KTH Signals
Sensors and Systems

Demand Responsive Resource Management for Cellular Networks

Link Asymmetry, Pricing and Multihopping

MAGNUS LINDSTRÖM

Doctoral Thesis
Stockholm, Sweden 2005

TRITA-S3-RST-0509
ISSN 1400-9137
ISRN KTH/RST/R--05/09--SE

KTH Signaler, sensorer och system
SE-100 44 Stockholm
SWEDEN

Akademisk avhandling som med tillstånd av Kungl Tekniska högskolan framlägges till offentlig granskning för avläggande av teknologie doktorsexamen fredagen den 29 april 2005 klockan 14:00 i sal C1, Electrum, Kungl Tekniska högskolan, Isafjordsgatan 22, Kista.

Tryck: Universitetsservice US AB

Abstract

Economic affordability of services and infrastructures has rapidly become one of the key issues in the evaluation and design of wireless access systems. The provisioning of high data rates, at an "affordable" price, constitutes a serious challenge to the structure and management of current and future wireless networks.

The management of radio resources, *Radio Resource Management* or RRM for short, has traditionally been benchmarked mostly by technical merits such as throughput (data delivery capability) and Quality of Service (QoS). When comparing different RRM schemes, the scheme that can deliver more bits per Hertz (unit of bandwidth) or per Euro is often assumed the more efficient. From an economic point of view, however, cost efficiency is not equivalent to profitability.

We conjecture that the economic efficiency and profitability can be improved both by better technical efficiency and by better accounting for users' service appreciation and willingness to pay. While we shall, primarily treat the operator's benefit of improved RRM, we will try to improve the RRM by means of being more responsive to the demands of the users. In eight conference and journal papers, we investigate: Provisioning of support for asymmetric traffic, Quality and pricing aware resource management and Creation of forwarding incentive in multihop cellular networks.

We show that implementing support for asymmetric links can improve the efficiency of (service) production in Time Division Duplexing (TDD) mode wireless networks with asymmetric traffic. That is, more traffic can be handled with the same system resources. Compared to Frequency Division Duplexing (FDD), TDD offers more flexible use of spectrum resources. The benefits of TDD and support for asymmetric links are readily available for systems providing high-rate spotty coverage. For systems aiming at full coverage and tight reuse, however, proper measures must be taken to control inter-mobile- and inter-base-station-interference.

We present the MEDUSA model framework for taking users' service appreciation and willingness to pay into account in performance evaluations of wireless networks with elastic traffic. Assuming that user satisfaction depends on both the quality and the price of the service, numerical experiments show that the economic efficiency of an RRM scheme is affected by the pricing scheme. We also introduce the concepts of speculative resource management to exploit traffic elasticity and improve resource utilisation. With speculative admission control, users with good propagation conditions may be admitted to a full system at the expense of a slight degradation of the QoS of some or all users, if the expected total revenue would thereby increase. Results indicate significant revenue gain with speculative admission control. Service perception aware scheduling was evaluated as a means to improve resource utilisation, but yielded only marginal gain compared to a weighted proportional fair scheduler.

For the third area studied in this Thesis, i.e. multihopping in cellular networks, economic efficiency was both the goal and one of the means to achieve it. By means of a resource re-distribution scheme called Resource Delegation we eliminated the bandwidth bottle neck of the relays. We combined Resource Delegation with economic compensation for the energy expenditures of the relays and were able to achieve significantly increased operator revenue with maintained or improved user utility. Assuming that the added complexity of keeping track of reward transactions is negligible, profitability was correspondingly improved.

Contents

Contents v

List of Figures viii

I 1

1 Introduction 3
1.1 Scope of the Thesis . . . 4
1.2 Monte Carlo Simulations . . . 6
1.3 Thesis Outline . . . 7

2 Radio Access in Cellular Networks 9
2.1 Multiple Access . . . 9
2.2 Signal-to-Interference-Ratios . . . 12
2.3 Radio Resource Management . . . 12

3 Supporting Asymmetric Traffic 15
3.1 Related Work . . . 16
3.2 Contribution . . . 18
3.3 Dynamic Link Asymmetry (Paper T1) . . . 19
3.4 Link Gain Estimation (Paper T2) . . . 24
3.5 Heterogeneous Link Asymmetry (Paper T3) . . . 27
3.6 Sensitivity to BS Placement (Paper T4) . . . 30
3.7 Conclusions . . . 32
3.8 Discussion . . . 32

4 Micro-economics and Game Theory 35
4.1 Game Theory . . . 36

5 Quality and Pricing Aware RRM 39
5.1 Related Work . . . 39
5.2 Contribution . . . 42

5.3 MEDUSA (Paper P5) 42
5.4 MEDUSA in Hard- and Soft-Capacity Systems (Paper P6) 46
5.5 Speculative Resource Allocation (Paper P7) 48
5.6 Conclusions 54
5.7 Discussion 56

6 Creation of Forwarding Incentive **59**
6.1 Related Work 60
6.2 Contribution 60
6.3 Resource Delegation and Rewards (Paper I8) 61
6.4 Conclusions 67
6.5 Discussion 68

7 Summary and Future Work **69**
7.1 Future Work 70

Bibliography **73**

Appendix **81**

A TDD Models and Performance Measures **83**
A.1 Radio Network 83
A.2 Scenario 85
A.3 Propagation Models 87
A.4 Performance Measures 89
A.5 Summary of System Parameters 90

II Paper Reprints (in chronological order) **91**

8 Dynamic Link Asymmetry in 'Bunched' Wireless Networks (Paper T1) **93**

9 Improved TDD Resource Allocation through Inter-Mobile Interference Avoidance (Paper T2) **101**

10 Heterogeneous Link Asymmetry in TDD Mode Cellular Systems (Paper T3) **109**

11 Demand and Pricing Effects on the Radio Resource Allocation of Multimedia Communication Systems (Paper P5) **117**

12 An Economic Model for the Radio Resource Management in Multimedia Wireless systems (Paper P6) **125**

13 Base Station Placement in Asymmetric TDD Mode Systems in a Manhattan Environment (Paper T4) **137**

14 Resource Delegation and Rewards to Stimulate Forwarding in Multihop Cellular Networks (Paper I8) **145**

15 Speculative Admission Control and Scheduling for Packet-Switched Wireless Networks (Paper P7) **153**

List of Figures

2.1	Hybrid FDMA/TDMA (F/TDMA) access protocol.	10
2.2	Duplexing methods	10
2.3	TDMA/TDD frame with downlink and uplink time-slots.	11
2.4	Bunch with Central Unit and RAUs	13
3.1	Conflicting switching point assignments	16
3.2	IMI and IBI free system	17
3.3	Movable Boundary slot search algorithm	21
3.4	Soft Switching Point slot search algorithm	21
3.5	Link gains in and between cells i and j	25
3.6	Strongest Base Stations sets	27
3.7	Asymmetry scenarios	29
3.8	Two different BS deployments.	31
4.1	Supply and demand	36
4.2	Example of a sigmoid utility function.	37
5.1	Provider revenue for flat price, 160 users, as a function of the price p.	46
5.2	Provider revenue for price $p(r) = kr$, 160 users, as a function of k.	47
5.3	TDMA–CDMA comparison, price $p(r) = kr$, admission rate, $\vartheta = 1.5$	48
5.4	Illustration of the MEDUSA extension.	50
5.5	Goodput plotted versus dissatisfaction rate.	55
6.1	Multihop wireless network.	59
6.2	Illustration to definition of data rates in multihop wireless network.	62
6.3	TDMA subslotting.	63
6.4	Regions of joint revenue,interference,utility and data rate improvement.	66
A.1	16-slot TDMA frame.	84
A.2	Ten-by-ten-block Manhattan environment	86
A.3	Illustration to recursive path loss model	88

Part I

Chapter 1

Introduction

Economic affordability of services and infrastructures has rapidly become one of the key issues in the evaluation and design of wireless access systems. The provisioning of high data rates, at an "affordable" price, constitutes a serious challenge to the structure and management of current and future wireless networks.

The management of radio resources, usually referred to as *Radio Resource Management* or RRM for short, has traditionally been benchmarked mostly by technical merits such as throughput (data delivery capability) and Quality of Service (QoS). In comparisons of different RRM schemes, it is often assumed that the scheme that can deliver more bits per Hertz (unit of bandwidth) or per Euro is the more efficient. From an economic point of view, however, cost efficiency is not equivalent to profitability [56]. While it is true that reduced cost leads to increased profitability, given that everything else remains the same (ceteris paribus), profitability also depends on revenue and thus the price that can be charged. Hence, from an economic point of view, an RRM scheme that is less technically efficient or more costly to implement may still prove to be more profitable. Since the price that can be charged (or users' willingness to pay) depends on the users' perceived value of the offered service, it seems plausible that the economic efficiency of RRM may be improved by accounting for users' perception of affordability and economic value of the service.

Increasing competition in the wireless access provisioning market and a growing plethora of diversified services, reduce the profit margins of the operators and improve users bargaining positions. There are, however, several potential means for an operator to improve its profitability. For convenience, we divide the schemes into two classes:

1. *'hard'* schemes which improve technical efficiency and
2. *'soft'* schemes which account for users service appreciation and willingness to pay.

The former aims at improving the economic efficiency by reducing cost, whereas the latter seeks to improve revenue by more accurately capturing users' preferences. In a perfectly competitive market [60] operators' efforts to improve their profitability indirectly also benefit the users. In practice, the user benefit is usually smaller. In this thesis we shall, however, primarily treat the operator benefit of improved RRM. We will try to improve it, though, by being more responsive to users' demands.

1.1 Scope of the Thesis

This thesis is a composite thesis, comprising eight papers which study three approaches to improve the efficiency of RRM in cellular systems.

Providing support for asymmetric traffic

First we observe that the increasing data traffic, as generated by information and multimedia services, may change the characteristics of the traffic loading our cellular systems. Most data applications either receive a lot more data than they send, or send a lot more data than they receive. Thus, making flexible resource allocation that can support asymmetric traffic valuable.

We investigate and propose RRM schemes to improve the *hard* efficiency of supporting asymmetric traffic in wireless networks in the four papers:

T1 "Dynamic Link Asymmetry in 'Bunched' Wireless Networks", M. Lindström and J. Zander, VTC 1999 Fall. [51]

> The original ideas were developed by the first author under the supervision of the second author. The paper was written jointly.

T2 "Improved TDD Resource Allocation through Inter-Mobile Interference Avoidance", M. Lindström, VTC 2001 Spring. [47]

T3 "Heterogeneous Link Asymmetry in TDD Mode Cellular Systems", M. Lindström, World Wireless Congress '03. [49]

T4 "Base Station Placement in Asymmetric TDD Mode Systems in a Manhattan Environment", M. Lindström, VTC 2004 Spring. [50]

It should be noted that some long-term symmetric services, such as speech, exhibit significant short-term asymmetric behaviour that can be exploited and more efficiently catered for with asymmetric links.

Quality and pricing aware resource management

While throughput maximisation maximises the chargeable goods and QoS-guarantees increase the attractiveness of the goods, it has also been shown that by accounting for users' perceptions of reasonable pricing of services, provider revenue may be increased [43, 16]. There are, however, few means in the literature that facilitate evaluation of the *soft* efficiency properties of RRM. Thus, we develop a model framework for *Quality and Pricing*

Aware Resource Management, demonstrate the implications of users' service-and-price dependent behaviour on the economic efficiency of a cellular system and further investigate potential benefits of revenue-based RRM:

P5 "Demand and Pricing Effects on the Radio Resource Allocation of Multimedia Communication Systems", L. Badia, M. Lindström, J. Zander and M. Zorzi, GLOBECOM2003. [6]

> The original ideas were jointly developed by the first and second authors. While author one was the responsible editor of the paper, author two provided most of the numerical examples. Authors three and four contributed valuable feedback.

P6 "An economic model for the radio resource management in multimedia wireless systems", L. Badia, M. Lindström, J. Zander and M. Zorzi, Computer Communications, vol. 27 (2004), no. 11, pp. 1056-1064, Elsevier B.V.. [8]

> The distribution of intellectual and labour contributions was the same as for Paper P5.

P7 "Speculative Admission Control and Scheduling for Packet-Switched Wireless Networks", M. Lindström, L. Badia, J. Zander, M. Zorzi, submitted to the IEEE Journal on Special Areas in Communications, Special issue: Price-Based Access Control and Economics for Communication Networks, February 2005. [45]

> The concept of speculative admission was seeded by discussions between the first and second authors. They later developed it independently: by author one as the AC component in a Speculative Resource Allocation framework, which also included scheduling, and by the second author isolated as Revenue-based Admission Control. The first author was the main author of the paper. Authors three and four contributed with fruitful discussions.

Creation of forwarding incentive in multihop cellular networks

Several researchers have shown that one can increase the range and/or capacity of a wireless communication system, without adding more expensive network equipment by means of a technique called *Multihopping* [30, 66]. Multihopping breaks long distance connections into multiple, shorter and more power efficient connections by letting users forward data of other users. Multihopping can potentially lower the cost of providing wireless access and thus increase the *hard* efficiency. Unfortunately, forwarding someone else's data is of little immediate benefit to the forwarding node. Unless overcome, this lack of forwarding incentive severely limits the feasibility of multihopping in cellular networks. Therefore, we propose and investigate a concept of *Resource Delegation* which is combined with a rewarding scheme to stimulate forwarding:

I8 "Resource Delegation and Rewards to Stimulate Forwarding in Multihop Cellular Networks", M. Lindström, P. Lungaro, VTC 2005 Spring. [46]

> The authors claim to have equally contributed to the original idea of resource delegation as a means to eliminate the bandwidth cost of the relay. It should be acknowledged, however, that 'subslotting' (see Section 6.3) was used, but with a different objective, by the second author in a previous work. The utility or objective functions of the different system entities were developed in intensive discussions among the authors. The first author acted as the main author of the paper and was also responsible for supplying the numerical examples.

Related papers

The following publications, not included in this Thesis, contain material similar to the above:

- "Improved TDD Resource Allocation through Link Gain Estimation", M. Lindström, 3Gwireless'01. [48]
- "An Economic Model for the Radio Resource Management in Multimedia Wireless Systems", L. Badia, M. Lindström, J. Zander and M. Zorzi, ASWN 2003. [7]
- "Resource Allocation for Asymmetric Traffic in Time Division Duplexing Mode Cellular Networks", M. Lindström, Licentiate Thesis, TRITA-S3-RST-0309. [44]

1.2 Monte Carlo Simulations

Most of the numerical results in the Thesis have been acquired from Monte Carlo simulations. Monte Carlo simulation is a powerful tool to find approximative solutions to quantitative problems by means of statistical sampling. Complex problems which are unwieldy to solve analytically can often easily be addressed with Monte Carlo methods. Slightly simplified, the procedure is to sample the studied system in a number of random configurations and then use statistical methods to describe the system as a whole. The solutions are, however, only estimates which can be significantly different from the 'true' solution. Fortunately, there are methods both to assess the accuracy of an estimates and to improve it. The accuracy can be improved by increasing the number of samples, but also by means of various so-called *variance reduction techniques*.

The accuracy of an estimate can be expressed with a *confidence interval* specifying the interval around the estimate in which the true solution, with a given probability (e.g. 95%), is located. Determining the confidence interval of an estimate of a probability is particularly easy[1]. Assume that $\hat{p}$ is an estimate of the probability p and that the estimate is the average of k samples. Then the 95% confidence interval is:

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{k}} \tag{1.1}$$

[1] We assume that the samples are Bernoulli random variables and that the number of samples is large.

Also, the relative length of the confidence interval is easily related to the number of occurences of the event for which the probability is being estimated. For a 95% confidence interval with a relative length of $2l$, the number of occurences needed is:

$$k\hat{p} = \frac{1.96}{l^2}(1 - \hat{p}) \tag{1.2}$$

When interpreting the results obtained from simulation studies, it is important to keep the limited accuracy in mind to avoid drawing false conclusions.

1.3 Thesis Outline

In Part I, Chapter 2 gives a brief introduction to radio access in cellular networks and present the concept of Radio Resource Management (RRM). Chapter 3 treats the support of asymmetric links in cellular networks. It presents the contribution of the Thesis in this area and relates it to previous work. A sampling of a few concepts from micro-economics and game theory is given in Chapter 4 as a preparation for Chapter 5 which presents the work on quality and pricing aware radio resource management. The last contribution area of the Thesis, forwarding incentives in multihop cellular networks, is covered in Chapter 6. Chapters 3, 5, 6 are all structured in a similar way. First the problem area is introduced. Then comes a review of previous work followed by a detailed description of the contribution of the author. Each original paper is briefly presented in a separate section. Conclusions are summarised at the end of the chapters. The last chapter, Chapter 7, summarises the work and discusses its applicability and extensions to further work.

Part II provides verbatim copies of the original papers presented in this Thesis.

Chapter 2

Radio Access in Cellular Networks

Because of the topology of modern mobile radio networks, they are commonly known as *cellular networks*. Analogously mobile phones are often referred to as cellular phones. A large geographical area is partitioned into smaller areas called *cells*. The mobile units, or *Mobile Stations* (MSs), in a cell are connected to a stationary unit, called *Base Station* (BS). The base station provides connections to other base stations and to the public wired network.

2.1 Multiple Access

To accommodate the communication of multiple users, *Multiple Access*, the radio resource is channelised. Users are given different *channels* in order to tell their signals apart. Channelisation can be accomplished in any of, at least, four domains and the corresponding multiple access protocols are:

- Frequency Division Multiple Access (FDMA) – The available spectrum is divided into smaller frequency bands and each user gets a separate frequency band. Though not entirely correct, the bands are commonly called *frequencies*, referring to the centre frequencies of the bands.

- Time Division Multiple Access (TDMA) – Users get to access the radio resource at different times. The entire resource may be used, but only for a predefined and limited period of time, a so-called *time-slot*.

- Code Division Multiple Access (CDMA) – Signals are encoded with special waveforms that expand the bandwidth of the signal, but also enable the separation of multiple concurrent signals in the same frequency.

- Space Division Multiple Access (SDMA) – The direction of transmissions and receptions are limited so that the spectrum can be simultaneously used by others in other directions.

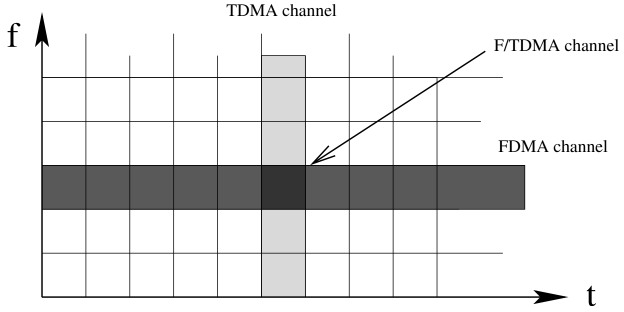


Figure 2.1: Hybrid FDMA/TDMA (F/TDMA) access protocol.

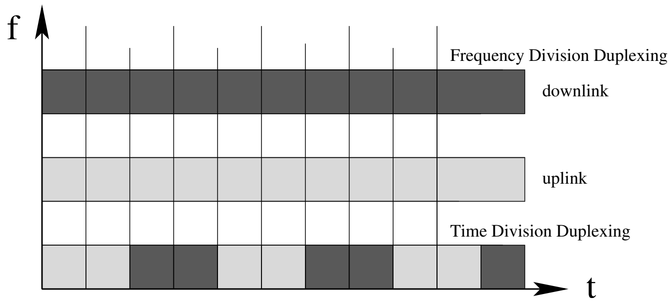


Figure 2.2: Duplexing methods

Different multiple access protocols may also be combined into *hybrid* protocols, Figure 2.1.

Bidirectional communication is often a requirement and, thus, the system must support simultaneous reception and transmission. This is known as *duplexing*. To make the direction of communication clear, we shall in the following use the word *uplink* for transmissions from MSs to BSs and the word *downlink* for transmissions from BSs to MSs. In theory, duplexing can be done by means of any of the described multiple access protocols. Due to physical constraints, however, only Frequency Division Duplexing (FDD) and Time Division Duplexing (TDD), Figure 2.2, remain viable options. The fact that we want transmitter and receiver to be at the same place makes space division a no-go for duplexing. For code division, the main culprit is the large power difference between the transmitted and received signals. The transmitted signal is typically some billion times stronger than the received signal and with today's amplifier technology the weaker received signal just cannot be recovered when mixed with the stronger transmitted signal. To be fair, it should be noted that there are limitations also to the applicability of frequency division and time division. FDD is possible only if the frequencies that we use are sufficiently separated by a *guard band*. This requirement is due to the power difference problem in combination with limitations in filtering technology; we lack the technology to make filters that can

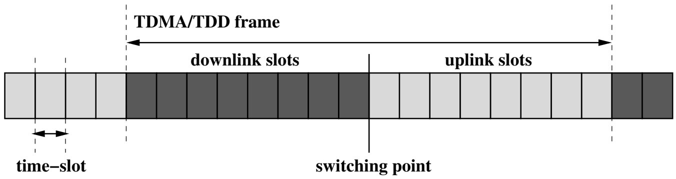


Figure 2.3: TDMA/TDD frame with downlink and uplink time-slots.

perfectly eliminate an unwanted signal that is too close in frequency to the wanted signal. With TDD, there is a delay associated with switching between the transmit and the receive modes. Thus, a small *guard interval/time* is needed not to be hit by the power difference problem.

The radio access method of a bidirectional wireless communication system is defined by a combination of a multiple access protocol and a duplexing method. Possible combinations include, but are not limited to:

- FDMA/FDD – used in the Nordic Mobile Telephone (NMT) and Advanced Mobile Phone (AMPS) systems,
- Hybrid FDMA/TDMA/FDD – used in the Global System for Mobile Communication (GSM) and Digital AMPS (D-AMPS) systems,
- CDMA/FDD – used in the American Interim Standard 95 (IS-95) and
- Hybrid FDMA/TDMA/TDD – used in the Digital Enhanced Cordless Telephone (DECT) system

In this thesis we shall, most of the time, consider a pure TDMA/TDD access method. TDMA time-slots are typically structured into groups called *frames*. Some of the time-slots in a TDMA/TDD frame are used for uplink communication and others for downlink, Figure 2.3. The point within the frame where the communication direction changes is called the *switching point*.

