement, simulation, and modeling.* John Wiley & Sons Ltd, 1991.

[34] K. Kim, Y. Han, and S. L. Kim. Joint subcarrier and power allocation in uplink OFDMA systems. *IEEE Communications Letters*, 9:526–52, 2005.

[35] K. Kumaran and L. Qian. Uplink scheduling in CDMA packet-data systems. In *Proc. of Wireless Networks*, volume 12, pages 33–43, 2006.

[36] R. Kwan, C. Leung, and J. Zhang. Multiuser scheduling on the downlink of an LTE cellular system. *Research Letters in Communications*, 2008:1–4, 2008.

[37] E. Lang, S. Redana, and B. Raaf. Business impact of relay deployment for coverage extension in 3GPP LTE-Advanced. In *Proc. of ICC Workshops 2009, Dresden, Germany*, pages 1–5, 2009.

[38] S. B. Lee, I. Pefkianakis, A. Meyerson, S. Xu, and S. Lu. Proportional fair frequency-domain packet scheduling for 3GPP LTE uplink. In *Proc. of Infocom 2009 mini-symposium, Rio de Janeiro, Brazil*, 2009.

[39] C. Li and S. Papavassilious. On the fairness and throughput trade-off of multi-user uplink scheduling in WCDMA systems. In *Proc. of VTC 2005 (Fall), Dallas, USA*, 2005.

[40] T. Liu and D. Everitt. Analytical approximation of other-cell interference in the uplink of CDMA cellular systems. In *Proc. of VTC 2006 (Spring), Melbourne, Australia*, volume 2, pages 693–697, 2006.

[41] T. Liu, A. Mäder, and D. Staehle. Analytical other-cell interference characterization over HSUPA-enabled multi-cell UMTS networks. In *Proc. of VTC 2007 (Fall) Baltimore, USA*, 2007.

[42] T. Liu, A. Mäder, D. Staehle, and D. Everitt. Analytic modeling of the UMTS enhanced uplink in multi-cell environments with volume-based best-effort traffic. In *Proc. of ISCIT 2007, Sydney, Australia*, 2007.

[43] A. Mäder and D. Staehle. An analytical model for best-effort traffic over the UMTS enhanced uplink. In *Proc. of VTC 2006 (Fall), Montreal, Canada*, 2006.

[44] L.A. Maestro Ruiz de Temino, G. Berardinelli, S. Frattasi, and P. Mogensen. Channel-aware scheduling algorithms for SC-FDMA in LTE uplink. In *Proc. of PIMRC 2008, Cannes, France*, 2008.

[45] X. Mao, A. Maaref, and K. H. Teo. Adaptive soft frequency reuse for inter-cell interference coordination in SC-FDMA based 3GPP LTE uplinks. In *Proc. of Globecom 2008, New Orleans, Louisiana, USA*, pages 1–6, 2008.

[46] P. Mogensen, W. Na, I. Z. Kovacs, F. Frederiksen, A. Pokhariyal, K.I. Pedersen, T. Kolding, K. Hugl, and M. Kuusela. LTE capacity compared to the Shannon bound. In *Proc. of VTC 2007 (Spring), Dublin, Ireland*, pages 1234–1238, 2007.

[47] H. G. Myung, J. Lim, and D. J. Goodman. Single carrier FDMA for uplink wireless transmission. *IEEE Vehicular Technology Magazine*, Vol. 48:30–38, 2006.

[48] C. Ng and C. Sung. Low complexity subcarrier and power allocation for utility maximization in uplink OFDMA systems. *IEEE Transactions on Wireless Communications*, Vol. 7:1667–1675, 2008.

[49] R. Nunez Queija, J. L. van den Berg, and M. R. H. Mandjes. Performance evaluation of strategies for integration of elastic and stream traffic. In *Proc. of ITC 16, Edinburgh, UK*, pages 1039–1050, 1999.

[50] R. Pabst, B. Walke, D. Schultz, and et al. Relay-based deployment concepts for wireless and mobile broadband radio. *IEEE Communications Magazine*, pages 80–89, 2004.

[51] S. Parkvall and D. Astely. The evolution of LTE towards IMT-advanced. *Journal of Communications*, 4, 2009.

[52] A. Racz, N. Reider, and G. Fodor. On the impact of inter-cell interference in LTE. In *Proc. of Globecom 2008, New Orleans, USA*, pages 1 –6, nov. 2008.