If the TDMA frame has N_{TS} time-slots, but more than N_{TS} channels are needed, TDMA can be combined with Frequency Division Multiple Access (FDMA), by adding more carriers, and/or Code Division Multiple Access (CDMA) by allowing multiple, code separated, transmissions in each time-slot. While these hybrid schemes are of practical importance, they are not considered in this thesis.

We call the smallest allocatable piece of the radio resource a Resource Unit (RU). In the hybrid access scheme this would correspond to a time-frequency-code slot. For plain TDMA an RU corresponds to a time-slot and the number of RUs equals the number of time-slots; i.e. $N_{\mathrm{RU}} = N_{\mathrm{TS}}$. When we are concerned with pure TDMA/TDD we will use the terms RU and time-slot interchangeably.

2.2 Signal-to-Interference-Ratios

The cellular concept was introduced in the 1970s to conserve bandwidth and increase capacity by means of frequency re-use [54]. When frequencies are reused, communication will be interfered by other concurrent communication in the same frequency. Successful communication is still possible, however, if the wanted signal, S, is sufficiently stronger than the interfering signals, I. We shall frequently refer to the *Signal-to-Interference-Ratio* (SIR), i.e. $\frac{S}{I}$, as a measure of the channel quality. Since radio signals are attenuated as they propagate, the SIR depends on the distance to the co-frequency cells; i.e., cells that use the same frequency. The concept of frequency reuse can be generalised to channel reuse by just replacing frequency with the dimension of choice. Therefore, we often speak about co-channels instead of co-frequencies.

2.3 Radio Resource Management

Since the transmission power directly affects the quality of the received signal, transmission power is considered a radio resource just as frequencies, time-slots, codes etc. are. Radio Resource Management (RRM) aims at making efficient use of the available radio resources. The RRM system is responsible for finding channels with sufficiently high quality, for assigning them to entities who need to communicate and for maintaining the channels for the duration of the communication. The definition of “efficient” may vary.

One often speak of centralised versus decentralised or distributed resource management. The resource management is usually considered to be centralised if it is based on information collected, at a central point, from all parts of the system. If the decisions are based only on information locally available, in for instance one cell, it is considered distributed. The more information that is available, the more fine grained control one has of the system. Thus, systems are generally believed to be able to achieve higher spectral efficiencies with centralised resource management than with decentralised. However, computational complexity and physical restrictions, such as delays caused by large distances, limits the size of centralised systems. For large cellular networks the complexity can be kept manageable by dividing the network into subnetworks, each of which is locally centralised thus retaining, to the extent possible, the performance benefits of a centralised system. It was shown by Berg and Pettersson, that locally centralised systems can achieve good performance even with very simple resource allocation algorithms [12]. Berg and Pettersson developed what they term “Bunched Radio Resource Management” where the subnetworks or bunches consist of groups of remote antenna units (RAUs) connected to and controlled by a hub or central unit (CU), figure 2.4.

Each bunch is equipped with an intra-bunch resource manager, which facilitates resource allocation and deallocation, power control and channel measurements.

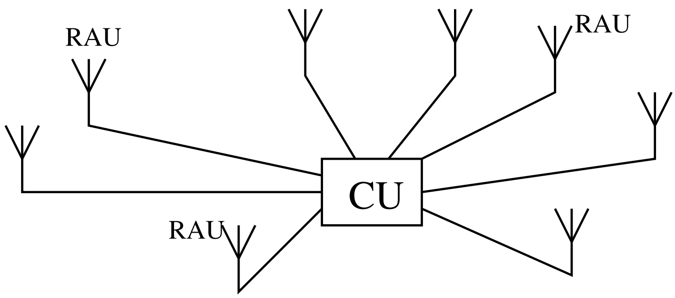


Figure 2.4: A small subnetwork or bunch. A number of Remote Antenna Units (RAUs) are connected to a Central Unit (CU) which is responsible for intra-bunch resource management.

Chapter 3

Supporting Asymmetric Traffic

In many current and future services there are and will be asymmetric data flows. Thus, an increasing fraction of the total offered traffic is expected to appear in the downlink segment of the system. When this kind of traffic is handled by a traditional symmetric full duplex system, there may be a lack of resources for communication in the high-volume direction while there is a surplus in the low-volume direction.

Asymmetric traffic could be handled more efficiently if the system would provide support for asymmetrical links, i.e. different effective data rates in uplink and downlink respectively. Ideally, no resources would be dedicated to a particular link direction. The resource manager would instead track the demand and allocate resources to uplink and downlink on an as-needed basis. Here we see a big difference between TDD and FDD mode systems. While in TDD mode systems, in principle, any number of the time-slots in the TDMA frame can be used for uplink or downlink, FDD is more limited. Due to the requirement of a guard band between the uplink and downlink bands, resources have to be a priori assigned to either uplink or downlink. With TDD, resources can be reallocated from uplink to downlink or vice versa, should that be necessary. This freedom is, however, as we shall see, somewhat limited. Adverse interference conditions easily arise. Especially if different asymmetries are required in neighbouring cells.

In traditional symmetric full duplex systems there is co-channel interference due to channel reuse. The uplink in a cell is interfered by uplinks in the co-channel cells. Let us call this interference *Same-Link Interference* (SLI). In TDD mode systems we can, however, have additional interference. Imagine that two cells are using the same time-slot for different link directions. Then, the uplink of one cell will be interfered by the downlink of the other cell and vice versa. That is, we are no longer interfered only by links of the same type, but also by links of the opposite direction, *Opposite-Link Interference* (OLI). The OLI can be further classified as *Inter-Mobile Interference* (IMI), when the downlink is interfered by the transmission of one or more MSs, or *Inter-Base-station Interference* (IBI), when the uplink is interfered by one or more BSs, Figure 3.1. The IMI case is particularly troublesome because MSs of neighbouring cells can be located very close to each other and, hence, end up in a situation where the interference is stronger than the wanted signal.

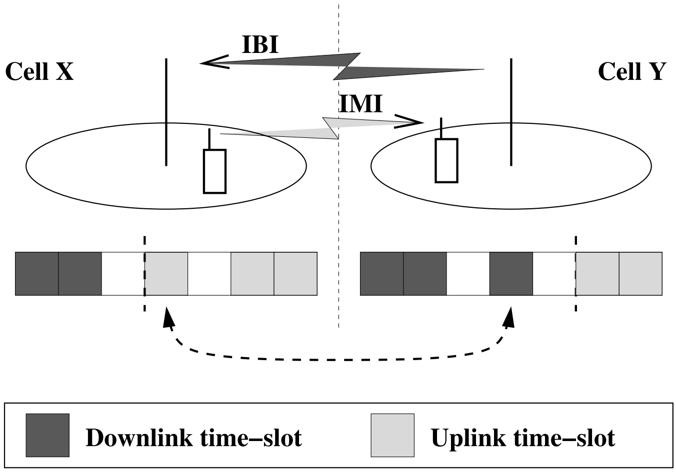


Figure 3.1: Conflicting switching point assignments in adjacent cells causes IMI and IBI.

To maintain a high channel quality, OLI need to be avoided. Therefore, a high degree of flexibility usually requires extensive knowledge about propagation losses between mobiles and base stations, respectively. Since this information is difficult to obtain, many resource allocation algorithms reduce IMI and IBI by limiting the flexibility with which the switching point can be selected. If a common switching point is used in all cells, outage caused by IMI and IBI becomes negligible, Figure 3.2. Unfortunately, in cells where the loads of the uplink and the downlink do not match the switching point, connections then have to be blocked even though there are still free time-slots. The free time-slots are reserved for the opposite link direction.

3.1 Related Work

The TDD interference dilemma was noticed already in the early development of the Digital European Cordless Telephony (DECT) system, later to become the Digital *Enhanced* Cordless Telephony (also DECT) system. DECT supports asymmetric traffic and 'solves' asymmetry and slot drift interference problems through slot handover [21] (section 3.10.2). Multiple-slot same-link other-cell interference as a result of lack of synchronisation was observed by Åkerberg [81], but other-link other-cell same-slot interference due to different asymmetries does not appear to have been paid much attention to at that time. The problem with base-to-base interference was acknowledged, however, in [21] (section D.4.5) with references to RES3-(90)178 and RES3-(90)185. Mobile-to-mobile interference was still not mentioned though.

A concept of adaptive duplexing for TDD mode systems, that can almost double the capacity by means of speech activity detection, was introduced by Nanda and Yue [57]. Unlike Packet Reservation Multiple Access (PRMA), which was also proposed to increase

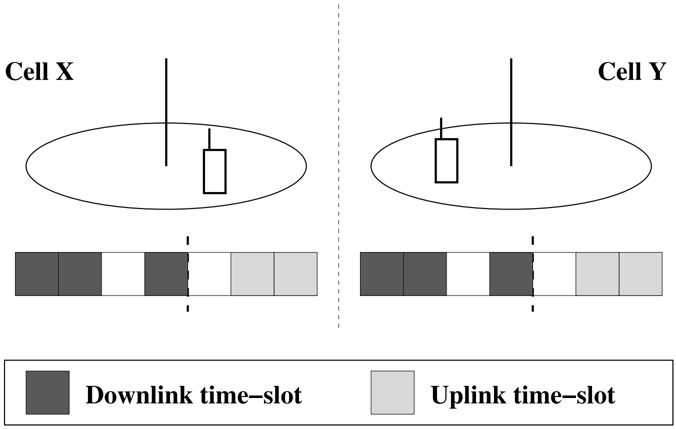


Figure 3.2: In a synchronised system with unified switching points there is no IMI or IBI.

capacity in TDMA cellular networks [28], adaptive duplexing does not depend on statistical multiplexing of many concurrent calls. Instead, the resources assigned to two conversing parties are dynamically shared among them. The authors noted that, because of the scheme's inherent frame asynchronism, downlink (uplink) transmissions are interfered by both uplink and downlink transmissions. Thus, requiring a larger reuse distance than would a frame synchronous TDD system. The authors expressed the belief that the main cause of the problem is line-of-sight propagation between base stations.

The performance limitations of frame asynchronous TDD was investigated in [15]. The degradation of both TDD and FDD mode systems due to frame asynchronism were studied. Both systems suffer from not being synchronised, though TDD significantly more than FDD.

A TDD-CDMA mode was proposed, and later accepted, for inclusion in the specification of the Universal Mobile Telecommunications System (UMTS) alongside with an FDD Wideband CDMA mode. The two modes are commonly known as UTRA/TDD and UTRA/FDD, respectively, where UTRA is an abbreviation of UMTS Terrestrial Radio Access. The specification for UMTS was proposed by the European Future Radio Wideband Multiple Access Systems (FRAMES) project. During the development of the proposal, the benefits of TDD mode systems to FDD mode systems were considered [61, 62, 63]. It was noted that the ease with which uplink and downlink resources can be reconfigured makes TDD particularly attractive for multimedia applications which frequently require an asymmetric channel. Many advantages of using TDD are attributed to the reciprocal nature of the channel, i.e since the uplink and downlink of the TDD channel utilise the same channel bandwidth, the impulse response for both links is essentially the same. Useful features arising from this fact are, among others:

- use of pre-equalisation,
- use of open-loop power control,

- more efficient transmit antenna diversity,
- no need for a duplex filter and
- sharing of hardware between receiver and transmitter.

In the FRAMES project, inter-cell and intra-cell interference problems originating from non-unified boundaries between uplink and downlink in adjacent cells were again identified. It was concluded that, to combat this, such cells must have the same asymmetry and be synchronised. Thus UTRA/TDD was suggested for isolated, short range, high bit rate cells within a full coverage UTRA/FDD. The interference problem was further investigated and it was found that both co-channel and adjacent-channel (mainly with respect to frequency) forms of OLI would be detrimental to performance [65, 33, 31]. Multi-operator scenarios and the importance of a separation distance between the BSs of different operators were also treated [33, 31].

There are several means to achieve frame/slot synchronisation. For example, all cells could be furnished with a GPS receiver or the timing could be provided by interconnecting the cells with a real time wire connection. Alternatively the system can be synchronised over the air. One over-the-air synchronisation scheme was proposed in [35] for use with UTRA/TDD. It was shown to be efficient and capable of achieving a synchronisation accuracy of 100 ns within a start-up time of approximately 400 seconds.

The significance of the BS placement to achieve good coverage and high capacity has been studied extensively in the literature. Though, almost exclusively for IMI and IBI free systems. Deployment strategies based on both optimisation methods [69, 5, 36, 29] and regular deployment patterns [55, 14] have been proposed and investigated. It has been established that the shadowing caused by buildings in urban environments can be used to reduce interference, but that in doing so the tracking of users when they turn around corners, becomes harder [14]. The implications of IMI and IBI were not considered.

The asymmetric properties of many data applications, such as web browsing, are rather obvious. Also voice communication systems over short durations of time exhibit an asymmetric behaviour since pairs of users engaged in conversation are not talking simultaneously. In [58, 77], a scheme where uplinks and downlinks are sharing time-slots is shown to give considerable capacity gains compared to traditional “static” TDD.

Since services with asymmetric data flows are increasing, it is desirable to have a dynamic direction allocation scheme, capable of tracking the traffic demand. For data, such a dynamic channel assignment algorithm, with an adjustable boundary or switching point between uplink and downlink, was proposed and evaluated in [13]. There too, significant capacity gains were found, but similar to [63] it was suggested that neighbouring cells should have the same asymmetry.

3.2 Contribution

The possibility to support different asymmetries, or just different slot configurations, in neighbouring cells is very attractive. Though several studies indicate that non-unified switching points are either unfeasible or don’t provide significant performance gain, there are also findings pointing to the contrary. Most studies refer to systems with distributed resource management and with no particular measures taken to avoid IMI or IBI. Also, locally centralised resource management has been shown to give substantial capacity gains in symmetric systems. Since in locally centralised systems, more information about the propagation conditions in the system is more readily available, it is reasonable to assume that this would be beneficial also to asymmetric TDD mode systems. It is believed that with appropriate interference avoidance measures, the ill effects of IMI and IBI can be limited.

Papers T1 through T4 (reproduced verbatim in Part II, Chapters 8, 9, 10 and 13, respectively) aim at establishing the performance gains possible in a TDD mode wireless network with locally centralised resource management and whether it would be worthwhile to implement special support for asymmetric traffic. The main contributions of the papers are as follows:

- T1 the proposal and evaluation of an extension from an adjustable boundary and decentralised resource management to multiple switching points and locally centralised resource management in order to increase the systems flexibility and ability to accommodate asymmetric traffic in a spectrally efficient way.
- T2 the proposal and investigation of a method to estimate and avoid inter-mobile interference based on measured MS-to-BS and BS-to-BS link gains.
- T3 an assessment of the implications of serving different asymmetries in different cells in a multiple cell TDD mode system also investigating the performance penalties incurred by spatially varying asymmetry.
- T4 an investigation of the significance of the BS placement for support of asymmetric traffic.

3.3 Dynamic Link Asymmetry (Paper T1)

In paper T1, "Dynamic Link Asymmetry in 'Bunched' Wireless Networks" we propose and evaluate a Radio Resource Allocation (RRA) scheme for flexible support of asymmetric traffic in TDD mode cellular systems with locally centralised resource management. We here summarily present and justify the proposal, some reference schemes and the results. A thorough description of the system models that were used is supplied in Appendix A.

Static Common Switching Point

As TDD has, traditionally, been used as an alternative to FDD, mostly to simplify and reduce the complexity of hardware [23], resource allocation has also followed the lines of that of FDD. That is, each time slot has been statically preallocated to either uplink or downlink; usually with the same pattern in all cells of the system. This kind of Static Common Switching Point (SCSP) effectively eliminates IMI and IBI, but limits the flexibility otherwise inherent in TDD. Due to temporal and spatial variations in the traffic, there will frequently be a lack of RUs for one link direction while, simultaneously, there is a surplus for the other link direction. Thus, RUs are wasted. Temporal variations can be taken into account by adjusting the switching point to reflect the time average of the link asymmetries in the system. Spatial variations, however, cannot as easily be accounted for while completely avoiding IMI and IBI.

The rules that, inserted in the resource allocator in Section A.1, would implement SCSP are simple. A single switching point is selected such that the ratio of the number of reserved downlink RUs to the number of reserved uplink RUs matches the average system link asymmetry as close as possible. Most of the time, an exact match is not possible. The number of reserved downlink and uplink RUs become:

$$N_{\text{DL}} = \lfloor N_{\text{RU}} A + 0.5 \rfloor \tag{3.1}$$

$$N_{\text{UL}} = N_{\text{RU}} - N_{\text{DL}} \,, \tag{3.2}$$

respectively.

When a downlink (uplink) RU is requested, the TDMA frame is searched sequentially for a free RU, from the beginning (end) of the frame towards, but not beyond, the switching point. The search is pre-empted at the switching point to prevent inter-cell interference. More formally, the nth candidate is, for downlink and uplink, respectively:

$$C_n^{\text{DL}} = \begin{cases} \{\text{RU}_n\} & \text{if } n \in [1, N_{\text{DL}}], \\ \{\} & \text{otherwise.} \end{cases} \tag{3.3}$$

and

$$C_n^{\text{UL}} = \begin{cases} \left\{\text{RU}_{(N_{\text{RU}}+1-n)}\right\} & \text{if } n \in [1, N_{\text{UL}}], \\ \{\} & \text{otherwise.} \end{cases} , \tag{3.4}$$

where RU_1 is the first RU in the frame and $\text{RU}_{N_{\text{RU}}}$ is the last.

Note that neither having only one switching point nor testing RUs sequentially are strictly requirements, but greatly simplify the implementation without loss of generality.

Movable Boundary

In order to support asymmetry ratios that vary or does not correspond to an integer number of time slots, schemes that assign time slots on an as-needed basis are compelling. Such schemes would not statically divide the TDMA frame into uplink and downlink segments, but delay the division decision for as long as possible. Ideally a decision to divide the frame would not have to be made at all.

The Movable Boundary (MB) scheme proposed by Chen et al [13] partially achieves the desired flexibility. It divides the frame, but not a priori. Similar to the SCSP rules above, the MB scheme sequentially searches the TDMA frame from one end to the other for a free RU. The search for downlink and uplink RUs, respectively, start from opposite ends of the frame. That is, a search for downlink RUs proceeds from the beginning of the

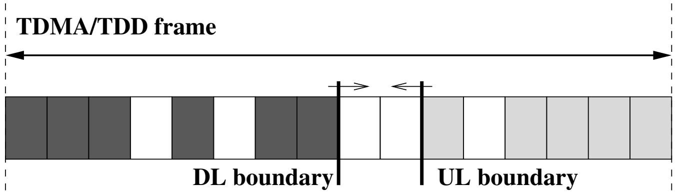


Figure 3.3: The principle of the Movable Boundary slot search algorithm. The search for free downlink and uplink slots, respectively, start at opposite ends of the TDMA frame and stop where uplink and downlink regions meet.

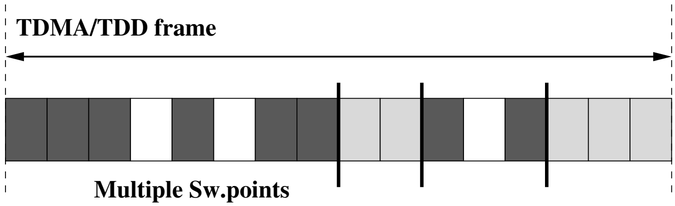


Figure 3.4: The SSP rule set allows the RU search to proceed beyond the switching point.

frame to the end while a search for uplink RUs starts at the end of the frame and works its way towards the beginning. Unlike SCSP, however, the search does not stop at a predefined switching point, but rather continues until a time-slot allocated to one or more RUs of the opposite link direction is encountered, Figure 3.3. This guarantees that there is only one switching point in a frame. It also implies that time slots near the beginning and the end of a frame will experience less IBI and IMI than more central slots. This was exploited in a patent [22] independent of the authors of the paper [13].

With a movable unsynchronised switching point the systems ability to accommodate nonuniform asymmetry is improved. The improved flexibility comes, however, at the expense of the introduction of IMI and IBI in the system. The single switching point, in contrast to multiple switching points, limits the interference, but also limits the flexibility. Freedom to use any of the available time-slots for any link direction would yield greater flexibility.

Soft Switching Point

Owing to our assumption of locally centralised resource management, more information is available to our resource manager about the propagation conditions than to the decentralised resource manager used in [13]. Thus we propose a different, enhanced, RU search rule where the stop condition has been relaxed and the search for a free RU can continue beyond the switching point Figure 3.4 [51]. Hence there may be multiple switching points within one TDMA frame. More switching points generally means more opportunities to be hit by IMI/IBI, but we shall rely on a feasibility check to avoid and to keep the interference on an acceptable level. We refer to this rule set as the Soft Switching Point (SSP) rule set.

The system performance resulting from the SSP rules is compared to the performance with the less flexible SCSP and MB rules defined above. This comparison, however, only shows the relative benefits of one scheme to the other. The schemes all belong to a class of algorithms where free RUs are discovered by means of sequential search from the beginning of the frame to the end, or from the end to the beginning, depending on the link direction. Thus, it is desirable to find an upper bound on the performance achievable within this class. Such a bound provides additional understanding of the performance and limitations of this class of resource allocation algorithms.

The resource allocator under consideration has access only to information about MS-to-BS link gains and BS-to-BS link gains for its feasibility check. Since information about MS-to-MS links is missing, the output from this feasibility check is not entirely accurate. IMI is not properly accounted for and, thus, can cause the resource allocator to make mistakes that result in outage.

Perfect Soft Switching Point

Relaxing possible complexity requirements/constraints, we define a reference resource allocator which has access to the missing MS-to-MS link gains. The feasibility check of this reference allocator is almost error free. Almost, because imperfections like measurement errors and stale measurement data can still adversely affect the accuracy of the feasibility check. In this work, however, we don't study measurement data errors and so consider the feasibility check with knowledge about all link gains to be error free. Thus, the reference allocator is capable of perfectly avoiding outage. We combine the perfect allocator with the SSP RU search rule, the most flexible of the ones we have defined, to form what we denote: Perfect SSP (PSSP), where perfect should be understood as "with knowledge about all link gains" and "able to completely avoid outage". PSSP would of course be difficult to realise, but may serve as a kind of upper bound on the performance of the class of algorithms studied in this work. That is, the class of resource allocation algorithms that consider random terminal order, ordered sequential RU search and first feasible RU allocation.

Algorithm Summary

In summary, we evaluate the possible performance gains obtained by providing Dynamic Link Asymmetry in a locally centralised cellular system. Four algorithms are studied. They belong to a class of algorithms that sequentially search the frame for free RU and choose the first RU that passes the integrated feasibility test. Depending on the link direction, the search starts either at the beginning or at the end of the frame. In increasing complexity order, the algorithms are:

- SCSP (Static Common Switching Point), where the resources are statically split between up- and downlinks according to the medium to long term average asymmetry of the load offered to the system.

- MB (Movable Boundary) [13] which has a single adjustable switching point.
- SSP (Soft Switching Point) [51] algorithm, which is similar to the Movable Boundary algorithm, but allows multiple switching points.
- PSSP (Perfect Soft Switching Point), where SSP has been combined with knowledge about all link gains and interference powers, including MS-to-MS links. This hypothetical system has been included to provide an upper bound on the performance and illustrate the magnitude of performance loss due to missing information about IMI.

Results

The proposed SSP scheme was evaluated in a cellular system, described in Appendix A, by means of snapshot simulations. The performance measures were:

a) the user blocking rate ν_{blk} , i.e. the rate at which terminals are unable to find a free time-slot,

b) the user outage rate ν_{out}, i.e. the rate at which users that are assigned channels fail to reach the required SIR,

c) the Grade of Annoyance (GoA), a weighted user assignment failure rate,

$$\text{GoA} = \nu_{\text{blk}} + 10\,(1 - \nu_{\text{blk}})\,\nu_{\text{out}} \tag{3.5}$$

and

d) the system capacity, C.

The capacity, C, is the load at which the GoA becomes 5% for either uplink or downlink, whichever occurs first. It should be noted that the definition of GoA which was used in the original Paper T1 used equal weights for blocking and outage. For this presentation we have, however, unified the performance measures to achieve consistency and comparability with Papers T2–T4.

While a measure of asymmetry can be defined in many ways, we chose a linear definition where asymmetry, A, is simply the fraction of the traffic that appears in the downlink. That is, $A = 0$, $A = 0.5$ and $A = 1$ correspond to uplink-only, symmetric and downlink-only traffic, respectively.

Results show that as expected the (hypothetical) PSSP allocation scheme where any resource unit can be used for any link direction, i.e. RUs are reserved neither for uplink nor for downlinks, outperforms a scheme with fixed, to the average asymmetry adapted, switching point. The PSSP scheme assumes full knowledge about all cross base and cross mobile link gains. Since this requires a huge amount of measurements and signalling, this algorithm is not considered feasible, but serves merely as an upper bound on performance. Further, we see that when there is more downlink traffic than uplink traffic, the SSP scheme provides a capacity, while not as high as for the PSSP scheme, still considerably higher than the capacity of the SCSP scheme. For symmetric traffic, however, the SCSP algorithm seems to provide higher capacity than the SSP and MB algorithms, contrary to the results of [51]. The difference is due to a change of performance measures. Here we have given outage a 10 times higher penalty than blocking. In [51], blocking and outage had equal weights.

Base stations are, as opposed to mobile stations, relatively few and have fixed locations. Since the SSP algorithm uses no knowledge about cross mobile link gains its complexity is only slightly higher than that of the SCSP scheme. The SSP scheme is also adaptive, i.e. tracks varying asymmetry ratios without the need for external adaptation mechanisms. The performance differences between the MB and SSP schemes are minor.

The low blocking rates of the Movable Boundary and Soft Switching Point schemes counteracts "starvation" effects at very skewed load distributions, but comes at the cost of higher outage in the downlink. Some means to avoid inter-mobile interference would improve the general performance.

Confidence of the Simulation Results

Regarding the confidence of the simulation results, it should be noted that the accuracy depends on the load of the system. Points corresponding to low load or capacity are less reliable than those corresponding to high load or capacity. For asymmetries close to 0 or 1 the variance in the results is higher than elsewhere because of the constraint that neither the uplink nor the downlink must exceed the GoA requirement; there are few users in the low-volume direction. It should also be noted that the weight of 10 assigned to outage in the GoA measure will affect the confidence in cases of outage, especially if the outage occurs in the low-volume direction. Uplink outage has been avoided by the feasibility check of our algorithms. Thus, results corresponding to asymmetries close to 1 are more reliable than those corresponding to asymmetries close to 0. Each data point is an average of 60 realisations of the system. Hence, for a relative load of $\bar{\omega}_c = 0.5$, there were $60\bar{\omega}_c N_{\text{c}} N_{\text{TS}} = 60 \cdot 0.5 \cdot 55 \cdot 16 = 26400$ users simulated. For an asymmetry of $A = 1/16$, however, only 1650 appear in the downlink.

For the reasons listed above, we avoid to draw conclusions about results corresponding to asymmetries less than 0.3 and loads or capacities less than 0.2.

3.4 Link Gain Estimation (Paper T2)

The SSP algorithm proposed and studied in Paper T1 [51] showed excellent blocking performance, but due to IM link gains being infeasible to measure, capacity improvement was not remarkable.

The key problem caused by moving the boundary is the introduction of IMI and IBI. This interference must be avoided in order to maintain a low outage probability. Thus, a high degree of flexibility usually requires extensive knowledge about propagation losses between mobiles and base stations, respectively. Since this information is difficult to obtain, most resource allocation algorithms reduce IMI and IBI by limiting the flexibility with which the switching point can be selected.

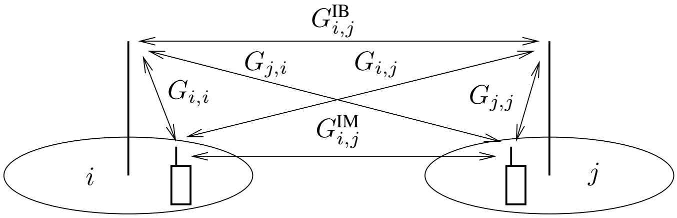


Figure 3.5: Link gains in and between cells i and j

In order to maintain the high flexibility of the SSP algorithm we want to solve it differently. In Paper T2, "Improved TDD Resource Allocation through Inter-Mobile Interference Avoidance", we propose and investigate techniques to estimate IM link gains and reduce IMI.

Location Based Estimates

Would the positions of the mobiles be known, the link gain between two mobiles could be estimated with the aid of maps and propagation models or they could simply be upper-bounded with a free space propagation loss. Finding the approximate locations of mobiles is rather straightforward [59]. Getting accurate positions is, however, very difficult [78, 38, 20, 32]. Fading introduces an uncertainty in the pin-pointing of a terminal and the positioning is further complicated by the non-line of sight problem [78]. Assuming that an IM link gain estimate is a function of two mobiles' positions, the uncertainty that propagates to the link gain estimate can be significant. The high complexity of the above method is also an issue. Therefore, location-based link gain estimates will not be further considered in this thesis.

Related-Links Based Estimates

Another option is to approximate the IM link gain by means of a function of some related link gains. The accuracy of this method would be low, though. Consider the scenario in Figure 3.5, where the link gains are: $G_{i,j}$ between BS j and MS i, $G^{\mathrm{IB}}_{i,j}$ between BSs i and j and $G^{\mathrm{IM}}_{i,j}$ between MSs i and j. The estimate

$$\widehat{G^{\mathrm{IM}}_{i,j}} = G^{\mathrm{IB}}_{i,j}, \tag{3.6}$$

is not useful because the propagation loss between the BSs is often at least 45 dB greater than the propagation loss between the MSs for interesting cases; cases where MSs are close and IMI would be severe. For these cases, the error would be too large. For non-adjacent cells, however, Equation 3.6 could be good enough.

Linear combinations of link gains fail as estimates because propagation losses are not additive. Even if they were, a MS could be located on any side of the BS. Assuming that MSs are located on a line between the BSs would yield a very pessimistic estimate.

Further, shadow fading would introduce a large error. Consider a case with log-normal shadow fading with a log-standard deviation of 10dB. Then the link gain estimate will contain a sum of several such contributions which will render the estimates rather useless.

'Disjoint BS-Set' Estimate

We propose a heuristic method where IM link gains are approximated with measured IB link gains. However, to further reduce IM interference, the scheme also restricts which mobiles may share time-slots. The proposal goes as follows:

For every mobile i, a set of only the n strongest base stations is considered. That is,

$$M_i^n = \left\{B_i^1, \dots B_i^n\right\}, \tag{3.7}$$

where B_i^j is the jth strongest base station as seen by mobile i. If the sets of two mobiles, k and l, are disjoint, the corresponding inter-mobile link gain, denoted $G_{k,l}^{\mathrm{IM}}$, is approximated with the largest cross-set inter-base station link gain,

$$G_{k,l}^{\mathrm{IM}} = \max_{i \in M_k^n,\ j \in M_l^m} \left(G_{i,j}^{\mathrm{IB}}\right), \tag{3.8}$$

where $G_{i,j}^{\mathrm{IB}}$ is the link gain between base stations i and j. If the sets of the two mobiles are not disjoint, the mobiles must only share a time-slot if both are transmitting or if both are receiving, Figure 3.6. The above scheme can be varied by changing the size of the base station sets. In the simplest variant, there is only one BS in each set and IM link gains are simply replaced with corresponding IB link gains. For mobiles in critical positions, i.e. mobiles that are near each other but connected to different base stations, little or nothing is gained with these estimates since the IM link gain will be greatly underestimated.

When two mobiles are near each other their link gains toward neighbouring base stations can be expected to be highly correlated and thus the probability that the two mobiles will have overlapping sets increases rapidly with the size of the sets. Since mobiles with overlapping sets may not share time-slots for different link directions, less IM interference is introduced. On the other hand, the application of sets of size three or greater is believed to become too restrictive. Even a set size of two could possibly be too restrictive. Thus, noticing that the set size for transmitting mobiles and the set size for receiving mobiles can actually be controlled separately, less restrictive, asymmetric, variants are also possible. However, in this work we shall limit our attention to BS sets of size two.

Results

Results show that with poor estimates, e.g. no estimates, system capacity is limited by outage. If the estimates are too conservative on the other hand, e.g. an assumption of free space propagation loss, capacity is limited by excessive blocking. The proposed 'Disjoint BS-Set' Estimate (DBSE) inter-mobile interference avoidance scheme compares the BSs that MSs have high gain to and provides the resource allocator with a guesstimate of which MSs are compatible from an IMI point of view. Application of the proposed scheme to the

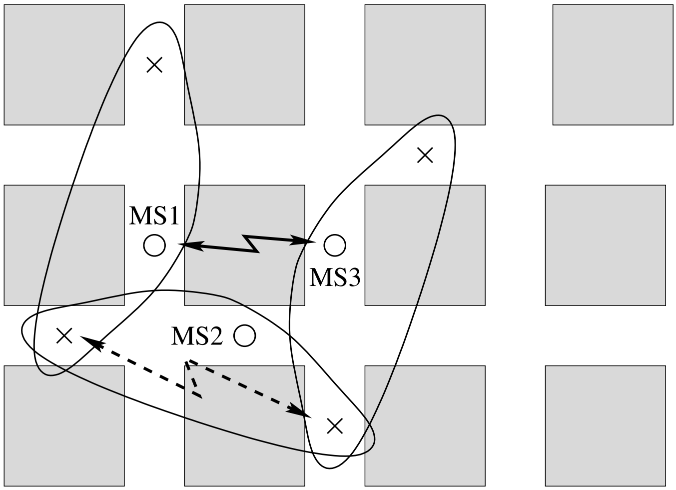


Figure 3.6: "Strongest base stations" sets for three mobile stations. The sets of MS1 and MS2 are not disjoint. Thus MS1 is not allowed to transmit in a time-slot where MS2 is receiving and vice versa. Since MS1 and MS3 have disjoint sets, we approximate the MS1-to-MS3 (solid arrow) link gain with the greatest inter-set base station gain (dashed arrow).

SSP algorithm in a centralised system results in reduced outage while blocking is increased. DBSE yields a fair system with respect to uplink and downlink GoA; i.e., neither link direction is favoured before the other. All in all, the capacity of the system is increased.

3.5 Heterogeneous Link Asymmetry (Paper T3)

With the increasingly diverse services offered in wireless communication systems, the asymmetry of the traffic offered to the system is likely to vary over the coverage area. For instance, voice traffic could dominate in some areas whereas event driven data transfers (i.e. web browsing) could dominate in other. The presented dynamic resource allocation schemes provide a means, not only to support asymmetric links, but also to support asymmetries which vary over the coverage area. Unfortunately, there is a penalty paid for this support; heterogeneous asymmetry gives rise to IMI and IBI which lowers the quality of the affected links, possibly resulting in outage.

Granted, supporting only one global asymmetry, A_{G}, throughout the system would efficiently avoid IBI and IMI. However, ignoring the variations would cause excessive blocking in parts of the system where the asymmetry of the offered traffic matches poorly with the asymmetry supported by the system.

In Paper T2 the Soft Switching Point strategy, presented in Paper T1, was improved with the introduction of a heuristic IMI avoidance scheme, the DBSE method. As was seen, performance was markedly improved. However, up till this point, no systematic variations in local expected asymmetry were introduced. In the previous papers, users were uniformly distributed over the area with identical expected asymmetries. Hence, asymmetry variations between different areas and cells were limited to instantaneous non-uniformities in the local user densities; i.e. different actual number of users in different areas at a certain instant.

In Paper T3 we evaluate and assess the performance gains of the SSP strategy with DBSE in environments with location dependent traffic asymmetry. The results are compared with the corresponding results of a system where the switching point is fixed in the same position throughout the system.

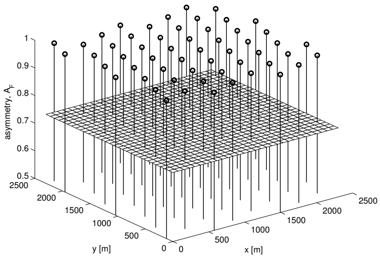
Asymmetry Scenarios

The asymmetry of the traffic offered to the system depends on the mix of services that are used by the users. This mix, in turn, is likely to vary depending on where users are located. In residential areas the mix may, for instance, consist of voice, web browsing and streaming media. It is our belief that most of the data traffic in residential areas would appear indoors. Outdoors there would mainly be voice communication and light browsing. The further away from home people get, the more inclined they will be to use wireless data services; wired infrastructure becomes less accessible. In shopping areas, voice services are expected to dominate. Asymmetric traffic will be present, however, at spots like cafes and restaurants etc.

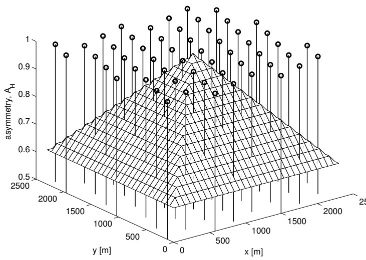
Private persons are usually cost sensitive and often sacrifice time and convenience to save money. To business users, time is money and convenience is therefore important. Hence, business users may, to a larger extent than private persons, use wireless data services also where wired solutions exist. Consequently, we expect more asymmetric services in business areas than in residential.

Accurate modelling of scenarios like the ones above is difficult and inconvenient. Partly because it is difficult to quantify the differences between the different environments and partly because it would be difficult to interpret the output of the study. Thus, we break the scenarios down into three simple models. The scenarios above can be thought of as a mix of the following three asymmetry scenarios:

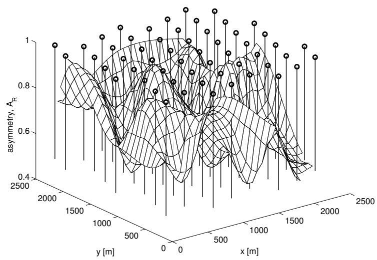
- A *flat* asymmetry, where the offered asymmetry, $A_{\mathrm{F},i}$, is equal in all cells i, Figure 3.7(a);
- a *hot spot* asymmetry, where there is a strong correlation between neighbouring cells, but asymmetry, $A_{\mathrm{H},i}$, changes gradually over the cells to reach its maximum, $A_{\max}$, in the centre of the service area, Figure 3.7(b) and
- a *random* asymmetry, where the expected asymmetries of different cells, $A_{\mathrm{R},i}$, are mutually independent, Figure 3.7(c).



(a) Flat asymmetry



(b) Hot Spot asymmetry



(c) Random asymmetry

Figure 3.7: Illustration of different asymmetry scenarios. The surface represent the expected local asymmetry as a function of the location. The vertical lines represent the locations of the base stations.

Assuming that the cellular system is deployed in the square area:

$$\begin{cases} -1 \leq x < 1 \\ -1 \leq y < 1, \end{cases} \tag{3.9}$$

the three scenarios can be defined more formally as follows:

$$A_{\mathrm{F},i} = A_G \tag{3.10}$$

$$A_{\mathrm{H},i} = \begin{cases} A_{\max} - |x| \, \frac{3}{2} \left(A_{\max} - A_{\mathrm{G}}\right) & \text{if } |x| \geq |y| \\ A_{\max} - |y| \, \frac{3}{2} \left(A_{\max} - A_{\mathrm{G}}\right) & \text{if } |x| < |y| \end{cases} \tag{3.11}$$

$$A_{\mathrm{R},i} \in \mathrm{U}(2A_{\mathrm{G}} - A_{\max}, A_{\max}), \tag{3.12}$$

where A_{G} is the global mean asymmetry of the offered traffic.

To determine how the system is affected by non-uniform asymmetry, we vary the peak asymmetry of the system while keeping the mean asymmetry constant. The mean asymmetry of the system is kept at a constant $A_{\mathrm{G}} = 0.75$ for all scenarios. The peak asymmetries for the hot spot and random scenarios are varied from $A_{\mathrm{max}} = 0.75$ to $A_{\mathrm{max}} = 1$, that is, from 75% downlink traffic to 100% downlink traffic.

Results

Heterogeneous asymmetry reduces the capacity of TDD mode cellular systems. Gradual changes can, however, be handled with almost no degradation by the Soft Switching Point algorithm in conjunction with the DBSE method. Random changes are harder to cope with and especially so if a Static Common Switching Point is enforced globally.

3.6 Sensitivity to BS Placement (Paper T4)

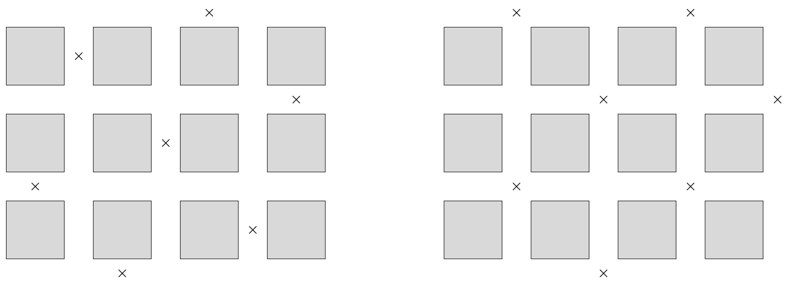
Results in the literature show that IBI is a significant source of outage in TDD mode systems in open macrocell environments and that inter-operator IBI may be the number one source of problems in multi-operator scenarios [33]. Since also Manhattan like environments are rather open along the streets, we are interested in determining how the interference situation depends on the location of the BSs within the Manhattan grid; both IBI and IMI.

Several system properties that affect system performance depend on the physical environment and the location of BSs and MSs; regardless of whether there is IMI and IBI or not. In Manhattan like environments, shadowing is pronounced. Buildings partially absorb and reflect the electromagnetic energy and have a wave-guiding effect. Thus, propagation occurs mainly along the streets and comparatively little energy leaks into side streets [55]. Whether BSs are placed *in* the intersections, Figure 3.8(b), or *between* the intersections, Figure 3.8(a), will influence the interference properties of the system. Both strategies have their merits. A between-intersections deployment can reduce interference compared to an in-intersections deployment, because interference from side streets is greatly attenuated. The drawback is that a turn around a corner in a between-intersections deployment is likely to be associated with a rather sudden cell change. Hence inter-cell handovers can become a problem. With BSs in intersections, on the other hand, the MS is often served by the same BS both before and after the turn. Thereby sudden inter-cell handovers may be avoided.

By supporting heterogeneous or non-unified switching points, OLI, i.e. IMI and IBI, is introduced in the system. In Paper T4, we evaluate how sensitive the system performance is to different BS placements. In particular, the influence of OLI is assessed.

Interference

For mobiles, IMI problems are often limited to a few nearby mobiles. The often clear path between BSs, however, makes BSs susceptible to IBI from several BSs. If BSs are located in the intersections of a Manhattan environment, they can have LOS to twice as many other



(a) BSs *between* intersections

(b) BSs *in* intersections

Figure 3.8: Two different BS deployments.

BSs as if they were placed halfway between the intersections. Thus, the closer BSs are to intersections, the more IBI they are subject to.

For MSs the situation is a bit different. Neither the locations of nor the number of IM interferers are affected by a change in the BS placement. Nevertheless, the amount of IMI may change because by altering the BS locations, the distribution of the transmission powers may change.

We investigate how the different types of interference and system performance depend on the locations of the BSs. Since the systems defined and studied in Papers T1–T3 are doing their best to keep the interference on a manageable level, they are not suitable for telling us which type of interference, i.e. IBI and IMI as well as base-to-mobile and mobile-to-base interference, is behaving how. Hence, to get an understanding of how the different types of intra-system interference behave as a function of the location of the BSs, we first study a worst case interference scenario where all nodes in the system except one create interference.