[53] S. Ramakrishna and J. M. Holtzman. A scheme for throughput maximization in a dual-class CDMA system. In *Proc. of ICUPC 1997, San Diego, USA*, 1997.

[54] E. Reetz, R. Hockmann, and R. Tonjes. Performance study on cooperative relaying topologies in beyond 3G systems. In *Proc. of ICT-MobileSummit 2008, Stockholm, Sweden*, 2008.

[55] C. Rosa, J. Outes, T.B. Sorensen, J. Wigard, and P. E. Mogensen. Combined time and code division scheduling for enhanced uplink packet access in WCDMA. In *Proc. of VTC 2004 (Fall), Los Angeles, USA*, 2004.

[56] S. M. Ross. *Introduction to probability models*. Academic press - An imprint of Elsevier, 2003.

[57] J. Schiller. *Mobile Communications*. Pearson Education Limited, 2003.

[58] R. Schoenen, R. Halfmann, and B. H. Walke. An FDD multihop cellular network for 3GPP-LTE. In *Proc. of VTC 2008 (Spring), Singapore-city, Singapore*, pages 1990–1994, 2008.

[59] R. Schoenen, W. Zirwas, and B. H. Walke. Raising coverage and capacity using fixed relays in a realistic scenario. In *Proc. of EW 2008, Prague, Czech Republic*, pages 1–6, 2008.

[60] M. Schwartz. *Mobile Wireless Communications*. Cambridge university press, 2005.

[61] S. Sengupta, M. Chatterjee, S. Ganguly, and R. Izmailov. WRN: improving system performance in 3g networks through fixed multi-hop relay nodes. In *Proc. of WCNC 2005, New Orleans, USA*, volume 3, pages 1708–1713, 2005.

[62] S. Sesia, I. Toufik, and M. Baker. *LTE, The UMTS Long Term Evolution: From Theory to Practice*. Wiley Publishing, 2009.

[63] D. Staehle, K. Leibnitz, K. Heck, B. Schröder, A. Weller, and P. Tran-Gia. Approximating the Othercell Interference Distribution in Inhomogeneous UMTS Networks. In *Proc. of VTC 2002 (Spring), Birmingham, USA*, 2002.

[64] M. Umlauft. Relay devices in UMTS networks - effects on application performance. In *Proc. of Med-Hoc-Net 2006, Lipari, Italy*, 2006.

[65] J. Vidal, O. Munoz, A. Agustin, E. Calvo, and A. Alcon. Enhancing 802.16 networks through infrastructure-based relays. In *Proc. of ICT-MobileSummit 2008, Stockholm, Sweden*, 2008.

[66] J. Voigt and K. Pannhorst. Optimizations on scheduling strategies for enhanced uplink on WCDMA. In *Proc. of VTC 2007 (Spring), Dublin, Ireland*, pages 1172–1176, 2007.

[67] L. Wang, Y. Ji, and F. Liu. A novel centralized resource scheduling scheme in OFDMA-based two-hop relay-enhanced cellular systems. In *Proc. of WiMob 2008, Avignon, France*, pages 113 –118, 2008.

[68] H.-Y. Wei, S. Ganguly, and R. Izmailov. Ad hoc relay network planning for improving cellular data coverage. In *Proc. of PIMRC 2004, Barcelona, Spain*, volume 2, pages 769 – 773, 2004.

[69] E. Weiss, S. Max, O. Klein, G. Hiertz, and B. Walke. Relay-based vs. Conventional Wireless Networks: Capacity and Spectrum efficiency. In *Proc. of PIMRC 2007, Athens, Greece*, pages 1–5, 2007.

[70] M. Wener, P. Moberg, and P. Skillermark. Cost assessment of radio access network deployment with relay nodes. In *Proc. of ICT-MobileSummit 2008, Stockholm, Sweden*, 2008.

[71] C. Wengerter, J. Ohlhorst, and A. G. E. von Elbwart. Fairness and throughput analysis for generalized proportional fair frequency scheduling in OFDMA. In *Proc. of VTC 2005 (Spring), Stockholm, Sweden*, 2005.

[72] M. Wodczak. Cooperative relaying in an indoor environment. In *Proc. of ICT-MobileSummit 2008, Stockholm, Sweden*, 2008.