Then, we evaluate the overall effect of the interference on the performance of a system with asymmetric links using the same system model, see Appendix A, as in Papers T1–T3. Resource allocation is assumed to be done by means of the SSP algorithm with DBSE. For comparisons we also consider the performance of the PSSP and SCSP schemes.

Results

Results show that a correct placement of the BSs is important for efficient use of the radio resource. In a Manhattan like environment, a placement between intersections provides for higher capacities than a placement in the intersections. System performance is rather insensitive as to the exact locations of the BSs, provided that they are not close to or in the intersections.

IBI is higher when BSs are located in the intersections than when they are located between intersections. However, under a constant-received-power power control regime, there is less IMI with BSs placed in the intersections. For a system using the SSP algorithm with DBSE, the capacity degradation caused by an in-intersections deployment is comparable to the degradation of a symmetric system with no IBI or IMI. The degradation is mainly due to increased Base-to-Mobile and Mobile-to-Base interference. The increase in IBI when BSs were placed closer to the intersections was compensated for by the decrease in IMI. No degradation particular to asymmetric traffic was seen.

3.7 Conclusions

We have found that providing native support for asymmetric data flows in a TDD mode system by means of locally adjustable switching points is beneficial for system capacity and also lowers the risk of starvation in the low-volume direction. The capacity increase is substantial, often 30 percent or more. Less for the simplest algorithms and more for the more elaborate ones.

Attempts to estimate missing inter-mobile link gains in order to be able to avoid inter-mobile interference and thereby lower the outage rate seem fruitless. The computational complexity of a feasibility check including inter-mobile link gains also tends to be very high (see discussion below). We have confirmed, however, that excess IMI can be avoided with good results by means of a heuristic protection scheme.

Performance drops when the local expected asymmetry is non-constant in the service area, but with the studied interference avoidance technique the advantage of adjustable switching points over a common switching point is further increased.

In a Manhattan like environment BSs are preferably placed between intersections. This provides for less interference and higher capacity than does an in-intersections deployment. The exact location between intersections is, however, not critical. Would an in-intersections deployment be necessary for other reasons, e.g. to simplify handover, the studied asymmetric systems seem to be no more vulnerable to an in-intersections placement than a symmetric system.

To conclude: if the offered traffic is asymmetric, there is significant capacity to be gained by allowing non-unified switching points. IMI and IBI can be reduced by a fair amount with simple protection schemes and proper placement of the BSs; hence, enabling us to cash in on the increased capacity.

3.8 Discussion

It should be noted that the conclusions arrived at in this Thesis apply in a single-operator scenario. In a multi-operator scenario, adjacent channel interference can seriously impair the performance of the system [31]. With today's filtering technology level, it is not possible to obtain adjacent channel protection much beyond 35 dB. Thus, BS placement, antenna directivity, synchronisation issues etc. become very important. In Manhattan like environments it is possible to use the buildings as shields to reduce the number of Line-Of-Sight (LOS) paths between BSs and between BSs and MSs. In macro-cell environments this is likely not possible and the performance difference between the between-intersections and in-intersections BS deployment schemes in Chapter 3.6 provides a hint to the performance loss that can be expected in more open environments. This supports the view that TDD mode systems are best suited for micro-cell municipal and pico-cell indoor environments.

We have studied a reuse-one TDMA/TDD system with 16 slots per frame. As seen in the results of Papers T1 and T2, the granularity of the resource, visibly affects the performance of the system that uses a common switching point. The larger the number of time-slots, the better the switching point can be matched to the asymmetry of the offered traffic and the smoother the sawtooth shaped curves become (the notches become less deep). Although not studied here, the TDMA/TDD access method can be combined with DS-CDMA such that each time-slot holds multiple RUs, separated by DS codes. Assuming that bandwidth, modulation method and power are fixed, time-slots have to be traded off to code-slots. A lower number of time-slots, however, implies less flexibility and the potential gain by allowing adjustable and multiple switching points decreases.

An important aspect of wireless mobile communication systems is mobility. When MSs move, propagation conditions change and resource re-allocation may become necessary. With multiple switching points, the interference situation is more complicated than with a single switching point. It is therefore reasonable to expect that re-allocations are more frequently necessary in conjunction with multiple switching points. Allocations and re-allocations are computationally intensive and require on-air signalling consuming some of the precious radio resource. The computations are mainly associated with the feasibility check, which calculates the expected SIRs for all links. In a locally centralised TDMA/FDD system, there are a maximum of $N_C(N_C - 1)/2$ MS-to-BS entries in the link gain matrix at any instance of time. Hence, a feasibility check can be done with $\mathcal{O}(N_C^2)$ floating point operations (flops) [11]. In a TDMA/TDD system with non-unified switching points, all MS-to-BS, BS-to-BS and MS-to-MS links must be considered in the feasibility check and the link gain matrix will therefore contain $2 * N_C(2 * N_C - 1)/2$ elements. Thus the computational complexity of the feasibility check of our TDMA/TDD system is also $\mathcal{O}(N_C^2)$. Since the advances in computing power still obeys Moore's law, the added complexity for TDD compared to FDD, is not seen as a problem. The bigger problem, for both FDD and TDD, remains the extra on-air signalling triggered by re-allocations.

One should also recall that there are applications where cellular full coverage is not needed or, from an economic point of view, even desirable. In these applications IMI and IBI is much less of a problem and, thus, the complexity and signalling overhead become low.

Since TDD mode operation only requires one frequency band for full duplex operation, finding usable spectrum for TDD is less involved than finding spectrum for FDD, which require paired spectrum; that is, two frequency bands which are sufficiently separated by a guard band. The ability to operate in unpaired spectrum makes TDD particularly suited for deployments in 'old' spectrum released for new use and for dynamic spectrum allocation.

Chapter 4

Micro-economics and Game Theory

Microeconomics is the study of the economic behaviour of individual consumers and producers and the distribution of goods and capital among them. While the consumers provide capital (and labour) and consume goods, the producers supply goods and consume capital (and labour). Microeconomics investigates the mechanisms that establish relative prices between goods and services and those that allocate the society's resources among its many alternative uses.

For the convenience of the reader, we shall in this chapter describe a few terms and concepts from microeconomics which we will occasionally use in the following chapters. The presentation will be very brief and does not aim at giving a full understanding of the underlying mechanisms, but rather an intuitive acceptance.

We begin with the concepts of market *supply* and *demand*. Figure 4.1 shows an example of basic supply and demand curves for one good in a market. The vertical axis indicates the price of a good; i.e. what sellers get and buyers pay (per unit) for a certain amount or quantity of the good. The supply curve S tells us how much of a good the producers are willing to sell at a given unit price and the demand curve D says how much the consumers are willing to buy of the good when it is offered at a certain price.

Very simplified, we can describe the curves in the following intuitive way: Different people (consumers) value a good differently. Some care for it more than others and are, thus, willing to pay more for it. Some people also have more money and can afford to pay more. When a good is offered at a very high price, only a few believe that the price is fair and can afford to buy it. With decreasing price, more and more can both appreciate the offer and afford it. Hence, the quantity demanded increases with decreasing price. A similar argument can be made for the supply side. If a good shall be sold at a very low price, only a few producers are skilled enough to produce it at a cost low enough to make the business profitable. At a high price, however, even a mediocre producer can make a profit. This is just a rough sketch indicating the general slope of the demand and supply curves, respectively. For a more detailed description of the shapes of the curves we refer to literature on microeconomics.

The demand for a good also depends on what other goods are available. For goods

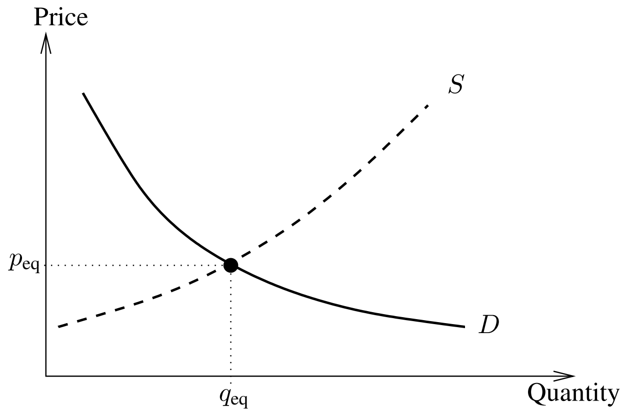


Figure 4.1: Supply and demand. In a perfectly competitive market, supply and demand are in equilibrium. The good is sold and bought at the equilibrium price p_{eq}.

which are usually used together, an increase in the price of one of the goods usually results in a reduced demand for the other good. Such goods are called *complementary* goods. If there are, instead, near equivalents or *substitutes* for a good, an increase in the price of the good results in an increased demand for the substitutes.

The responsiveness of the quantity demanded to a change in the price, ceteris paribus, is called the *price elasticity of demand* and is defined as the relative change in the quantity demanded divided by the relative change in the price, that is, $\frac{\Delta\text{quantity}/\text{quantity}}{\Delta\text{price}/\text{price}}$. While the price elasticity of demand depends on many factors, we will only comment on substitutes. When close substitutes are available, elasticity tends to be high. Conversely, when there are no substitutes, the elasticity is usually low. Price elasticity of demand is only one of a number of elasticities defined in economics. When we refer to elasticity or price elasticity, we shall, however, henceforth mean the price elasticity of demand.

In a perfectly competitive market supply and demand tend to an equilibrium where the price and quantity of the good is given by the intersection of the supply and demand curves, Figure 4.1. A perfectly competitive market is a market with many buyers and sellers, homogeneous good, no entry or exit barriers and perfect information about prices and alternatives in the market. With this definition in mind, we observe that the wireless access market, with few sellers, enormous entry barriers and difficult-to-compare prices, does not meet the requirements of a perfectly competitive market[1].

4.1 Game Theory

Game theory is a branch of mathematics and economics which study interactions based on formal incentive structures, so-called “games”. It is closely related to economics in

[1]There are quite a few Mobile Virtual Network Operators (MVNO), but they are buying network access from only a few sources.

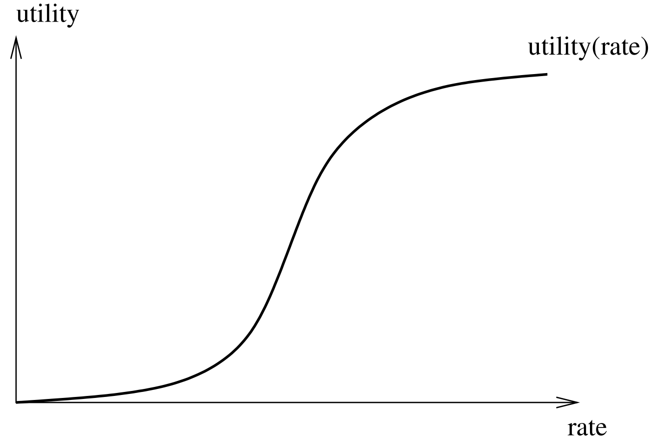


Figure 4.2: Example of a sigmoid utility function.

that rational strategies are sought for situations where the outcome depends, not only on a player's own strategy, but also on the strategies of other players; players who may or may not have different goals. Game theory was first developed by John von Neumann and Oskar Morgenstern [75]. They restricted their attention to a class of problems called zero-sum games, that is, games in which players can only gain at the expense of other players. John Nash extended the theory to non-zero-sum games in the 1950s. Game theorists now search for so-called *Nash equilibria* which are sets of strategies such that no player can improve its own outcome by unilaterally changing strategies while the strategies of the other players remain fixed. Nash equilibria are stable, but not always desirable. It is possible that one or more of the players can have its outcome improved by lifting the restriction of only changing their own strategies. An outcome that cannot be improved without hurting at least one player is called *Pareto efficient*. Nash equilibria are often not Pareto efficient.

The concept of *utility* is commonly used in microeconomics and game theory to describe the satisfaction the users or players experience as a result of their actions. *Utility functions* describe the satisfaction of users as functions of their decision variables. The basic property of a utility function is that it orders the options of a player in order of preference.

In Chapter 5 we will use so-called *sigmoid functions* to describe users' satisfaction as functions of the data rate. A sigmoid function is an S shaped function, Figure 4.2. By using sigmoid functions we assume that for very low rates a small rate increase is not perceived as a great improvement; i.e., the quality of the service will still be poor. For very high data rates, the quality is already so high that an improvement makes no sense. In the region in between, however, the quality is quite dynamic.

Chapter 5

Quality and Pricing Aware RRM

The use of multimedia services is expected to increase rapidly. With new devices and applications appearing every so often, the introduction of more flexible applications and forms of information exchange seems inevitable. Soft or adaptive applications adjust their information exchange according to their ability to conform data transmission to current link conditions so that radio resources can be conserved. However, the added flexibility also increases the complexity of determining the best usage of the radio resources. The issue is further complicated by the fact that user behaviour is influenced by the pricing of the services. Taking users service appreciation into account in the RRM may improve both quality of service and resource utilisation. Thereby also enabling increased operator revenue.

5.1 Related Work

While traditional metrics of RRM performance and Quality of Service (QoS) have been technical, an increasing number of studies consider economic contexts, in which, not only parameters of technical nature, but also tariffs and provider revenue are of importance [43, 80].

In the literature, various researchers have adopted micro-economic considerations and concepts such as, for instance, utilities, profit maximisation and game theory to analyse, and to possibly improve, the performance of next generation wireless communication systems [80, 42, 25, 79, 72, 26].

While all taking somewhat of a micro-economic approach, the objectives of the studies vary:

- stabilisation of rate assignment (in an ATM network) [42][1]),
- alternative implementations of distributed power control [25, 79],

[1] This paper does not treat a wireless communication system, but is cited here because of its early adoption of game theory and use of pricing for communication systems resource management.

- rate allocation in a "fair way" and such that total network revenue is maximised [80] and

- maximisation of total system throughput [72].

In these descriptions, non-technical parameters as user satisfaction or pricing have predominantly been considered only from the users' point-of-view. The role of the provider has mostly been neglected, or the provider has merely been considered to act as an arbitrator or mediator that guarantees the welfare of the users. Recently, however, studies on joint network-centric and user-centric RRM have also been presented [26]. In fact, these studies usually apply game-theoretical concepts, like utility functions, to optimise or simply improve the usage of the radio resource. Some works also use *virtual* prices to improve user politeness so as to control system load and interference [67, 26]. Little is, however, said about the *real* pricing policy of the operator, how the users react to it, and which could be the best strategy for the provider.

It has been shown that by accounting for users' perceptions of reasonable pricing of services, provider revenue may be increased [43, 16]. Also, the introduction of appropriate pricing schemes makes it possible for the service provider to improve the efficiency and robustness of the radio resource management [27, 67]. In fact, when the resource to allocate is scarce and users compete for it, the pricing strategy might be aimed at increasing cooperation among users and preventing allocation to users who have low satisfaction from the service. To the best of the author's knowledge, there is, however, a lack of models that describe the joint impact of pricing and QoS on users' satisfaction. Such a model would be particularly useful for evaluation of RRM performance in systems with elastic traffic [70], i.e. where perceived quality is a more or less continuous function of the throughput (or delay).

Admission Control and Scheduling

Admission Control (AC) is a means to limit the amount of simultaneous communication, and hence, the amount of created interference, in a wireless network in order to protect the communication from excessive interference [9]. Traffic should only be admitted into the network if it does not compromise the QoS of the users who are currently being served, and rejected otherwise. Admission control is particularly important for interference limited systems where the local capacity is determined by the current interference level rather than by the number of available channels. Due to the local capacity varying with the interference level, interference limited systems are sometimes referred to as *soft capacity* systems and channel limited systems, analogously, as *hard capacity* systems. AC has been extensively studied in the literature [9, 53, 34, 4, 39, 74]. In [9], a centralised AC problem in a noise-less system with unconstrained power was approached with an SIR feasibility check prior to admission. An AC scheme based on SIR-thresholds was presented in [53]. Transmitted and Received Power Based Call Admission Control (TPCAC and RPCAC, respectively) blocking new calls when exceeding a pre-set threshold was proposed in [34]. With the SIR and power based schemes, admission decisions were instantaneous and based on local information. They did not capture the strike-back effect of active connections raising their powers as a consequence of an admission. The Soft and Safe (SAS) AC algorithm proposed in [4] introduced interactive distributed AC which let a new user interact with the active connections using limited power before the admission decision is taken. The new user's power constraint is gradually relaxed and the new user is admitted if no target SIRs drop below their threshold, and blocked otherwise. A framework for Multi-Service AC is proposed in [39]. In [74] pricing is considered in multi-service AC for revenue optimisation with QoS guarantees. Pricing was used to enable revenue maximisation in the admission process, but the effects of pricing on users' quality requirements were not considered; nor were the effects of a degradation of the supplied service quality studied. It seems plausible that awareness of and information about users' service appreciation could be used to improve the match between the requested service quality and the supplied service quality.

In a packet switched system, it is the task of a scheduler to distribute the radio resource among users so as to make the "best use" of the resource under the constraint that QoS requirements and/or some fairness criteria are met. The "best use" of the resource is generally taken to be equivalent to maximising throughput, revenue or some fairness criterion. To this end, numerous scheduling algorithms have been presented in the literature, a few of the most important being: Round Robin (RR), Equal Throughput (ET), Fractional Fair (FF), Proportionally Fair (PF) and Relatively Best (RB) scheduling.

The RR scheduler cycles through the users, giving them equal time on the channel. ET scheduling, as the name implies, seeks to provide users with equal throughput and can be viewed as a weighted RR scheduler with (per user) weights that are inversely proportional to the achievable rates. An FF scheduler tries to improve total throughput by giving priority to users with higher achievable rates and can also be described as a weighted RR scheme, with weights that are proportional to the achievable rates. More intricate schemes try to combine the objectives of these simple schedulers. The PF scheduler seeks to improve system throughput by exploiting multiuser diversity, while at the same time providing some degree of (long term) throughput fairness [41]. Similarly, RB scheduling exploits multiuser diversity, but maintains time fraction fairness [52]. There are also schemes that consider delay fairness, for instance: the Largest Weighted Delay First (LWDF) scheduler and its derivatives.

The motivations for these algorithms all assume that higher system throughput translates to greater revenue and/or that fairness is important (or even matters). This is, however, arguable. First because services might not be priced directly proportional to their rate requirements. Second, and maybe more important, users are generally not aware of the situations of other users and so can't compare. Thus, they have no notion of fairness. A user is likely to judge the system based on the perception of receiving decent service, rather than based on fairness from a time or throughput or even price point of view.

In this thesis, we focus on users' service appreciation and take fairness to mean that users perceive the combination of QoS and price they are offered to be decent. Then, from this fairness point of view and assuming that disappointed users may shorten their sessions, the "best use" of the resource could be to maximise the willingness of users to continue their sessions.

The provider may, however, also allow a speculative allocation of resource to some users at the expense of lowering the service level of other users. This can be beneficial if the expected marginal revenue associated with the variation is positive. However, lowering of the service level of a user is risky. If the user no longer perceives the service as usable and affordable he may leave the system, resulting in a loss of revenue. Thus, to enable the use of speculative over-assignment, service appreciation of and resources allocated to already admitted users must be carefully monitored and managed to maximise users willingness to stay in the system.

5.2 Contribution

Papers P5 through P7 (reproduced verbatim in Part II, Chapters 11, 12 and 15, respectively) present a framework for combining service quality and service price as they are perceived by the users into a single metric for user satisfaction, illustrate the impact of users' behaviour on overall system performance and operator revenue and demonstrate how information about user satisfaction can be used in admission control and scheduling to improve user satisfaction and operator revenue. Paper

P5 proposes a general framework, named Micro-Economic Distributed User Service Acceptance (MEDUSA), for joint consideration of QoS and pricing and illustrates it with an example service acceptance model. The service acceptance model describes the probability with which a random user with a particular utility function will accept a particular service offer (a QoS-Price pair). The usefulness of the acceptance model for system performance evaluations is demonstrated and the effect of the system load in a soft-capacity system is discussed.

P6 extends the analysis of P5 to cover both hard- and soft-capacity systems. Focus is on capacity results rather than the particular effect of varying load.

P7 conjectures that speculative over-admission may be used to increase revenue and that most schedulers are suboptimal in that they don't account for pricing effects in their dealing with fairness and determination of the best use of system resources. A speculative admission scheme and a packet scheduler that assigns transmission opportunity to the user that would improve its probability to continue service the most, is proposed and studied.

5.3 MEDUSA (Paper P5)

As both resource management and pricing affect the behaviour of the users, it is reasonable to consider them to be related. Users who are faced with inadequate QoS or with (unjustifiably) high QoS at a very high price are likely dissatisfied. Such dissatisfaction also has an impact on the provider. It affects revenue generation, number of customers and resource allocation efficiency. In an economic sense, an allocation is efficient only if the serviced users are satisfied. An appropriate model of user satisfaction would account for both perceived quality and price. Such a model is still missing in the literature, or there are only simplified models valid for a limited number of cases. Our objective is to outline an, as general as possible, framework that enables us to represent the user satisfaction by means of a probability of accepting a service offer.

In Paper P5 we develop a framework that we call Micro-Economic Distributed User Service Acceptance, or MEDUSA for short, and exemplify its usefulness in system performance evaluations. In this section we summarise the framework and the observations that are made in the paper.

The MEDUSA Framework

The MEDUSA framework describes users' acceptance of service offers from the provider with a probability A which accounts for the trade-off between utility (or perceived QoS), u, and price, p; that is $A = A(u, p)$.

The values of u and p are determined as functions of a parameter r which describes the allocated amount of resource. In our context it is reasonable to identify r with the achieved data rate. The utility functions must satisfy certain properties. In particular, as every user is willing to have as much resource as possible,

$$\frac{\mathrm{d}u(r)}{\mathrm{d}r} \geq 0\,. \tag{5.1}$$

The law of diminishing marginal utilities from economics, stating that:

$$\lim_{r\to\infty} \frac{\mathrm{d}u(r)}{\mathrm{d}r} = 0\,, \tag{5.2}$$

should also apply. In fact, there are intrinsic limitations which prevent the users from experiencing QoS beyond a certain limit; i.e. there is an upper bound to the appreciation of a service. Thus, we replace Equation (5.2) with the stricter requirement:

$$\lim_{r\to\infty} u(r) = l. \tag{5.3}$$

Also the price is represented by a function $p(r)$ (in general, dependent on the rate) for which no particular assumptions are made, even though it seems reasonable to require that $p'(r) \geq 0$.

The acceptance probability function need to meet certain requirements. It should increase for increasing utility and decreasing price. That is,

$$\frac{\partial A}{\partial u} \geq 0 \qquad \frac{\partial A}{\partial p} \leq 0. \tag{5.4}$$

Furthermore, there are also asymptotic conditions for the probability A. Being a probability, it has to be limited to the range $[0, 1]$. In more detail, let us consider $A(u, p)$ when the price is fixed to a finite non-zero value. In this case, by varying u, we can say that:

$$\lim_{u\to 0} A(u,p) = 0 \qquad \lim_{u\to\infty} A(u,p) = 1. \tag{5.5}$$

The second part of relationship (5.5) does not represent a real case because an infinite utility is not reachable, see Equation (5.3). It is, however, useful to introduce it to have a more suitable mathematical expression. For a given finite value of the utility greater than zero, it is possible to write the dual relationships for the price, i.e.:

$$\lim_{p\to 0} A(u,p) = 1 \qquad \lim_{p\to\infty} A(u,p) = 0. \tag{5.6}$$

Note that the above properties do not state the behaviour of $A(u,p)$ when both u and p are zero or infinite. This is because these values are arbitrary. In fact, $A(0,0)$ is the acceptance probability of a blocked user; i.e a user that is not admitted, regardless of its value of A. Similarly, $A(\infty,\infty)$ is a case that can not occur, as it would violate both conditions of limited utility and limited allocable resource.

A suitable expression for the MEDUSA model, which assures the validity of Equations (5.5)–(5.6) is:

$$A(u,p) \triangleq 1 - e^{-C\cdot u^{\mu}\cdot p^{-\epsilon}}, \tag{5.7}$$

where C, μ, ϵ, are appropriate positive constants. μ and ϵ describe the users sensitivity to changes in utility and price, respectively. The value of C can be determined as follows: if it is estimated that the best rate assignment (giving utility l) with a certain price p has a residual probability q of being *refused*, then

$$C = -\frac{p^{\epsilon}\log q}{l^{\mu}}. \tag{5.8}$$

Observations

A strong point of the MEDUSA model is the tunability, which makes it similar to the Cobb-Douglas curves [73], widely used in economics. They are curves which represent the connection between price p and demand d, by assuming a relationship like:

$$d(p) = \kappa p^{-\alpha}, \tag{5.9}$$

where κ and α are appropriate constants. The advantage of using Cobb-Douglas curves in economics is that they are easily tractable even though the true relationship might not be exactly a direct proportionality between d and the αth power of p. The parameter α, called "elasticity" in micro-economics, must, on the other hand, be appropriately chosen in order to have a good match of the theoretical curve.

Also in our case, the parameters ϵ and μ represent, roughly speaking, the sensitivities[2] to the price and the utility. Thus, the parameter can be easily adapted to different kinds of markets, in particular ϵ and μ can be changed (possibly through realistic measurements) if the users become more or less sensitive to price or quality variation, respectively.

[2]The proper economic term should be "elasticity" (like the Cobb-Douglas exponent α). However, in the following we will call these parameters "sensitivities" to avoid confusion with the property of the traffic being elastic.

When it comes to pricing, one should observe that pricing strategies include a lot of different proposals [68, 17] and it is not clear whether all of them can be considered realistic. A basic property of a realistic pricing policy is a conceptual simplicity which allows users to predict the economic consequences of their actions[3]. Thus, we require the pricing policy to be set a priori to entering a service agreement. In this work we discuss two different policies:

- a simple usage-based pricing where the price $p(r)$ is linearly related to the rate r, i.e., $p(r) = kr$, with a given constant k, and
- flat price strategy, where the price p is the same for all rates.

The simplicity of these two policies implies their probable presence in next generation networks, even though a more complicated pricing scheme may turn out to be better for both the users and the provider. However, the model can be applied to every fixed pricing relationship known by the users a priori.

In the numerical examples of this paper, we use the probability A to model the behaviour of users in a resource assignment scheme in which the only choice left to the users is whether they want to accept the service or not. Thus, the operator revenue becomes:

$$R = \sum_{i=1}^{N} R_i = \sum_{i=1}^{N} p_i A(u_i, p_i), \tag{5.10}$$

where the users are considered to be numbered from 1 to N and their corresponding utilities and prices u_i and p_i, respectively. We assume that the operator has acquired knowledge about the relationship between assigned rate and user utility; a relationship which may be different for different users. Users' utilities are modelled by sigmoid functions. We further assume that the provider adopts a centralised and greedy rate assignment, where users are assigned the rates that correspond to a certain fixed marginal utility ϑ. For more details about the evaluation environment, we refer to Paper P5 in Part II of this Thesis. Figures 5.1 and 5.2 show the revenue behaviours of the flat and the usage-based pricing schemes, respectively. Price variations affect the revenue through the acceptance probability. One should note, however, that price is not the only factor which determines the revenue variation. In fact, the revenue varies even if the price is kept constant when the marginal utility target ϑ is changed. A lower value of ϑ corresponds to a higher perceived quality and, thus, a higher price can be justified. Though all curves exhibit the same general behaviour, both the maximum revenue and the price for which the maximum revenue is achieved are different. As a general comment, it is shown that the pricing and the management strategy heavily affect each other. A good choice of the price for one RRM approach may not be as good for another RRM approach. A fact to keep in mind when planning or evaluating pricing and RRM schemes.

[3] This does not, however, imply that the pricing schemes of two operators should be easily comparable. Operators may, in fact, want to limit the cross-operator comparability of their pricing schemes to make users' decision to change operators more difficult. We restrict our concern to predictability within the realms of a single operator.

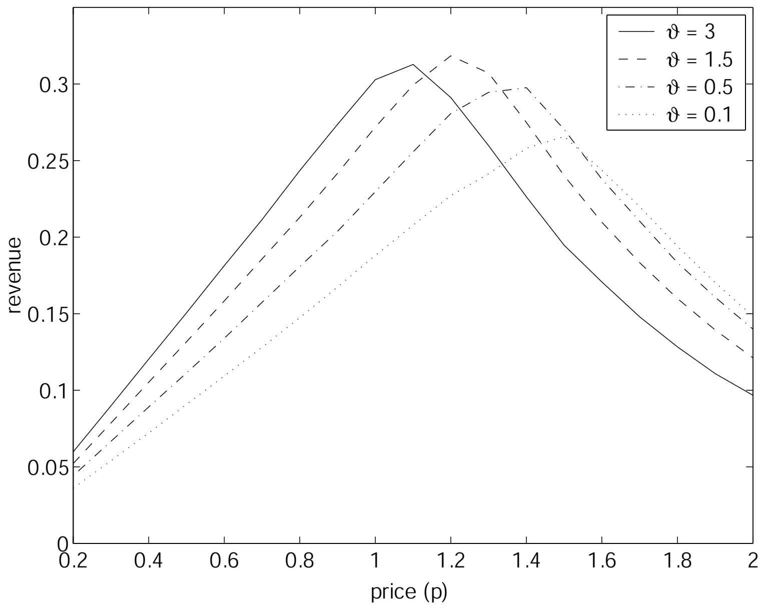


Figure 5.1: Provider revenue for flat price, 160 users, as a function of the price p.

The MEDUSA model can also be explored analytically to give useful insights. For example, Equation 5.10 can be used as a basis for a maximisation problem. Consider the two scenarios where the capacity is either larger or smaller than the demand. In the former case, the users which accept the service can always be accommodated within the constrained capacity. If the network is conservatively dimensioned, it is likely that this situation occurs for every case of low load. In this case we can consider the RRM as an unconstrained optimisation problem. Thus, our model can be used to derive properties of a suitable pricing policy that at least gives a satisfactory revenue for the case of no congestion. For a congested system, these results have to be modified according to the capacity constraint.

5.4 MEDUSA in Hard- and Soft-Capacity Systems (Paper P6)

Paper P6 extends the investigation of Paper P5 to cover both CDMA-like and TDMA-like systems; i.e., systems with hard and soft capacity respectively.

It is observed that the peak performance (with respect to price) of the CDMA-like network is slightly higher in terms of achievable revenue. In general, however, the performance is similar. A more interesting phenomenon can be observed in Figure 5.3, where it is shown that the "knee", which is also seen in Paper P5, is not present in the TDMA-like network. That is, the CDMA network presents a counter-intuitive behaviour where the

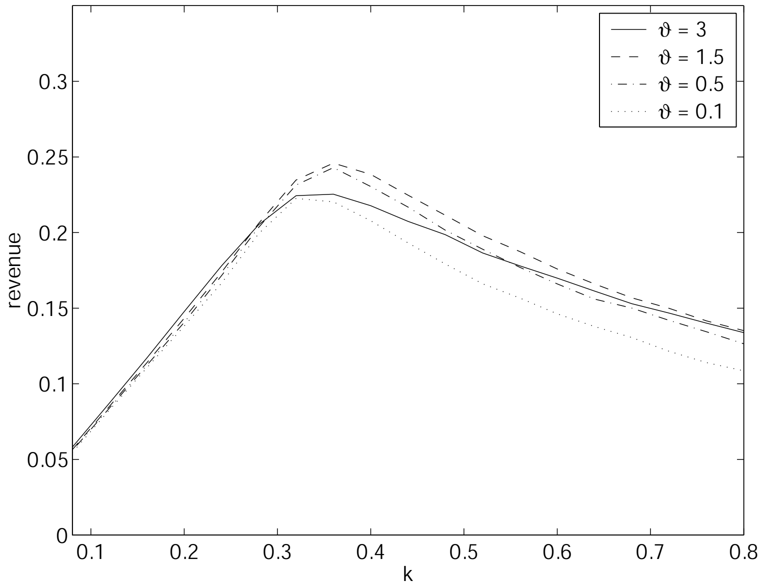


Figure 5.2: Provider revenue for price $p(r) = kr$, 160 users, as a function of k.

admission rate can increase with higher price.

The explanation is related to the fact that a CDMA system is interference limited and users with poor propagation conditions consume more of the system's power and interference budgets than users with good propagation conditions. Some users are unable to achieve a decent QoS even at maximum transmission power. If the price is sufficiently low they will, however, still find the price to performance ratio reasonable and accept the service offer. Since servicing these users is expensive in terms of system resources, it limits the number of users that can be admitted.

With a higher price, on the other hand, less of the users with bad propagation conditions agree with the price to performance ratio and, hence, less of them contend for service. Therefore, users with better channels who were previously blocked because of lack of resources can now be admitted. Users with better channels consume less resources and, thus, can be admitted in larger number than the users with bad channels that now decline the service offer due to the higher price.

This indicates that soft-capacity networks are not suited for a greedy allocation strategy, but that pricing can be used to "select" better users and, hence, increase both admission rate and throughput/revenue.

The TDMA system is more robust to a greedy allocation, but since it is channel-limited, it can not fully exploit favourable path loss. This may eventually explain the lower maximum system throughput and revenue compared to the CDMA system.

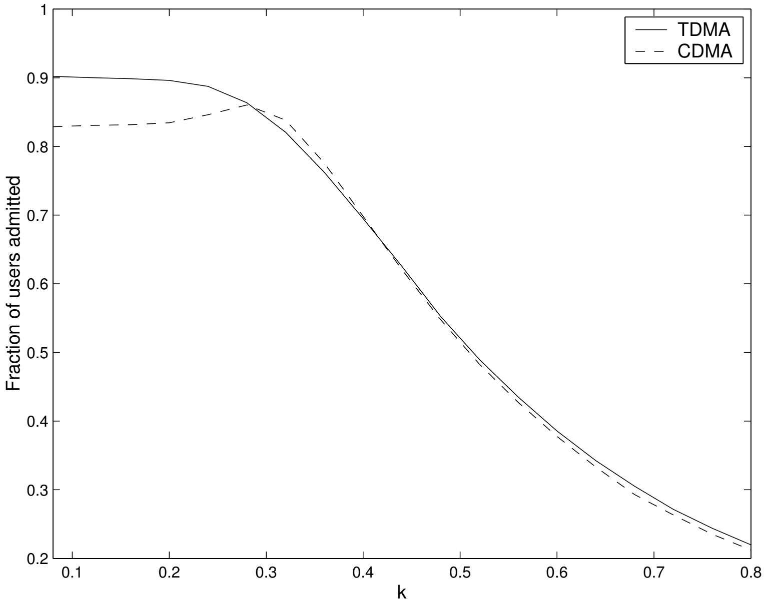


Figure 5.3: TDMA–CDMA comparison, price $p(r) = kr$, admission rate, $\vartheta = 1.5$

5.5 Speculative Resource Allocation (Paper P7)

The scheduling of packets can be seen as a particular case of resource assignment, the goal of which is to obtain a satisfactory provider's revenue. To this end, the scheduler should, according to our previous discussion, give priority to assignments which leave the users satisfied with respect to both QoS and price paid. Unsatisfied users are expected to be likely to leave the service and, thus, deteriorating revenue.

If the scheduling is efficient in terms of maintaining high user satisfaction per resource unit, we may admit more users. We may in fact even attempt to admit users past the capacity of the system. With information about users' service appreciation we may make a qualified guess about whether or not users currently in the system would accept a quality degradation and thereby allow more admissions. In Paper P7 we introduce these kinds of *Speculative Admission Control* (SAC) and *Service Perception Aware* (SPA) scheduling and investigate what can be gained in terms of user satisfaction and operator revenue. In particular, we focus on the forward direction of a high speed packet access system and realtime streaming services. For the purpose of the system evaluation we also extend and use the MEDUSA model. Special handling of handoffs is not treated in this work.

Extension of MEDUSA

The original proposal of the MEDUSA framework only treated *initial* service offerings and suggested a service *acceptance probability*. Here we are, however, concerned not only with users' initial acceptance of a service contract at the time of admission, but also with users' continuing willingness to hang on to the contract once they have entered. Hence we extend the Acceptance Probability model to an *Accept and Stay Probability* model defined on the service interval $[t_{\text{in}}, t_{\text{out}}]$. We do this as follows: First we notice that if a user, with a priori information about the price, requests the rate r_{req}, it is reasonable to assume that the user will accept the utility and price corresponding to r_{req}. That is, the user is satisfied if she gets what she asked for. Second, would the rate actually supplied to the user improve the satisfaction of the user, she will keep using the service. However, would r change in a way that would reduce the satisfaction of the user, we assume that she only sticks to the service with a probability conditioned on her previously being satisfied. Thus, we define the probability that a user hangs on to the service at time t_2, given that the contract at time t_1 has been accepted, to be:

$$S(t_1, t_2) = \begin{cases} \min\left(1, \frac{\min\limits_{t_1 \le t \le t_2} A(t)}{H(t_1)}\right) & \text{if } t_{\text{in}} \le t_2 < t_{\text{out}} \\ 0 & \text{otherwise,} \end{cases} \tag{5.11}$$

where $t_{\text{in}} \le t_1 \le t_2 \le t_{\text{out}}$, $A(t)$ is $A(u, p)$ evaluated at time t and

$$H(t_1) = \begin{cases} A(u(r_{\text{req}}), p(r_{\text{req}})) & \text{if } t_1 = t_{\text{in}} \\ \min\limits_{t_{\text{in}} \le t \le t_1} A(t) & \text{if } t_{\text{in}} < t_1 \le t_{\text{out}} \end{cases} \tag{5.12}$$

represents the lowest satisfaction level that has already been accepted, Figure 5.4.

The probability of accepting the *session* for the wanted interval $[t_{\text{in}}, t_{\text{done}}]$, i.e. the probability of a *successful exit*, then becomes:

$$A^S(t_{\text{in}}, t_{\text{done}}) = \min\left(1, \frac{\min\limits_{t_{\text{in}} \le t \le t_{\text{done}}} A(t)}{A(u(r_{\text{req}}), p(r_{\text{req}}))}\right), \tag{5.13}$$

where $t_{\text{done}} \ge t_{\text{out}}$.

If we further assume that the user always requests the rate, $\hat{r}$, that yields the best compromise between the given utility and price, and therefore maximises $A(u, p)$, Equation (5.13) simplifies to:

$$A^S(t_{\text{in}}, t_{\text{done}}) = \frac{\min\limits_{t_{\text{in}} \le t \le t_{\text{done}}} A(t)}{A(u(\hat{r}), p(\hat{r}))}. \tag{5.14}$$

That is, the probability of successful exit equals the MEDUSA acceptance probability for the worst utility-price combination of the session, given that the requested combination is satisfactory. The probability of successful exit is difficult to determine a priori. In fact,

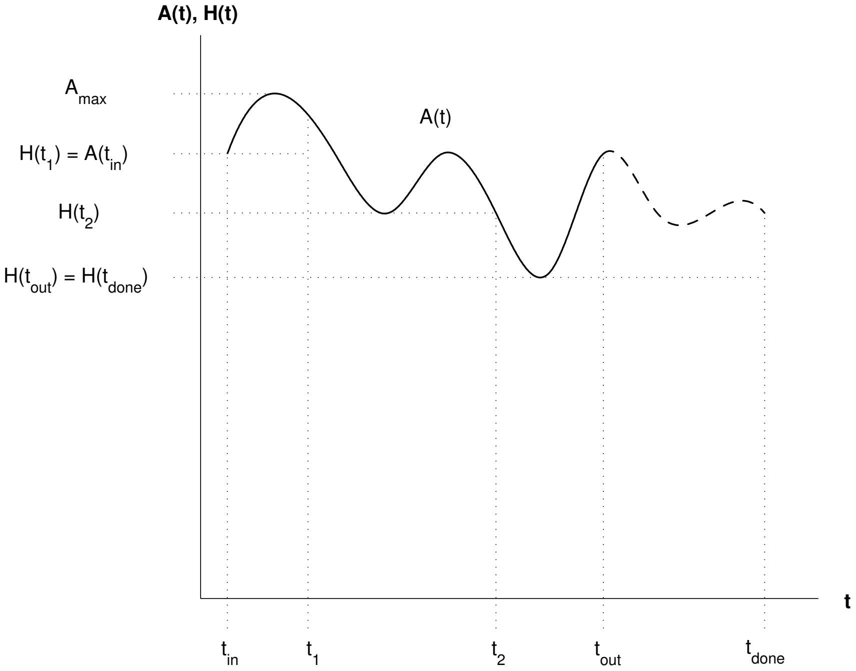


Figure 5.4: Illustration of the MEDUSA extension.

if the session is prematurely aborted ($t_{\text{out}} < t_{\text{done}}$), determining it a posteriori is equally difficult.

This far, the equations have referred to a single user. When dealing with multiple users in the following, we shall use indices, i, to refer to individual users; e.g. $S_i(t_1, t_2)$ for the stay probability of user i at time t_2, given that the service was satisfactory at time t_1, and $u_i(r_{\text{req},i})$ for user i's utility of the rate requested by the same user.

The above model does not explicitly account for the fact that user decisions are neither instantaneous nor based on instantaneous values. This is, however, partially hidden by averaging in the scheduler. More involved modelling is outside the scope of this paper and left for further research.

Speculative Admission Control

When a session request arrives at the base station, an admission control mechanism evaluates if there is enough free resources to accommodate the request. Let us assume, for a while, that the session is admitted if the data rate corresponding to the expected value of the C/I on the link, minus some fade margin, is sufficient to carry the rate r_{req} in the available time-slots and refused or blocked otherwise. The objective of this AC is simply to maintain a high user satisfaction by means of avoiding congestion in the network. Due to fading, the QoS will, however, vary over time. One way to stabilise the QoS is to increase the fade margin in the admission test, effectively reserving some resources. This, however, limits the efficiency of the resource utilisation since a large fraction of the time there will be 'reserved' but unused resources. While the assumption of a C/I-based AC is convenient for illustrating the problem, the same holds for other AC schemes where resources are reserved to protect active sessions.

To improve on this situation from both technical and economic efficiency points of views, we propose an admission control mechanism that not only considers the technical capacity of the system, but also proactively estimates and considers users' willingness to accept a provided QoS and price. Since it involves forecasts, of both user and system behaviours, we refer to this type of admission control as *Speculative Admission Control* (SAC). Similar to the example above, SAC estimates whether there are enough free resources to accommodate the new request. If there are enough resources and the revenue that is predicted to come from the session is deemed sufficient, all is well an the user is admitted. If, however, the system is unable to meet a request from a user, i.e. the estimated resource requirement is greater than the available resources, the admission control system may offer the user an admission at *reduced rate* if the expected revenue would thereby increase. The operator may reduce the rate only for the user under consideration or *speculate* in the effects of a global reduction of the QoS and distribute the reduction over some or all of the users. In this paper, we assume that the operator bases his speculation on equal relative rate reduction for all users. A user which rejects the counteroffer is classified as being denied service.

If the pricing strategy is such that

$$\sum_{n=0}^{N'} R_i' > \sum_{n=0}^{N} R_i \quad \text{when} \quad \sum_{n=0}^{N'} r_i' = \sum_{n=0}^{N} r_i \text{ and } N' > N, \tag{5.15}$$

where N (N') is the number of users and R_i (R_i') and r_i (r_i') are the revenue contribution and rate of user i, respectively, it may be beneficial for an operator to offer reduced rates even when there are sufficient resources. This opportunity is, however, not available in paper P7 where we consider the price to be directly proportional to the rate.