[73] X. Xiang, F. Liu, and Y. Ji. Simulation based performance evaluation of ICI mitigation schemes for broadband wireless access networks. In *Proc. of CNS 2008, Montreal, Canada*, pages 181–187, 2008.

[74] F. Xiangning, C. Si, and Z Xiaodong. An inter-cell interference coordination technique based on users' ratio and multi-level frequency allocations. In *Proc. of WiCom 2007, Shanghai, China*, pages 799–802, 2007.

[75] Z. Xie and B. Walke. Resource allocation and reuse for inter-cell interference mitigation in OFDMA based communication networks. In *Proc. of WICON 2010, Singapore-city, Singapore*, pages 1–6, 2010.

[76] E. Yaacoub and Z. Dawy. Centralized and distributed LTE uplink scheduling in a distributed base station scenario. In *Proc. of ACTEA 2009, Zouk Mosbeh, Lebanon*, 2009.

[77] E. Yaacoub, H. Al-Asadi, and Z. Dawy. Low complexity scheduling algorithms for the LTE uplink. In *Proc. of ISCC 2009, Sousse, Tunisia*, pages 266–270, 2009.

[78] Y. Yu, S. Murphy, and L. Murphy. Interference aware relay station location planning for ieee 802.16j mobile multi-hop relay network. In *Proc. of PM2HW2N 2009, Tenerife, Canary Islands, Spain*, pages 201–208, 2009.

[79] T. Zhang, Z. Zeng, C. Feng, J. Cheng, and L. Song. Uplink power allocation for interference coordination in multi-cell OFDM systems. In *Proc. of ChinaCom 2008, Hangzhou, China*, pages 716–720, 2008.

Acronyms

AMPS	Advanced Mobile Phone System
BS	Base Station
CDF	Cumulative Distribution Function
CDMA	Code Division Multiple Access
CTMC	Continuous Time Markov Chain
DCH	Dedicated Channel
EDCH	Enhanced Dedicated Channel
EDGE	Enhanced Data Rates for GSM Evolution
EPC	Evolved Packet Core
EUL	Enhanced Uplink
E-UTRAN	Evolved UTRAN
FDD	Frequency Division Duplex
FDMA	Frequency Division Multiple Access
GGSN	Gateway GPRS Support Node
GMSC	Gateway MSC
GPRS	General Packet Radio Service
GSM	Global System for Mobile communications
HLR	Home Location Register
HSDPA	High Speed Downlink Packet Access
HS-DSCH	High Speed Downlink Shared Channel
HSPA	High Speed Packet Access
HSUPA	High Speed Uplink Packet Access
HSS	Home Subscriber Server

ICIC	Inter-Cell Interference Coordination
LTE	Long Term Evolution
MME	Mobility Management Entity
MS	Mobile Station
MSC	Mobile Switching Centre
NMT	Nordic Mobile Telephony
OBO	One-by-one
OFDMA	Orthogonal Frequency Division Multiple Access
PCRF	Policy and Charging Resource Function
PDCCH	Physical Downlink Control CHannel
P-GW	Packet Data Network Gateway
QAM	Quadrature Amplitude Modulation
QoS	Quality of Service
RB	Resource Block
RNC	Radio Network Controller
RNS	Radio Network Subsystem
RR	Round Robin
RRM	Radio Resource Management
RS	Relay Station
SAE	System Architecture Evolution
SC-FDMA	Single Carrier Frequency Division Multiple Access
SINR	Signal to Interference plus Noise Ratio
SGSN	Serving GPRS Support Node
S-GW	Serving Gateway
SMS	Short Message Service
TDD	Time Division Duplex
TDMA	Time Division Multiple Access
TTI	Transmission Time Interval
UMTS	Universal Mobile Telecommunications System
UTRAN	UMTS Terrestrial Radio Access Network
WCDMA	Wideband Code Division Multiple Access

About the author

I, Desislava Cvetanova Dimitrova (Десислава Цветанова Димитрова), was born on the 20th of September 1981 in the beautiful city of Plovdiv in Bulgaria. My professional education started at the Technical University of Sofia (Bulgaria), where I completed my Bachelor of Science study in *Telecommunications* in 2003. My study *Telematics*, which lead to my Master of Science degree, was completed in 2006 at the University of Twente (The Netherlands). I continued in the same year to pursue my Doctor of Philosophy degree again at the University of Twente as part of the Design and Analysis of Communication Systems research group. My project was a common initiative of the University of Twente and TNO in the framework of the expertise centre of e-Quality.

Throughout the years of my PhD I not only further developed my interest in mobile networks (which was the focus of my research) but also broadened it towards sensor and vehicular networks. The results of these interests are, among others, the two editions of the Workshop on Pervasive Application of Wireless Technologies, which I organised in 2009 and 2010. I have also participated in the COST 290 project and collaborated with TNO researchers in the Socrates project. Further, a detailed list of my publications in chronological order is given below. Two of these publications were given a best paper award.

- Dimitrova D.C., van den Berg H., Heijenk G. and Litjens R. *Flow level performance comparison of packet scheduling schemes for EUL UMTS* In: Proceedings of the 6th International Conference on Wired/Wireless Internet Communications, 28-30 May 2008, Tampere, Finland. pp. 27-40. Lecture Notes in Computer Science LNCS 5031. Springer Verlag.
  Best paper award

- Dimitrova D.C., van den Berg H., Heijenk G. and Litjens R. *Impact of inter-cell interference on flow level performance of scheduling schemes for the UMTS EUL* In: Proceedings of the 4th IEEE International Conference on Wireless and Mobile Computing, Networking and Communications, 12-14 October 2008, Avignon, France. pp. 106-112. IEEE Computer Society Press.

- Dimitrova D.C., van den Berg H. and Heijenk G. *Analysing the impact of relay station characteristics on uplink performance in cellular networks* In: Proceedings of the 3rd ERCIM Workshop on eMobility, 27-28 May 2009, Enschede, The Netherlands. pp. 31-42. ISBN 978-90-365-2846-7.

- Dimitrova D.C., van den Berg H. and Heijenk G. *Performance analysis of uplink packet schedulers in cellular networks with relaying* In: Proceedings of the 2nd Joint IFIP Wireless and Mobile Networking Conference, 09-11 September 2009, Gdansk, Poland. pp. 263-273. IFIP Advances in Information and Communication Technology 308. Springer Verlag.

- Dimitrova D.C., van den Berg H. and Heijenk G. *Scheduler dependent modeling of inter-cell interference in UMTS EUL* In: Proceedings of the 3rd IEEE International Conference on Next Generation Mobile Appliation, Services and Technologies, 15-18 September 2009, Cardiff, Wales, UK. pp. 259-264. IEEE Computer Society Press.

- Dimitrova D.C., van den Berg H. and Heijenk G. *Performance of relay-enabled uplink in cellular networks - a flow level analysis* In: Proceedings of the 1st International Conference on Ultra Modern Communications, 12-14 October 2009, St Petersburg, Russia. pp. 1-8. IEEE Computer Society Press.
  Best paper award

- Georgiev K. and Dimitrova D.C. *Impact of relaying on inter-cell interference in mobile cellular networks* In: Proceedings of 16th European Wireless, 12-15 April 2010, Lucca, Italy. pp. 398-405. IEEE Computer Society Press.

- Fernandez Diaz I., Litjens R., van den Berg H., Dimitrova D. and Spaey K. *Sensitivity analysis of the optimal parameter settings of an LTE packet scheduler* In: Proceedings of VTC Spring 2010, 16-19 May 2010, Taipei, Taiwan. pp. 1-6. IEEE Computer Society Press.

- Dimitrova, D.C., van den Berg, J.L., Heijenk, G. and Litjens, R. *Scheduling strategies for LTE uplink with flow behaviour analysis* In: Fourth ERCIM workshop on eMobility, Lulea, Sweden. pp. 15-26. ISBN 978-91-7439-103-9.

- Dimitrova D.C., van den Berg H. and Heijenk G. *Uplink packet scheduling in cellular networks with relaying - comparative study* In: Special Issue of Telecommunication Systems (accepted for publication).

- Dimitrova, D.C. and van den Berg, J.L., Heijenk, G. and Litjens, R. *Analysis of packet scheduling for UMTE EUL - design decisions and performance evaluation* In: Journal of Internet Engineering (submitted for publication).

- Dimitrova, D.C. and van den Berg, J.L., Heijenk, G. and Litjens, R. *LTE uplink scheduling - flow level analysis* In: Networking 2011 (submitted for publication).