To formalise the admission rule with an example, we first assume that users arrive to the system one at the time and that the operator makes an on-line admission decision without information about future arrivals (neither their number nor the average size of their requests). This leads to a sequential greedy allocation of resources to the users. A process, in which the operator can prevent users from being admitted, i.e. reject users, by assigning them the zero rate, $r_i = 0$. The acceptance probability of the zero rate is 0. Consider now the system where there are $i-1$ active users and a new request for the rate $r_{\text{req},i}$ arrives from user i at time t_i. Let us, hypothetically, assume that the user is admitted and let t_i^- and t_i^+ be the two instants of time just before and just after the admission. At t_i^- the rate assignment vector is $\mathbf{r}^{(t_i^-)} = (r_1^{(t_i^-)}, r_2^{(t_i^-)}, \ldots, r_{i-1}^{(t_i^-)}, r_i^{(t_i^-)} = 0)$, and at t_i^+ it is $\mathbf{r}^{(t_i^+)} = (r_1^{(t_i^+)}, r_2^{(t_i^+)}, \ldots, r_{i-1}^{(t_i^+)}, r_i^{(t_i^+)})$. If $\mathbf{r}^{(t_i^+)} = (r_1^{(t_i^-)}, r_2^{(t_i^-)}, \ldots, r_{i-1}^{(t_i^-)}, r_{\text{req},i})$ is not feasible, the operator may exploit traffic elasticity [70], i.e. rate tunability of data services, and adjust the rate assignments in any justifiable[4] way to make the rate assignment feasible. In this work we have restricted the study to adjustments of the kind

$$\mathbf{r}^{(t_i^+)} = k \cdot (r_1^{(t_i^-)}, r_2^{(t_i^-)}, \ldots, r_{i-1}^{(t_i^-)}, r_{\text{req},i}), \tag{5.16}$$

where k is a value in the range $(0, 1]$. The revenues corresponding to t_i^- and t_i^+ are

$$R^{(t_i^-)} = \sum_{j=1}^{i-1} p(r_j^{(t_i^-)}) + 0 \qquad \text{and} \tag{5.17}$$

$$R^{(t_i^+)} = \sum_{j=1}^{i-1} p(r_j^{(t_i^+)}) S_j(t_i^-, t_i^+) + p(r_i^{(t_i^+)}) A(u_i(r_i^{(t_i^+)}), p(r_i^{(t_i^+)})), \tag{5.18}$$

respectively, and the admission rule formally becomes:

$$\begin{aligned} \text{if } R^{(t_i^-)} < R^{(t_i^+)} \quad &\rightarrow \quad \text{admit user } i \\ \text{if } R^{(t_i^-)} \geq R^{(t_i^+)} \quad &\rightarrow \quad \text{reject user } i. \end{aligned} \tag{5.19}$$

Service Perception Aware Scheduling

Maximising the provider benefit involves determining which users should be served and at what service level (rate, priority etc.) they should be served. It also involves actually delivering on the determined service level. While a packet scheduler could possibly be designed to take all these issues into account and continuously optimise the resource allocation, it can be argued that the objectives of AC and scheduling are sufficiently different to vouch for a separation of the functionalities. We shall assume the position that, except for best-effort services where there are no QoS guarantees, users expect the QoS to remain approximately constant for the duration of a session. Unsatisfied users are expected to be likely to leave the service, thus, deteriorating revenue. Hence, the main objective of the scheduler is to achieve a high user satisfaction in a resource efficient way. To this end the scheduler should, according to our previous discussion, give priority to assignments which leave the users satisfied with respect to both QoS and price paid. Revenue and profitability considerations are privileges of the AC. Discrimination among users may be implemented by the scheduler, but is a decision made by the AC, or a mobility management function.

We propose and evaluate a packet scheduler for realtime streaming services with the aim to maximise user satisfaction for the set of admitted users. In particular, we focus on the forward direction of a high speed packet access system and realtime streaming services. Rather than equalising the data rates or on-air times we want to maintain a high service appreciation. Time-slots are, therefore, assigned to the users that would end up the

[4] When adjusting the rate assignments of admitted users, one has to consider both the value added by revenue from the additional user and the potential loss of revenue when users who become unsatisfied with the service leave as a result of the adjustment.

least satisfied would they *not* get the resource. More formally, assume that users initially request the rates, $\hat{r}_i$, that maximise their service acceptance probabilities, that is:

$$\hat{r}_i = \arg\max_r A(u_i(r), p_i(r)), \tag{5.20}$$

resulting in the initial acceptance probabilities

$$\hat{A}_i = A(u_i(\hat{r}_i), p_i(\hat{r}_i)). \tag{5.21}$$

Let $\bar{r}_i(n)$ be a moving average of the achieved data rate by the time of slot n:

$$\bar{r}_i(n+1) = \begin{cases} [1 - \frac{1}{\tau}]\bar{r}_i(n) + \frac{r_i(n)}{\tau} & \text{if } i \text{ served in slot n} \\ [1 - \frac{1}{\tau}]\bar{r}_i(n) & \text{otherwise} \end{cases}, \tag{5.22}$$

where $r_i(n)$ is the instantaneous data rate during slot n and τ is the time constant of the smoothing filter. Then, the user to be scheduled for slot n, $j(n)$, is

$$j(n) = \arg\max_{i:\ \bar{r}_i(n)[1-\frac{1}{\tau}] < \hat{r}_i} \frac{\hat{A}_i - A\left(u_i(\bar{r}_i(n)[1 - \frac{1}{\tau}]), p_i(\bar{r}_i(n)[1 - \frac{1}{\tau}])\right)}{\hat{A}_i}. \tag{5.23}$$

Results

The SPA scheduler is compared to a weighted proportional fair scheduler with and without SAC in a single cell hexagonal environment where both log-normal shadow fading and Rayleigh fading are considered. When SAC is not employed, feasibility is only checked versus average SIR as described above. This simple feasibility check AC is, in the following, denoted fAC. Both SAC and the simplistic AC use a 3 dB fade margin.

The benefits of SAC is also evaluated in a 19-cell environment with perfect scheduling; i.e. assuming that the resource allocations made by the base station can always be sustained. The evaluations are performed by means of simulation with the system parameters summarised in Tables 5.1 and 5.2.

As performance measures we use *dissatisfaction rate* and *goodput*. We take the dissatisfaction rate to be the sum of blocking and the *Premature Session Termination Rate* (PSTR), i.e. the fraction of the users that decide to leave the system earlier than they first intended because of poor service. The goodput is defined as the chargeable throughput which is the throughput that users have requested and, thus, are willing to pay for. Throughput in excess of the requested is considered nonchargeable.

Results show that the SPA scheduler performs similarly to the modified weighted PF scheduler which can be assumed to use the channel efficiently. Taking users' service perception into account, the SPA scheduler fares better than the PF scheduler with respect to PSTR. Without SAC, the blocking is significantly higher than the PSTR, however, and, thus, the difference between SPA and PF may or may not be important. It will, ultimately, depend on how users respond to terminated sessions. For comparison, one may note that dropping is commonly considered to be more of a nuisance than blocking. Users who are

Table 5.1: Summary of system parameters

Parameter	Symbol	Value
cell radius	r_{cell}	500 m
loss @ 1 m from tx	$L_{1\text{m}}$	28 dB
propagation loss exponent	α	3.5
log-normal fading std.dev.	σ	8 dB
Doppler frequency	f_{Doppler}	40 Hz
slot duration	t_{slot}	5 ms
max tx power	p_{max}	33 dBm
bandwidth	W	5 MHz
interference + noise (constant)	I_{const}	-80 dBm
available rates	$\mathcal{S}$	{ 0, 64, 128, 256, 512,1024,2048} kbps
C/I thresholds	$\mathcal{G}$	$2^{R/W} - 1,\ R \in \mathcal{S}$
smoothing filter time constant	τ	50 slots
session duration	T_{dur}	1000 slots
$P_{\text{tx}} = 33\text{dBm} \rightarrow SINR = -9.4\text{dB}$ @ cell border $\rightarrow$ 773 kbps		

Table 5.2: Accept-and-Stay probability parameters

Parameter	Symbol	Value
utility parameter	ζ	$2 \div 20$
utility parameter	K	$[0.05 \div 1] \cdot 256$ kbps
acceptance prob. parameter	μ	2.0
acceptance prob. parameter	ϵ	4.0
acceptance prob. parameter	C	$-(2048 \cdot 10^3)^4 \cdot \log(0.9)$
price	p	1 unit per bps

dissatisfied with the service have many options which affects the system on different time scales. For instance: they may retry, leave the system temporarily or permanently etc..

With SAC, blocking and, thus, the dissatisfaction rate is reduced. Results show that this translates to significant capacity gains. Figure 5.5 shows the effective chargeable throughputs (the "goodputs") vs dissatisfaction rates for SPA and PF scheduling with fAC and SAC. Results show that, with a scheduler that provides the users with acceptable QoS-to-Price performance, the use of SAC may increase revenue by around 30% compared to fAC while keeping user satisfaction approximately constant. Alternatively, the user satisfaction can be correspondingly improved for fixed revenue.

5.6 Conclusions

Operator revenue depends on how users respond to both radio resource management and pricing. The MEDUSA model, which defines an acceptance probability considering the joint effect of user utility and price, enables us to include economic considerations in the study of RRM. Results show that even similar RRM strategies can behave differently when economic parameters like pricing strategies and user demand are taken into account. Thus,

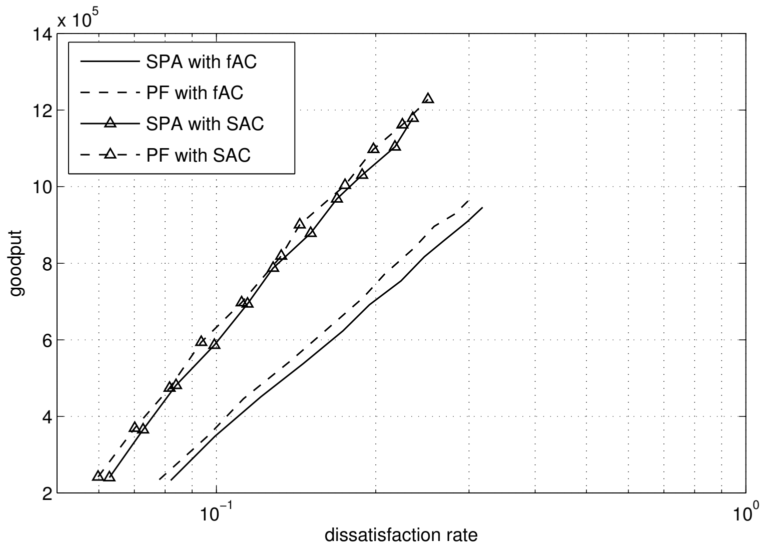


Figure 5.5: Goodput plotted versus dissatisfaction rate.

to efficiently control the performance of the system, it is suggested that the selection and tuning of RRM and pricing policies be addressed jointly.

The *Premature Session Termination Rate*, PSTR, is an important addition to the traditional system performance measures: blocking rate and outage rate. Unlike the outage rate, which is a network-centric measure of service availability, the PSTR is a user-centric measure of the perceived service quality. A high PSTR indicates unreliable or unstable service and may, in the long run, motivate users to change service providers.

Results show that, with a scheduler that provides the users with acceptable QoS-to-Price performance, the use of SAC may increase revenue with about 30% compared to a more conservative feasibility check admission control while keeping user satisfaction approximately constant. Alternatively, the user satisfaction can be correspondingly improved for fixed revenue. The importance of the scheduling order varies with the stringency of the quality requirements. With loose requirements, trying to increase users appreciation on a packet basis is less beneficial. Taking user appreciation into account is still important, however, for improving fairness with respect to user location. Results also show that, while a good scheduling policy is essential, proper admission control is important too.

To sum up, the MEDUSA model allows useful insights to be gained about the RRM strategy. The economic aspects of RRM should not be neglected, for they not only affect performance, but also require several strategic choices to be made. It is imperative for the provider to take into account these aspects; thus, our model can be useful to gain understanding of them and improve the RRM in real systems.

5.7 Discussion

The results were obtained by means of simulations. Presented values are averages of many realisations with a total number of users ranging from 10000 to 30000. Rate estimates, such as those of blocking, admission, and PSTR, were generally based on at least a few hundred occurences of the studied events. The PSTR, for the case of SPA with fAC, was an exception counting only approximately 10 occurences. Thus, one should avoid drawing conclusions from the PSTR for SPA with fAC.

One may argue that a user who requests service would deterministically accept a service offer which matches the requested level of service. However, the assumption that the tariff is fixed and made available to a user prior to entering a service contract[5] does not necessarily mean that the tariff is available to the user prior to the service request. Thus, the meaning and application of a service acceptance probability depend on the service negotiation procedure. If the tariff is unknown to the user at the time of the service request, the user will defer its decision to use the service until the user has seen the *service offer* of the operator. Hence, in this scenario, a service request initiated by a user has a non-zero probability of being declined by the same user. If, on the other hand, the tariff is available to the user prior to the service request, one could expect the user to have evaluated it before placing the request. In this case, the decision to use the service is made prior to placing the service request and the probability of accepting what the operator eventually provides the user with would rather be conditioned on the user accepting the requested service level (or any level resulting in a greater satisfaction); that is, only service levels giving lower satisfaction than the requested have a non-zero probability of being rejected. To summarise, we note that, at some point, users always decide whether or not to accept an offered service. However, whether this point can be directly observed by the operator/system or is hidden, in for instance the arrival process, depends on the scenario. A scenario with multiple operators offering subscription free services may have observable decisions whereas the decisions in a single operator scenario with long term subscriptions may be hidden in the arrival process.

In the investigations of Papers P5–P7 we have focused on the potential gains of accounting for users' service appreciation in the RRM. We have not thoroughly studied how this information about the users can be acquired. Thus, we do not suggest that the presented schemes are good for implementation. They are meant to indicate how much could be gained by, or rather, how much is lost by not accounting for users' service appreciation in the RRM. How much information is needed and how it can be acquired need to be studied further.

Effects of user dissatisfaction on the input traffic have not been considered in this work. Just as blocking and dropping result in new arrivals when users retry, one can imagine that some of the dissatisfied users who terminated their sessions want to retry and, thus, increase the load experienced by the system, which may in turn lead to more sessions being terminated due to poor quality if the admission control is inadequate. While we

[5] With a service contract we here mean an agreement valid for a session or a burst of packets rather than a long term subscription agreement granting the user the right to submit service requests to the network of a specific operator.

believe that, with adequate admission control and reasonable levels of terminated sessions, the effects will be marginal, we leave the issue open for further investigation.

The sensitivities to utility and price need not be constants. They may for instance vary with the type of service. From economics we know that the price elasticity of a good depends on the availability of substitutes. If there are close substitutes, price elasticity tends to be high, whereas if there are no close substitutes, price elasticity is low. In 2003 the voice service price elasticity was expected to exceed 1.5 in the North American market in what could be an increasing trend [76]. We believe, however, that the price elasticities of wireless data and multimedia traffic are higher than that of voice. There are more alternatives and we have not yet come to depend on them as heavily as we do on wireless voice. Thus, we have assumed a higher price sensitivity. Our contribution here is a general framework, which is an engineering issue. We believe that this also needs integration with other research fields such as economics and operational research. It may have implications on these fields as well as on the technical implementation of real networks.

Chapter 6

Creation of Forwarding Incentive

It has been shown that the capacity and/or the coverage of a wireless network can be improved by breaking long distance connections down into multiple shorter connections or links by means of a technique called multihopping, Figure 6.1. The capacity and coverage gains are achieved when the aggregate energy efficiency of the short paths is higher than that of the long path. This is true for many practical situations. Multihopping, or more precisely ad hoc multihopping, was initially investigated for the purpose of tactical communication in areas where little or no infrastructure is deployed.

In recent literature, the application of ad hoc multihop principles to cellular systems has shown that the potential gains of coverage extension and capacity enhancement are substantial [30, 66]. These studies, however, implicitly assume that the users of the network are willing to cooperate in the exchange of information.

A cooperative user behaviour is typically achievable in situations where the different units invest their resources for achieving a common goal. This may be the case in military networks, but is far less obvious in a cellular network, where relaying is expected to be done through other selfish terminals. On one hand, the power savings and rate improvement from letting one's traffic be relayed by an intermediate node to the base station are rather tangible. On the other hand, there is, little immediate incentive to participate

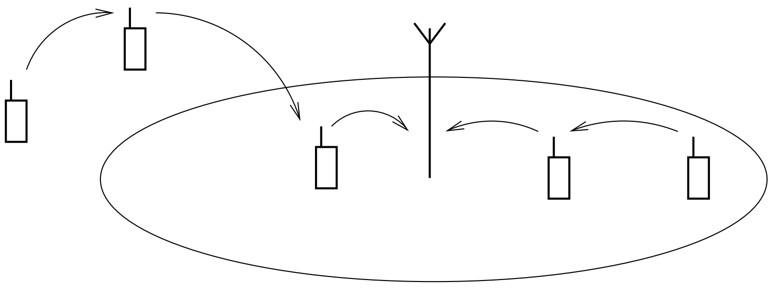


Figure 6.1: Multihop wireless network.

in the forwarding of someone else's traffic, since in forwarding, a relay sacrifices its own energy and, traditionally, also its own bandwidth; may it be a time-slot in a TDMA system or a code in a CDMA system. The loss of bandwidth limits the opportunity to own communication and we shall refer to it as an opportunity cost.

6.1 Related Work

Finding mechanisms that enable cooperation between terminals is essential for reaping the benefits of ad hoc cellular systems. Several efforts to create forwarding incentive and improve fairness have been made in the literature. Yet, we do not believe that an ultimate solution is provided. Many of the proposed solutions are based on micro-payment schemes [40]. There, compensation for relaying information is achieved by transferring transmission rights (tokens) from the source node to the nodes that cooperated in the delivery. However, in cellular networks multihopping is mainly beneficial to users far from the base stations. Therefore tokens are expected only to be transferred from the cell borders to the inner regions and not vice versa, resulting in an unfair system and an eventually stagnant business model.

There are also several examples where pricing has been suggested for solving resource allocation and sharing problems [67, 18]. A pricing scheme designed for improving cooperation in multihop networks is presented in [37]. In this scheme terminals which invest energy and part of their own bandwidth are reimbursed when they are relaying traffic. The incentives created this way, however, barely balance the large opportunity costs, and cooperation is only partially achieved.

Compensating relays by means of economic rewards is not entirely straightforward. Partially because, in practical systems, bandwidth and energy are not substitutes. For instance, limited peak power and limited modulation options restrict the extent to which lost bandwidth can be compensated for with energy. In the context of economic compensation, the result may be a high price. The fact that the sacrifice is two-dimensional also complicates the compensation of it. Yet, it seems that the trade-offs between bandwidth, energy and price commonly made in the literature are unnecessarily complex.

6.2 Contribution

Paper I8 (reproduced verbatim in Part II, Chapter 14) proposes a reduction of the trade-off space for relay nodes in multihop cellular networks. It introduces a concept of "resource delegation", where a source node owning a particular resource may choose to transfer the authority to use the resource, or a fraction thereof, to a relaying node. Thus, the relaying node does not have to sacrifice its own bandwidth; i.e., bandwidth is removed from the trade-off space of a relay. To stimulate forwarding, the paper also proposes a pricing scheme where a relay is paid a reward by either the source terminal or the operator whenever packets are forwarded through the relay. Since communication with the base station is associated with a good or a perceived utility for the terminals, the reward needs to sufficiently compensate for an *expected* loss of utility implied by using energy for forwarding instead of selfish transmission. The benefits of the proposed schemes are evaluated from both user and operator perspectives and compared to a regular, single-hop, TDMA cellular network.

6.3 Resource Delegation and Rewards (Paper I8)

To eliminate the bandwidth investment of the relay we introduce the concept of *resource delegation*. Resource delegation is a transfer, from one node to another, of the authority to use a particular resource, while preserving the interests of the first node. Instead of the relay sacrificing its own bandwidth, the source node delegates part of its bandwidth resource to the relay. In a TDMA system this can be accomplished by means of subslotting; i.e. the basic resource unit, the time-slot, is divided into source and relay subslots. The source node restricts its transmission to the source-subslot and delegates the relay subslot to be used by the relay. We shall focus on TDMA systems, but note that resource delegation can be done similarly in a CDMA system by means of subcoding and multicode transmissions. With the introduction of resource delegation, energy remains the only resource to be compensated. Thus, both the problem of determining a proper compensation is simplified and a lesser compensation is required.

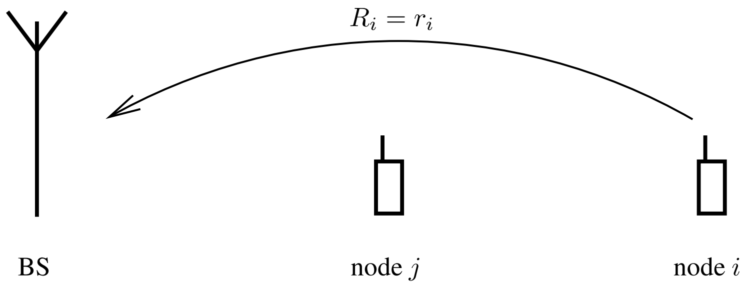
We supplement the resource delegation with a pricing scheme where a relay is paid a *reward* whenever packets are forwarded through the relay. The rewards can be directly transferred from each transmitter to its relays, or they can be provided by the network operator investing part of its revenues. In the latter case, the users will jointly pay for the total energy expenditures in the network[1]. In order to quantify the rewards we define utility functions which describe the different nodes' benefits and costs of communication. Economic compensation is related to energy and opportunity loss through the utility functions.

Our incentive mechanism is general enough to be applied in many different cellular systems. However, in order to gain insight in how the resource delegation and rewarding scheme performs, we provide an example that targets the uplink part of a full-coverage TDMA system for low/medium bit rates. Assuming that the base stations are the sinks of all ongoing communications, the resource delegation and rewarding schemes are directly extended into a routing scheme. In this implementation, multihopping (MH) is never forced. Rather, it is the result of an agreement between the different entities which use utility functions to quantify the good associated with establishing a multihopping path. For interpretational convenience connections between a source terminal and a BS are restricted to be either single-hop (SH) direct connections or two-hop connections through one other terminal (relay).