Acknowledgements

Dear reader, this thesis described the work I did in the four years of my PhD. Without a doubt I learnt how to approach a research topic, I considerably extended my technical vocabulary and I mastered the skill to form sentences undecipherable for the commoners. Yet, my biggest achievement lie beyond the written word. I broadened my horizons. I learnt to cherish myself and others. I became a better person. All these I owe to ...

Beste Hans, mijn grootste geluk tijdens de PhD is mijn supervisor geweest. Waarschijnlijk kan ik niet alle aspecten beseffen waarmee je me veranderd hebt. Van jou heb ik geleerd beter te schrijven, aandacht zelfs aan de details te geven, altijd voor kwaliteit te gaan. Ik ben dankbaar voor alle tijd en aandacht die je in mij geïnvesteerd hebt. De grootste les die ik van jou geleerd heb is dat men zijn beste spiegel is. Ik zal trots zijn als ik in de toekomst zo een volledig professionalist en begeleider als jij kan zijn.

Ik wil ook Geert en Remco bedanken als de mensen die dichtbij mijn onderzoek hebben gestaan en met wie ik samen werk gedaan heb waar ik trots op ben. Ieder heeft bijgedragen aan mijn professionele ontwikkeling en heeft me veel geleerd.

I would also like to thank my other committee members for the time and attention they gave me and I can only hope that being in my committee has been a pleasant experience.

My source of pride, my students! Thank you all for enduring my comments and persistence in doing better work; for your desire to learn; for teaching me how to be a better supervisor. Most of all thank you for being my friends and for

your appreciation. It means a lot! Ambes, you were the best start a PhD can ask for - enthusiastic, eager to know more and responsible. I am sure that you have a great future ahead and I can not wait to see it. На българските еразмусчета – Камен и Стоян – едно голямо благодаря. Момчета, с вас свършихме чудесна работа, с която се гордея като изследовател и ръководител. Надявам се и за вас съвместната ни дейност да е била ползотворна и да е оставила приятни и незабравими преживявания независимо от моя постоянен стремеж за повече и по-добре. На нашето бисерче Криси благодаря за голямото внимание и емоционалната подкрепа. У всеки един от вас намерих незаменим приятел.

In kamer 5022 werd er meest gehoord ... Marijn. Jij bent een integraal deel uit mijn PhD geworden. Mijn progress en succes is ook aan jou te danken - al onze discussies met wiskundige arde hebben me erg veel gehoplen. Ik heb van jou ook geleerd om beter te luisteren en minder emotioneel problemen te benaderen. Ik kan niet genoeg worden vinden om jou te bedanken.

Dearest girls, Anna and Yimeng, my good friends, I have been fortunate to share a room with you. Every day I spent at work has been a pleasant experience majorly thanks to you and Marijn. I enjoyed all the professional and personal conversations we had and I hope that wherever life may take us, we will always stay close. Cara ricciolina, tantissime grazie per le innumerevoli conversazioni e tutto cio che hai fatto per me. E per tutte le volte che mi hai salvata da me stessa.

Dear DACS colleagues but above all friends, each one of you have touched me in a special way and left his or her personal imprint on me. You also made my work my second family, which is a rare jewel. For this I can only be grateful to you Tiago, Anne, Ramin, Fei, Rafael, Idilio, Giovane, Anja, Martijn, Wouter, Daniel and Stephan.

To my other DACS colleagues - Boudewijn, Aiko, Georgios and Pieter-Tjerk - I am grateful for all professional advices. Each of you, unique in your own way, showed me the skills that an accomplished professional should possess.

Nirvana, thank you for all your support and believing in an idea. I enjoyed every moment working with you. I hope our endeavours in the wireless world will continue in the future.

Beste Jan, ik ben je zo dankvaar voor de enthousiasme die je in het Erasmus programa gestopt hebt. Ik geloof dat we iets zeer waardevol gedaan hebben. Hopelijk hebben we in de toekomst nog andere kansen om weer iets goeds te doen.

To my other UT colleagues - Laura, Luis Olavo, Ramon and Rita - I have enjoyed a lot our various experiences ranging from cultural events over active housing sport to traveling accross Europe by train.

Caro Eduardo, a nossa amizade surgiu como uma surpresa para mim. Alguém que, inicialmente, pareceu pensar tão diferentemente, mostrou-se ser o mesmo filósofo que eu sou. Também gostaria de dizer que eu sou extremamente grata pelas conversas que nós tivemos - o meu universo interno ganhou uma supernova.