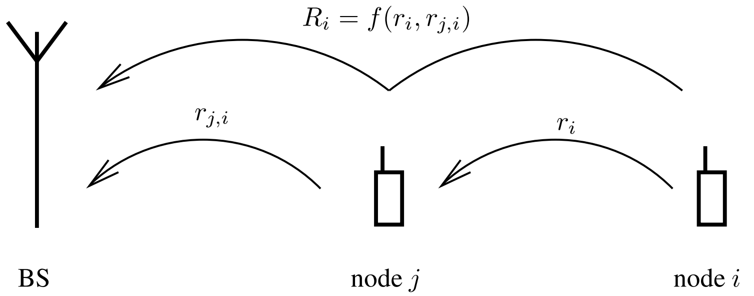
Data Rates

In order to describe the utility functions of the nodes in the network, we need to define some data rates. The *path* data rate R_i, experienced in the end-to-end connection between

[1] This is justified by the expectation that most users will, while not directly, at least indirectly benefit from the reduced interference and/or increased capacity.



(a) single-hop connection



(b) multihop connection

Figure 6.2: Illustration to definition of data rates in multihop wireless network.

a source node i and its assigned BS, may be different depending on if the connection is performed through a direct link or by means of multihopping. In the first case, R_i coincides with the *link* data rate r_i, achieved in the direct link with the BS, Figure 6.2(a). In the multihopping case, illustrated in Figure 6.2(b), R_i depends on a composition of the rate r_i, between source node i and relay node j, and the rate $r_{j,i}$, between relay node j and the BS. For consistency, r_i always denotes the rate on the first link and $r_{j,i}$ shall be interpreted as the rate that is achieved on the link between relay node j and the BS on the behalf of source node i (one relay may forward traffic from more than one source).

The resource delegation is then performed by dividing the original time-slot T_S into two adjacent parts with shares θ_S^i, for the source-subslot, and $\theta_R^{j,i}$, for the relay-subslot, Figure 6.3. The values of these slot shares are set so that the same amount of bits is carried in both links during T_S. This leads to the following expression for R_i:

$$R_i(p_i, p_{j,i}) = r_i \cdot (1 - M_i) + \frac{r_i \cdot r_{j,i}}{r_i + r_{j,i}} \cdot M_i, \tag{6.1}$$

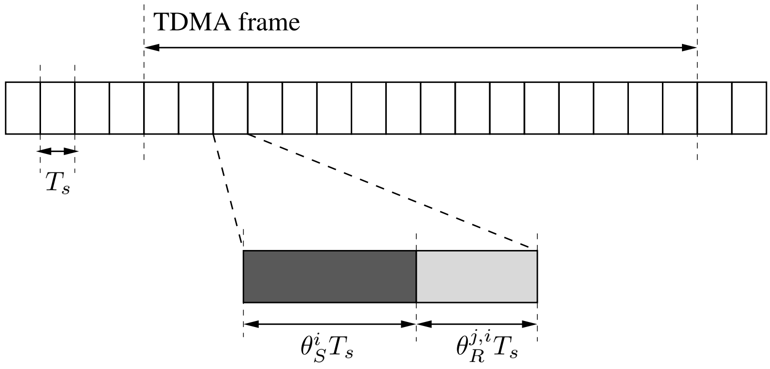


Figure 6.3: TDMA subslotting.

where, M_i represents a boolean that is equal to one whenever i selects multihopping and j participates. Assuming fixed interference levels and stationary user locations, r_i is a function of the power p_i used by source i and of the distance between source i and the receiving node, be it the BS for the case of a direct connection, or relay j for the case of a multihop connection. Under the same assumptions, $r_{j,i}$ is determined by the relay power $p_{j,i}$ and the distance between relay j and the BS.

Utility Functions

The utility functions of the nodes assign numerical values to levels of satisfaction and form the basis for routing and forwarding decisions. For informed decisions the utility functions should capture both good and ill associated with transmission. For a source node, there is the good of outputting data and the ills of battery drainage and traffic charge.

As previously considered in the literature [67], we identify the source's objective with the maximisation of the number of transferred bits per unit of energy spent in the transmission process. In TDMA-like systems, where slots have fixed duration, this goal is equivalent to maximising the transmission rate per unit of power. Further, in our model, we assume that the network operator has adopted a pricing scheme in which the service fee is proportional to the provided path data rate, through a constant λ. Any given source, i, is aware of this pricing policy and can select its preferred path data rate, R_i, by tuning p_i. In case of multihopping through another node, j, i has to pay an additional forwarding reward. On the other hand, for a given R_i, multihopping may lead to a reduced energy expenditure. In our implementation, the reward is assumed to be proportional to R_i with a proportionality constant of μ_S. The utility formula that is introduced in this paper has the following structure:

$$U_S^i(\lambda, \mu_S, p_i, p_{j,i}) = \frac{R_i}{c + \theta_S^i \cdot p_i} - (\lambda + \mu_S \cdot M_i)\, R_i. \qquad (6.2)$$

The denominator of the first term represents the total power invested in the transmission. It has two components: p_i and a constant c. The constant represents a fixed power dissipation in the terminal units, and also stabilises the utility function in the proximity of $p_i = 0$. We assume that terminals, or user agents within them, are responsible for setting the value of c so that it reflects the fixed dissipation cost of the terminal.

Any terminal that has not been scheduled for transmission in the current time slot can potentially act as a relay. In our scheme, a relay, j, that forwards traffic on behalf of a source, i, obtains a reward that is proportional to R_i through a proportionality constant μ. Furthermore, R_i and the amount of invested energy, $E_{j,i} = p_{j,i}\theta_R^{j,i}T_S$, can both be tuned by varying $p_{j,i}$.

The selection of $p_{j,i} = 0$ represents the decision of not participating in the forwarding; this choice does not provide any net benefit or loss, and therefore sets a zero utility level for the relay. Conversely, the relay joins the forwarding if there exists a $p_{j,i} > 0$ that results in a positive utility. In both these cases the relay's decisions are selfishly made with the purpose of maximising its benefits. This means that a relay selects the value of $p_{j,i}$ that provides the largest reward while considering the benefit that is potentially lost by investing $E_{j,i}$ in relaying instead of in own transmissions.

In order to quantify the benefit lost, denoted $U_{Lost}^{j,i}$, we need to estimate the utility that can be obtained through selfish communication. In fact, if $E_{j,i}$ is not invested in the current slot, it will be used by the relay in a future instant of time. This means that both the physical location in which this energy will be used and the type of the future connection (SH or MH) are unknown. For these reasons, we propose to quantify $U_{Lost}^{j,i}$ in terms of the expected benefit obtainable in a direct connection $\overline{U_{\text{self}}}$, and the expected energy cost $\overline{E_{\text{self}}} = \overline{p_{\text{self}}}T_S$ associated with it. Thus, the relay utility function is described by the following expression:

$$U_R^{j,i}(\lambda, \mu, p_i, p_{j,i}) = \underbrace{\mu \cdot R_i}_{\text{Reward}} - \underbrace{\frac{\theta_R^{j,i} \cdot p_{j,i}}{c + \overline{p_{\text{self}}}(\lambda)} \cdot \overline{U_{\text{self}}}(\lambda)}_{U_{\text{lost}}^{j,i}}, \tag{6.3}$$

where, the expectations $\overline{U_{\text{self}}}$ and $\overline{p_{\text{self}}}$ are both evaluated under the assumption of uniform terminal distribution over a circular cell area of radius d_{cell}. Using Equation (6.2), these quantities are computed in the following way:

$$\overline{U_{\text{self}}}(\lambda) = \int_0^{d_{\text{cell}}} \max_{p_i(x)} U_S^i(\lambda, 0, p_i(x), 0)\, \frac{2x}{d_{\text{cell}}^2}\, dx \tag{6.4}$$

$$\overline{p_{\text{self}}}(\lambda) = \int_0^{d_{\text{cell}}} \arg\max_{p_i(x)} U_S^i(\lambda, 0, p_i(x), 0)\, \frac{2x}{d_{\text{cell}}^2}\, dx. \tag{6.5}$$

For the network operator we assume that the utility coincides with the network revenues. By varying the coefficients of the pricing scheme, the operator influences the data rates selected by the users, and in turn its revenues. Thus, the operator utility function may be written:

$$U_O(\lambda, \mu, \mu_T, \boldsymbol{R}) = \sum_{i=1}^{N} R_i \cdot [\lambda - M_i \underbrace{(\mu - \mu_S)}_{\mu_O}], \tag{6.6}$$

where $\boldsymbol{R}$ is the vector of source node data rates, $(R_1, R_2, \ldots, R_N)$, and $\mu_O R_i M_i$ represents the amount of forwarding reward provided by the operator to the relay supporting user i. The triple $t = (\lambda, \mu, \mu_S)$ is here defined as a set of tariff coefficients. This set defines a broad class of pricing schemes, that include different degrees of collaboration between the operator and the transmitter in the provision of the relay reward.

Routing Scheme

Above we have defined utility functions for the different entities in the system. These utility functions form the basis of the path selection or routing scheme. It is the objective of each entity to maximise its utility. The means to achieve the objective varies with the function of the entity. Terminals and relays vary their transmitting powers, p_is and $p_{j,i}$s, respectively. The network adjusts the price, λ, and the relaying rewards, μ_O and μ_S. We assume that all entities consider only their own interests in a simultaneous search for their respective maxima. The source node selects the path that yields the highest utility. Would none of the possible multihop routes provide a utility greater than that of a direct connection, the source node will unilaterally decide to make a direct connection.

Results

The proposed resource delegation and rewarding schemes are evaluated in terms of user utility, user data rate, operator revenue and the interference seen by neighbouring base stations; all as functions of price, λ, and reward, μ. For reference, we use a single-hop system where λ and μ are set so that operator revenue is maximised. The evaluation is performed by means of uplink simulations of a macrocell TDMA system where terminals and relays are assumed to be uniformly scattered over the system area. So as to reduce complexity and not to unnecessarily obfuscate the fundamental behaviour of the proposed schemes, a single-cell system without fading is studied[2]. Interference is estimated by averaging the powers that are seen by a virtual first tier of neighbouring cells. Data rates are further assumed to follow the Shannon capacity formula. The system parameters are summarised in Table 6.1.

Results show that, for both user induced rewards ($\mu_S = \mu$) and operator induced rewards ($\mu_S = 0$), there exist domains, as functions of λ and μ, where all of our four performance measures are jointly improved. The domains are depicted in Figure 6.4. It can be seen that, while the regions of mutual benefit are smaller when one considers worst case interference, they are still significant. For small rewards, the area is bounded upwards in price by decreasing transmitter utility. For greater rewards, the limit is instead set by increasing interference. Since revenue decreases with lower price, price is in general the parameter that limits the regions of mutual benefit downwards. In Figure 6.4 we have indicated the four points, in the t-planes ($\mu_S = \mu$) and ($\mu_S = 0$), which:

I) maximise SH revenue,

[2]Large scale fading in a single-cell system with intra-cell channel orthogonality, would mean little more than a randomisation of terminal locations.

Table 6.1: Summary of system parameters

Parameter	Symbol	Value
cell radius	r_{cell}	1000 m
loss @ 1 m from tx	$L_{1\text{m}}$	21 dB
propagation loss exponent	α	3.5
bandwidth	B	5 MHz
noise density	N_0	-174 dBm/Hz
tx independent power expenditure	P_{const}	160 mW
# of users	N_u	100
$P_{\text{tx}} = 30\text{dBm} \rightarrow SNR = 7\text{dB}$ @ cell border		

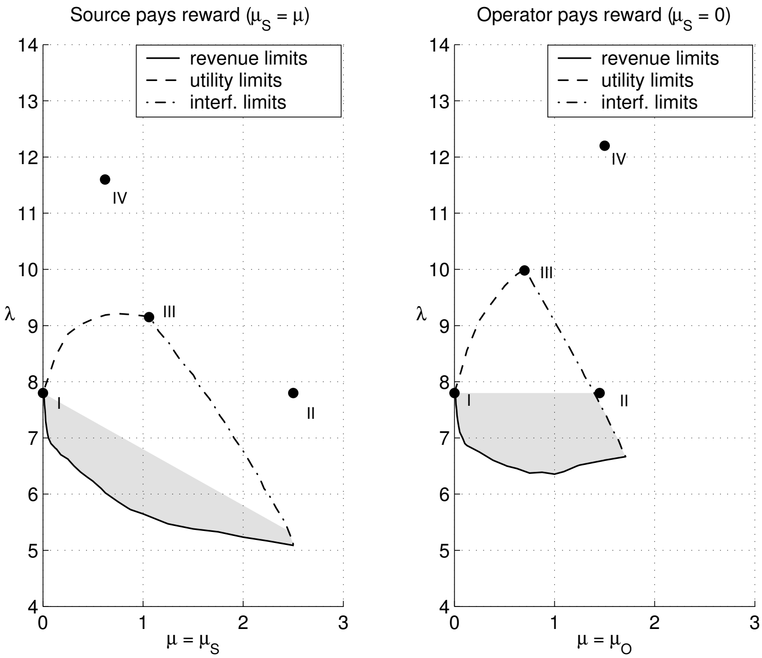


Figure 6.4: Regions of joint improvement in the four variables: revenue, interference, user utility and data rate. Shaded area indicate reduced effective price as experienced by the source node.

II) maximise MH revenue given that λ is equal to the price that maximises SH revenue,

III) maximise MH revenue within the domain of joint improvement and

IV) maximise MH revenue without constraints on λ and μ.

The gains associated with each point, relative to point I, are presented in Table 6.2. The price-reward pair that maximises operator revenue virtually deletes the user utility and as

Table 6.2: Multihopping gains relative to case I ($\lambda = 7.8$)

Source pays reward		revenue	data rate	utility	interference
II	$\mu = 2.5$ $\lambda = 7.8$	1.5	1.5	1.2	1.6
III	$\mu = 1.1$ $\lambda = 9.1$	1.6	1.4	1.0	1.0
IV	$\mu = 0.62$ $\lambda = 11.6$	2.1	1.4	0.15	4.0

Operator pays reward		revenue	data rate	utility	interference
II	$\mu = 1.5$ $\lambda = 7.8$	1.2	1.5	2.0	1.1
III	$\mu = 0.70$ $\lambda = 9.8$	1.7	1.4	1.0	1.0
IV	$\mu = 1.5$ $\lambda = 12.2$	2.9	2.1	0.19	23

such does not appear to be a good choice for an operator who wish to maintain a good customer relation. There is, however, significant revenue gain also in the domain of joint improvement. The shaded areas in Figure 6.4 indicate the pricing parameters that correspond to a reduction of the effective price paid by the originating terminal. The existence of such a region is important since price may be more tangible to users than is energy or power and because users are usually hostile to price raises.

Looking at a single pair of one source node and one relay node, results show that multihopping provides substantial gains compared to single-hopping for a wide range of locations of the relay. In the direction towards the base station, the relay locations, for which the originating and relaying terminals agree on multihopping, are limited by the sourcing terminal, because of diminishing power savings. In the direction away from the base station, the limit is, instead, set by the relay, because of unjustifiably high power consumption.

6.4 Conclusions

By transferring bandwidth from an originating node to a relay, the opportunity cost for a relay becomes purely energy related. Thus, the assessment of the cost for forwarding is simplified. With simplified cost assessment, pricing based rewarding schemes, such as the proposed ones, become efficient forwarding incentivisers. In fact, when used in conjunction with resource delegation, the proposed pricing schemes create an opportunity for the operator and the users to mutually benefit; the operator from reduced extra-cell interference and increased revenue and the users from higher data rate and lower price.

6.5 Discussion

When we introduced the concept of Resource Delegation in Paper I8, we assumed that a slot could be partitioned into two subslots with arbitrary precision anywhere in the slot. This is a rather optimistic assumption which needs to be addressed in future work. A more realistic assumption would be to define a few fixed boundaries where a time-slot can be split. Switching from receive to transmit mode also introduces some overhead since the switch introduces an interval during which data can neither be received nor transmitted. For reference, IEEE Std 802.16-2004[3] requires a subscriber station to switch from receive to transmit mode in less than 50μs [1].

Another issue that need to be addressed is the route discovery. Potential relays need to be found and a contract negotiated. The implementational aspects and overhead of route discovery was not considered in Paper I8, but is work in progress. We currently consider an on-line route discovery process where the source node first sets up a direct connection to the BS and then, at a regular or random interval, piggy-backs a relay discovery message on his data stream[4]. The relay discovery message contains information which enable a potential relay to calculate the link gain between the source and itself and to perform a utility optimisation locally. If the potential relay concludes that a contract would be mutually beneficial, it replies to the source node in a time-slot, specified in the discovery message, to which the source node has committed to listen. If the potential relay decides to not participate it remains silent. If the source node is out of coverage, no direct connection is setup, but discovery messages are sent and potential relays requested to reply in one of their own slots. If a deal is made, the relay forwards the source node's service request to the BS which assigns resources to the source terminal[5]. Due to mobility the source and/or the relay may want to terminate the forwarding contract and the source node has the options of either finding a better relay, re-establishing a direct connection or terminating the connection. The source node may also terminate a forwarding contract at any time, would it find a better relay.

So far we have only studied resource delegation on error free channels. In wireless communication, there are, however, frequently errors and retransmissions are therefore needed. How retransmissions should be handled in conjunction with resource delegation has yet to be investigated.

[3] Air interface for fixed broadband wireless access systems.

[4] If idle nodes only listen occasionally to the BS, the initial discovery message may be sent by the BS on behalf of the source node. This, however, prevents out of coverage operation.

[5] The source node was itself unable to request service since it was out of range.

Chapter 7

Summary and Future Work

In the introduction we conjectured that the economic efficiency and profitability could be improved both by better 'hard', technical, efficiency and by better accounting for users' 'soft' properties of service appreciation and willingness to pay.

On the 'hard' efficiency account, we have shown that implementing support for asymmetric links can improve the efficiency of (service) production in TDD mode wireless network with asymmetric traffic. That is, more traffic can be handled with the same system resources. Compared to FDD, TDD also enables a reduction of hardware complexity by not needing a duplex filter and by allowing more hardware to be shared between the receiver and the transmitter of a unit. This may increase the economic efficiency in the production of the equipment. Further, as spectrum for TDD can be made available more readily, spectrum related costs associated with system deployment and operation may be reduced. The benefits of TDD and support for asymmetric links are readily available for systems providing high-rate spotty coverage. For systems aiming at full coverage and tight reuse, however, proper measures must be taken to control inter-mobile- and inter-base-station-interference.

We presented the MEDUSA model framework for taking users' service appreciation and willingness to pay into account in performance evaluations of wireless networks with elastic traffic. Assuming that user satisfaction depends on both the quality and the price of the service, numerical experiments show that the economic efficiency of an RRM scheme is affected by the pricing scheme.

While the model is useful for evaluation of the efficiency of pricing and RRM schemes, one should be careful when trying to find a pricing policy that maximises profitability for a particular RRM scheme. Willingness to pay is based on the economic value to a user of an offered service, not on the cost of production of the service. Thus, it is likely wiser to first get a good idea of what pricing policies are viable and then use the model to assess the economic efficiency of RRM schemes under these pricing policies. This procedure may require a study of user values, but as a by-product the results of such a study can also be used to improve the user satisfaction model.

We introduced the concepts of speculative resource management to exploit traffic elasticity and improve resource utilisation. With speculative admission control, users may be admitted to a full system at the expense of a degradation of the QoS of some or all users, if the expected total revenue would thereby increase. A user's willingness to pay is related to the provided good. In a wireless network there is, however, no one-to-one correspondence between produced good and resource consumption. The cost of production heavily depends on the propagation conditions of the link over which the good is delivered. Thus, from an economic efficiency point of view, giving priority to users who pay relatively more for the resource is beneficial. Adding a user with good propagation conditions can improve the total revenue even though the contributions of the previous users decrease. Results indicated significant revenue gain with speculative admission control. Service perception aware scheduling was evaluated as a means to improve resource utilisation, but yielded only marginal gain compared to a weighted proportional fair scheduler.

For the third area studied in this Thesis, i.e. multihopping in cellular networks, economic efficiency was both the goal and one of the means to achieve it. By means of a resource re-distribution scheme called Resource Delegation we eliminated the bandwidth bottle neck of the relays. We combined Resource Delegation with economic compensation for the energy expenditures of the relays and were able to achieve significantly increased operator revenue with maintained or improved user utility. Assuming that the added complexity of keeping track of reward transactions is negligible, profitability was correspondingly improved.

7.1 Future Work

The algorithms and schemes which we have studied have, mostly, illustrated principles and are not ready for deployment. Several implementational issues which may affect real-world performance remain to be further investigated. The results of our work indicate, however, that further studies are worthwhile. We believe that the effects of mobility and protocol issues are topics of particular interest for future work.

In combination with delays, mobility causes resource allocation decisions to be based on information which is no longer accurate and, thus, reduce the performance of the RRM. Mobility also generates signalling overhead since changing propagation conditions require modifications to the resource assignment and since terminals leaving or entering a cell need to be handed over to or from another BS. Thus, before proceeding too far with solving other implementational issues, the performance implications of mobility need to be studied.

The resource delegation and rewarding schemes of Paper I8 need a protocol for route discovery and set up. There will be a trade-off between how fast routes are discovered and the amount of overhead that is introduced. The overall performance penalty needs to be studied to assess how much of the gains remain. Investigation of: if and how 'subslotting' is best implemented is also needed.

For quality and pricing aware RRM, one should study which information about users' satisfaction can be made available to the system and how. We would also like to refine the user satisfaction model, match it to real data and evaluate the performance of different RRM schemes in a more realistic setting.