Иване, благодаря сърдечно за цялата ти помощ и подкрепа в израстването ми като всестранно развит изследовател. От теб получих много ценни съвети и напътствия. Незабравими ще останат приятните ни разговори над мисо супа.

Роси, дано семенцето, което посяхме и поливахме да продължи да расте. Благодаря ти също и за многото идеи и напътствия.

Lieve Sara, Menno en Almer, ik kan alleen dankbaar zijn dat ik vrienden als jullie heb. Door de jaren heen hebben jullie nast me gestaan in goede en slechte tijden. Ik hoop dat ik voor jullie ooit ook zo een goede vriend ben. Ik weet niet wat de toekomst te bieden heeft, maar ik hoop dat we elkaar niet uit het oog verliezen.

Ook vergeet ik Dennis en Gilly en Wico en Renske niet. Ondanks dat we elkaar niet heel vaak zien, hebben jullie altijd een oor voor mijn verhalen. Ik geniet ieder keer dat we samen afspreken.

Dushi, chikitin ta kla! Despite the circumstances in which we met knowing you has been a big event in my life. Rarely I have met someone so complex and so much alike me. From you I have not only learnt a lot about life and people but also about myself. The future, as full of uncertainties as it is, I hope will leave open the door to friendship. This star will always shine.

To my neighbours - Juan Carlos, Yuri, Yonathan, Yuda, Sander, Hanneke, Mohsen, Boyan, Vasko, Trajce, Paula, Gintas - discussing the twists of life over a cup of tea was very enjoyable and added to my daily routine the spice that tingles the brain. A special thanks to Sha - talking to you have thought me some vary valuable lessons.

Girly!!! You have been a fantastic neighbour and a very good friend. Thank you for all that you have shared with me and all the cups of tea you have offered me.

A special thank you also to Karen, Blas, Nico, Suen, Liga, Zdravko, Garabed, Edwin and Camilo! The friendship of each is a shinny pearl in the neckless of

my life.

To all my other friends that have been around me, independently of when or for how long, I am grateful to you for seeing in me an interesting person and for sharing parts of your lives with me.

Мили момичета - Донка, Милена, Мария и Ками - с вас сме минали през много житейски радости и неволи и винаги сме си давали рамо да се подкрепим. През годините аз се научих да ценя разговорите си с вас макар и понякога кратки. Вие винаги сте ми носили утеха, радост и убеждението, че има и прекрасни хора на тази земя. Продължавайте да сте такива искрящи перли!

Слънце, няма да те забравя, разбира се. Нашите разговори поддържат пламъка на дълбоко философското начало у мен. Благодаря ти за високата оценка и уважение, които винаги съм получавала.

Когато завършвахме училище, не можех да предполагам, че от всичките ми съученици именно вие, момчета, ще останете до мен през годините. Особено ти, Лазо, с твоето постоянство в нашите срещи. Вик, независимо от спорадичните ни разговори от теб съм научила много за човешката природа и винаги ще останеш мой добър приятел.

Lieve Simon en Adri, jullie zijn mijn tweede familie gewoorden; een familie waar ik erg gelukkig mee ben; een familie die altijd een goed advies kan geven. Jullie steun waardeeer ik veel! Jullie hebben altijd voor gezorgd dat we samen een gezellige tijd doorbrengen. Dank jullie wel!

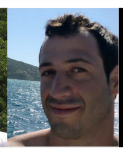
Хората, които винаги са стояли зад мен и са давали безграничната си обич, дори и когато не винаги са разбирали изборите ми са моето семейството. Мамо и тате, вие сте в основата на моя успех като човек и професионалист. С непрекъснатото си насърчаване, подкрепа и критика, когато е необходимо, вие допринесохте за моето израстване. Гордея се да имам такива несломими и даряващи родители. Надявам се и вие се гордеете с мен и вярвам знаете, че винаги ще съм до вас.

Буца, Фило, винаги съм била много щастлива, че те има и че се разбираме така добре. Ти си най-добрия ми приятел, към който винаги мога да се обърна за съвет, да си помрънкам или да ми стане весело. Независимо от километрите, които ни делят, винаги ще си бъдем кака и буца.

To all of you who taught my heart to cry and laugh, who tore down the barricade in my mind, who made me who I am ... a humble thank you!







frame 6