In all contexts of co-operation there is a risk that someone will try to gain an advantage at somebody else's expense. Thus, it would be interesting to study the possibility of cheating in the contexts of both quality and pricing aware RRM and resource delegation with rewards. If there are possible exploits, how can they be prevented?

Bibliography

[1] Air interface for fixed broadband wireless access systems. IEEE Std 802.16-2004, The Institute of Electrical and Electronics Engineers, Inc., 3 Park Avenue, New York, NY 10016-5997, USA, October 2004. ISBN 0-7381-4070-8.

[2] COST Action 231. Digital mobile radio towards future generation systems. Final report EUR 18957, The Office for Official Publications of the European Communities, L-2985 Luxemburg, 1999. ISBN 92-828-5416-7.

[3] E. Anderlind and J. Zander. A traffic model for non-real-time data users in wireless radio networks. *IEEE Communications Letters*, 1(2):37–39, March 1997.

[4] M. Andersin, Z. Rosberg, and J. Zander. Soft and safe admission control in cellular networks. *IEEE/ACM Transactions on Networking*, 5(2):255–265, 1997.

[5] H. R. Anderson and J. P. McGeehan. Optimizing microcell base station locations using simulated annealing techniques. In *Proc. IEEE VTC 1994*, volume 2, pages 858–862, 1994.

[6] L. Badia, M. Lindström, J. Zander, and M. Zorzi. Demand and pricing effects on the radio resource allocation of multimedia communication systems. In *Proceedings Globecom 2003*, 2003.

[7] L. Badia, M. Lindström, J. Zander, and M. Zorzi. An economic model for the radio resource management in multi-media wireless systems. In *Proc. ASWN 2003*, 2003.

[8] L. Badia, M. Lindström, J. Zander, and M. Zorzi. An economic model for the radio resource management in multi-media wireless systems. *Computer Communications*, 27(11):1056–1064, July 2004.

[9] N. Bambos and G.J. Pottie. Power control based admission policies in cellular radio networks. In *Proc. IEEE GLOBECOM '92*, volume 2, pages 863–867, 1992.

[10] J.-E. Berg. A recursive method for street microcell path loss calculations. In *Proc. IEEE PIMRC'95*, pages 140–143, 1995.

[11] M. Berg. *Capacity, Robustness, and Complexity of Radio Resource Management in Bunched Personal Communication Systems*. Licentiate Thesis, KTH, Royal Institute of Technology, 1999.

[12] M. Berg, S. Pettersson, and J. Zander. A radio resource management concept for 'bunched' personal communication systems. In *Proc. Multiaccess, Mobility and Teletraffic for Personal Communications Workshop, MMT '97*, 1997.

[13] L. Chen *et al.* A dynamic channel assignment algorithm for asymmetric traffic in voice/data integrated TDMA/TDD mobile radio. In *International Conference on Information, Communications and Signal Processing (ICICS)*, pages 215–219, 1997.

[14] H.-S. Cho, M. Y. Chung, S. H. Kang, and D. K. Sung. Performance analysis of cross- and cigar-shaped urban microcells considering user mobility characteristics. *IEEE Transactions on Vehicular Technology*, 49(1):105–116, January 2000.

[15] J.C.-I. Chuang. Performance limitations of TDD wireless personal communications with asynchronous radio ports. *Electronics Letters*, 28(6):532–534, March 1992.

[16] R. Cocchi, S. Shenker, D. Estrin, and L. Zhang. Pricing in computer networks: Motivation, formulation and example. *IEEE/ACM Transactions on Networking*, 1(6): 614–627, 1993.

[17] C. Courcoubetis, F. P. Kelly, V. A. Siris, and R. Weber. A study of simple usage-based charging schemes for broadband networks. *Telecommunications Systems*, 15: 323–343, 2000.

[18] C. Courcoubetis and R. Weber. *Pricing Communication Networks (Economics, Technology and Modeling)*. John Wiley & Sons Ltd, Chichester, West Sussex, England, 2003.

[19] Shuang Deng. Empirical model of www document arrivals at access link. In *Conf. Record IEEE ICC'96*, volume 3, pages 1797–1802, 1996.

[20] C. Drane, M. Macnaughtan, and C. Scott. Positioning GSM telephones. *IEEE Communications Magazine*, 36(4):46–59, April 1998.

[21] Radio Equipment and Systems (RES). Digital european cordless telecommunications (DECT). a guide to DECT features that influence the traffic capacity and the maintenance of high radio llink transmission quality, including the results of simulations. Technical report ETR 042, ETSI Secretariat, F-06921 Sophia Antipolis CEDEX - FRANCE, July 1992.

[22] Telefonaktiebolaget LM Ericsson. Channel allocation in a telecommunications system with asymmetric uplink and downlink traffic. International Application Published under the Patent Cooperation Treaty (PCT), January 2000. International Publication Number: WO 00/01188.

[23] R. Esmailzadeh, M. Nakagawwa, and E. A. Sourour. Time-division duplex cdma communications. *IEEE Personal Communications*, 4(2):51–56, April 1997.

[24] ETSI. Selection procedures for the choice of radio transmission technologies of the UMTS (UMTS 30.03). Technical Report 101 112 version 3.2.0, ETSI, April 1998.

[25] D. Famolari, N. Mandayam, D. Goodman, and V. Shah. *Wireless Multimedia Network Technologies*, chapter 1: A New Framework for Power Control in Wireless Data Networks: Games, Utility and Pricing, pages 289–310. Kluwer Academic Publishers, September 1999.

[26] N. Feng, S.-C. Mau, and N.B. Mandayam. Pricing and power control for joint user-centric and network-centric resource allocation. *IEEE Trans. Commun.*, 52(9):1547–1557, September 2004.

[27] R. J. Gibbens and F. P. Kelly. Resource pricing and the evolution of congestion control. *Automatica*, 35:1969–1985, 1999.

[28] D.J. Goodman, R.A. R.A. Valenzuela, K.T. Gayliard, and B. Ramamurthi. Packet reservation multiple access for local wireless communications. *IEEE Transactions on Communications*, 37(8):885–890, August 1989.

[29] J. K. Han, B. S. Park, Y. S. Choi, and H. K. Park. Genetic approach with a new representation for base station placement in mobile communications. In *Proc. IEEE VTC 2001 – Fall*, volume 4, pages 2703–2707, 2001.

[30] T.J. Harrold and A.R. Nix. Capacity enhancement using intelligent relaying for future personal communication systems. In *Proc. IEEE VTC'00 Fall*, volume 5, pages 2115–2120, 2000.

[31] S. M. Heikkinen, H. Haas, and G. J. R. Povey. Investigation of adjacent channel interference in UTRA-TDD system. *UMTS Terminals and Software Radio, IEE Colloquium on*, (055):13/1–13/6, 1999.

[32] E. Hepsaydir. Mobile positioning in CDMA cellular networks. In *Proc. IEEE VTC'99 – Fall*, volume 2, pages 795–799, 1999.

[33] H. Holma, G. J. R. Povey, and A. Toskala. Evaluation of interference between uplink and downlink in UTRA/TDD. In *Proc. IEEE VTC'99 – Fall*, volume 5, pages 2616–2620, 1999.

[34] C. Y. Huang and R. D. Yates. Call admission in power controlled cdma systems. In *Proc. IEEE VTC 1996*, volume 3, pages 1665–1669, 1996.

[35] A. P. Hulbert, M. Purat, and S. Oestreich. Network synchronisation for UTRA TDD. In *Proc. IEEE VTC 2001 – Spring*, volume 4, pages 2804–2807, 2001.

[36] S. Hurley. Planning effective cellular mobile radio networks. *IEEE Transactions on Vehicular Technology*, 51(2):243–253, March 2002.

[37] O. Ileri, S.-C. Mau, and N. Mandayam. Pricing for enabling forwarding in self-configuring ad hoc networks. In *Proc. WCNC 2004*, volume 2, pages 1034–1039, 2004.

[38] Jr. J. Caffery and G. L. Stüber. Subscriber location in CDMA cellular networks. *IEEE Transactions on Vehicular Technology*, 47(2):406–416, May 1998.

[39] R. Jain and E. W. Knightly. A framework for design and evaluation of admission control algorithms in multi-service mobile networks. In *Proc. IEEE INFOCOM '99*, volume 3, pages 1027–1035, 1999.

[40] M. Jakobsson, J. P. Hubaux, and L. Buttyan. A micro-payment scheme encouraging collaboration in multi-hop cellular networks. In *Proceedings of Financial Crypto 2003*, January 2003.

[41] F. Kelly. Charging and rate control for elastic traffic. *European Transactions on Telecommunications*, 8(3):33–37, 1997.

[42] F. P. Kelly, A. Maulloo, and D. Tan. Rate control in communication networks: shadow prices, proportional fairness and stability. *Journal of the operational research society*, 49:237–252, 1998.

[43] C. Lindemann, M. Lohmann, and A. Thümmler. A unified approach for improving QoS and provider revenue in 3g mobile networks. *ACM Journal on Special Topics in Mobile Networks and Applications (MONET)*, 2003.

[44] M. Lindström. *Resource Allocation for Asymmetric Traffic in Time Division Duplexing Mode Cellular Networks*. Licentiate Thesis, TRITA-S3-RST-0309, KTH, Royal Institute of Technology, 2003.

[45] M. Lindström, L. Badia, J. Zander, and M. Zorzi. An economic model for the radio resource management in multi-media wireless systems. February 2005. submitted to the IEEE JSAC Special issue: Price-Based Access Control and Economics for Communication Networks.

[46] M. Lindström and P. Lungaro. Resource delegation and rewards to stimulate forwarding in multihop cellular networks. In *Proc. IEEE VTC 2005 – Spring*, 2005.

[47] M. Lindström. Improved TDD resource allocation through inter-mobile interference avoidance. In *Proc. IEEE VTC 2001 – Spring*, volume 2, pages 1027–1031, 2001.

[48] M. Lindström. Improved TDD resource allocation through link gain estimation. In *Proc. 3Gwireless'01*, pages 245–249, San Francisco, U.S.A., 2001.

[49] M. Lindström. Heterogeneous link asymmetry in TDD mode cellular systems. In *Proc. (electronic) World Wireless Congress '03*, San Francisco, U.S.A., May 2003.

[50] M. Lindström. Base station placement in asymmetric TDD mode systems in a manhattan environment. In *Proc. IEEE VTC 2004 – Spring*, volume 4, pages 1968–1972, 2004.

[51] M. Lindström and J. Zander. Dynamic link asymmetry in 'bunched' wireless networks. In *Proc. IEEE VTC 1999 – Fall*, volume 1, pages 352–356, 1999.

[52] X. Liu, E.K.P. Chong, and N.B. Shroff. Opportunistic transmission scheduling with resource-sharing constraints in wireless networks. *IEEE Journal on Selected Areas in Communications*, 19(10):2053–2064, October 2001.

[53] Z. Liu and M. El Zarki. Sir-based call admission control for ds-cdma cellular systems. *IEEE Journal on Selected Areas in Communications*, 12(4):638–644, May 1994.

[54] V.H. Mac Donald. Advanced mobile phone service: The cellular concept. *The Bell System Technical Journal*, 58(1):15–41, January 1979.

[55] L. R. Maciel and H.L. Bertoni. Cell shape for microcellular systems in residential and commercial environments. *IEEE Transactions on Vehicular Technology*, 43(2): 270–278, May 1994.

[56] T. T. Nagle and R. K. Holden. *The Strategy and Tactics of Pricing*. Pearson Education, Prentice Hall, Upper Saddle River, NJ 07458, U.S.A., 2002. ISBN 0-13-026248-X.

[57] S. Nanda and O.-C. Yue. Variable partition duplexing for wireless communications. In *Proc. IEEE GLOBECOM '90*, volume 2, pages 1147–1153, 1990.

[58] H. C. Papadopoulos and C-E. W. Sundberg. Shared time division duplexing (STDD) with fast speech activity detection. In *Conf. Record of IEEE ICC'96*, volume 3, pages 1745–1749, 1996.

[59] M. Pent, M. A. Spirito, and E. Turco. Method for positioning GSM mobile stations using absolute time delay measurements. *Electronics Letters*, 33(24):2019–2020, November 1997.

[60] R. S. Pindyck and D. L. Rubinfeld. *Microeconomics*. Prentice Hall International, Upper Saddle River, NJ 07458, U.S.A., fourth edition, 1998. ISBN 0-13-896184-0.

[61] G. J. R. Povey. Time division duplex-code division multiple access for mobile multimedia services. In *Proc. PIMRC'97*, volume 3, pages 1034–1037, 1997.

[62] G. J. R. Povey, H. Holma, and A. Toskala. Hybrid FDD/TDD-CDMA for third generation cellular systems. *CDMA Techniques and Applications for Third Generation Mobile Systems, IEE Colloquium on*, (129):2/1–2/6, 1997.

[63] G. J. R. Povey, H. Holma, and A. Toskala. TDD-CDMA extension to FDD-CDMA based third generation cellular system. In *Universal Personal Communications Record, 1997 IEEE 6th International Conference on*, volume 2, pages 813–817, 1997.

[64] R. Prasad, W. Mohr, and W. Konhäuser, editors. *Third Generation Mobile Communication Systems*. Artech House Publishers, Norwood, MA, U.S.A., 2000.

[65] RN6. Influences of the new TDD concept to rn6 work. Technical Report AC090/SM/RN6/PI/I/101/a1, FRAMES, June 1998.

[66] T. Rouse, S. McLaughlin, and H. Haas. Coverage-capacity analysis of opportunity driven multiple access (ODMA) in UTRA TDD. In *3G Mobile Communication Technologies*, pages 252 – 256, 2001.

[67] C.U. Saraydar, N.B. Mandayam, and D.J. Goodman. Efficient power control via pricing in wireless data networks. *IEEE Trans. Comm.*, 50(2):291–303, February 2002.

[68] S. Shenker, D. D. Clark, D. Estrin, and S. Herzog. Pricing in computer networks: Reshaping the research agenda. *ACM Computer Communication Review*, 26:19–43, 1996.

[69] H. D. Sherali, C. M. Pendyala, and T. S. Rappaport. Optimal location of transmitters for micro-cellular radio communication system design. *IEEE Journal on Selected Areas in Communications*, 14(4):662–673, May 1996.

[70] V. Siris. Resource control for elastic traffic in CDMA networks. In *Proceedings ACM MobiCom 2002*, pages 193–204, 2002.

[71] S.N. Subramanian and Tho Le-Ngoc. Traffic modeling in a multi-media environment. In *Canadian Conference on Electrical and Computer Engineering 1995*, volume 2, pages 838–841, 1995.

[72] C.W. Sung and W.S. Wong. A noncooperative power control game for multirate cdma data networks. *Transaction on Wireless Communications*, 2(1):186–194, January 2003.

[73] H. R. Varian. *Intermediate Microeconomics: A Modern Approach*. W. W. Norton & Company, New York, U.S.A., fifth edition, 1999. ISBN 0-393-97370-0.

[74] N. Verema and I.R. Chen. Admission control algorithms integrated with pricing for revenue optimization with qos guarantees in mobile wireless networks. In *Proc. International Conference on Parallell and Distributed Systems (ICPADS) 2004*, pages 495–502, 2004.

[75] J. von Neumann and O. Morgenstern. *The Theory of Games and Economic Behaviour*. Princeton University Press, Princeton, N.J., U.S.A., 1944.

[76] J. Werding. The wireless world – vision or utopia? CIC open seminar, November 2003. Center for Information and Communication Research (CIC) at the Stockholm School of Economics.

[77] W.-C. Wong, C.-E. W. Sundberg, and N. Seshadri. Shared time division duplexing: an approach to low-delay high-quality wireless digital speech communications. *IEEE transactions on vehicular technology*, 43(4):934–945, November 1994.

[78] M. P. Wylie and J. Holtzman. The non-line of sight problem in mobile location estimation. In *Proc. IEEE ICUPC 1996*, volume 2, pages 827–831, 1996.

[79] M. Xiao, N. B. Shroff, and E. K.-P. Chong. Utility-based power control in cellular wireless systems. In *Proc. INFOCOM 2001*, pages 412–421, 2001.

[80] H. Yaïche, R. Mazumdar, and C. Rosenberg. A game theoretic framework for bandwidth allocation and pricing in broadband networks. *IEEE/ACM Transactions on Networking*, 8(5):667–678, October 2000.

[81] D. Åkerberg. Properties of a TDMA pico cellular office communication system. In *Proc. IEEE GLOBECOM '88*, volume 3, pages 1343–1349, 1988.

Appendix

Appendix A

TDD Models and Performance Measures

In this chapter we develop the framework of the TDD mode system in which our resource allocation algorithms for asymmetric traffic will be evaluated. We describe the radio network with its basic resource management system, the cellular environment and the propagation models. The traffic model is then introduced and asymmetry defined in more detail before the performance measures are defined.

A.1 Radio Network

The access method of our test system is Time Division Multiple Access (TDMA) with Time Division Duplexing (TDD). Each TDMA carrier has a frame structure that consists of $N_{\mathrm{TS}} = 16$ time-slots, Figure A.1. We try to isolate the TDD characteristics and stress the system as much as possible by using only one carrier and a frequency reuse of one, i.e. all cells use the same single carrier. Hence, the test system is a pure TDMA/TDD system and we shall in the following use the terms Resource Unit (RU) and time-slot interchangeably. We assume that the system is frame synchronised so that frames start at the same time in all cells. This greatly simplifies the analysis since there is negligible fractional slot overlap to consider. Frame synchronism has been shown to be easily achievable [35]. Thus, it is also a reasonable assumption.

Resource Allocation

Our primary concern is here to assess the feasibility and usefulness of supporting asymmetric traffic loads, i.e. loads where the information volumes are different in up and down links respectively. For this purpose, we have chosen to adopt the locally centralised "Bunched" approach to radio resource management [12]. Each bunch is equipped with an intra-bunch resource manager, which facilitates resource allocation and deallocation, power control and channel measurements. In this work our attention is restricted to a single subnetwork and the resource allocation subfunction of the resource manager.

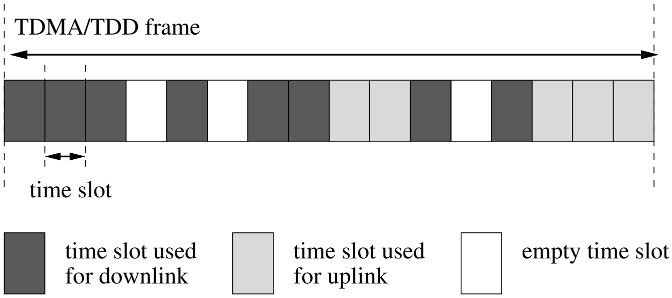


Figure A.1: 16-slot TDMA frame.

All gains between mobile stations and base stations are assumed to be known through intelligent measuring. Assuming that the transmitter powers of all transmitters are also known, received signal and interference powers can be calculated by the central unit (CU). Using this information, the resource allocator then attempts to locate a resource unit that, as far as the CU can tell, will provide the link under consideration with a sufficiently high SIR,

$$\Gamma_i = \frac{P_{\text{Rx},i}}{\sum\limits_{j \neq i} P_{\text{Rx},j} + \mathcal{N}} \quad , \tag{A.1}$$

where $P_{\text{Rx},i}$ is the received power from transmitter i and $\mathcal{N}$ is the receiver noise. Before the resource unit is actually assigned to a link, the CU also evaluates the effect that the assignment will have on already existing links. If the quality requirements of all links are seemingly satisfied, an assignment is judged feasible and is committed. If not, the resource allocator reprocesses the request for a resource unit until a feasible assignment is found or all free resource units have been evaluated, whichever comes first. Note that measurement imperfections may yield the feasibility check imperfect. Hence, our use of the word seemingly.

The following pseudo code describes the resource allocator and feasibility check:

1. Initialise:
   - Number of users that want to communicate: N
   - Set of requested links: $M = \{M_1, M_2, \ldots, M_N\}$
   - Number of considered users: $k = 0$
   - Set of admitted users: $U = \{\}$
2. Test if all users $\{U, M_{k+1}\}$ can seemingly be supported.
   a) Find a free RU according to a rule set R and temporarily assign it to user M_{k+1}.
   b) Under the current power control regime, calculate/estimate the received SIR, Γ, for all links. Interference from sources to which the path losses are known is accounted for. Interference subject to unknown path loss is neglected.

c) Compare SIRs to a target, γ_{tg}. If all SIRs are greater than this target, all users are considered to be supported: go to 3.

d) Find another free RU, according to R, temporarily assign it to user M_{k+1} and repeat from 2b. When no more RUs can be found, block user M_{k+1} and go to 4.

3. Add user M_{k+1} to the set of admitted users U.

4. Increase k by 1.

5. If $k \leq N$ then repeat from 2.

The rule sets, R, are described and evaluated along with a few reference rule sets in papers T1 and T2. In this appendix we merely intend to establish the properties of the system in which they have been evaluated. MSs are assumed to try to connect only to the BS to which it has the greatest link gain. Should that be unsuccessful, due to no admission or to a too low SIR, no attempts are made to connect to other BSs.

The link quality is assumed to be satisfactory when the SIR exceeds a threshold $\gamma_{\text{th}} = 7$ dB. With appropriate coding and modulation this should be enough to provide bit rates in the 1–2 Mbit/s range in a 4 MHz bandwidth with Bit Error Rates (BER) between 10^{-6} and 10^{-3}. These values are estimates based on results from link level simulations for UTRA/TDD [64] with spreading removed and a 3 dB margin added. Exact bit rates are not critical to this study, however. The feasibility check targets the SIR threshold with a 1 dB margin for round off errors etc. Thus, $\gamma_{\text{tg}} = 8$ dB. Transmission powers are controlled towards a Constant Received Power (CRP) target, P_{tg}, 20 dB above the noise level, $\mathcal{N}$. With a bandwidth of around 4 MHz, the power control target and the noise level become: $P_{\text{tg}} = -88$ dBm and $\mathcal{N} = -108$ dBm, respectively. The maximum transmission power is limited to 32 dBm

A.2 Scenario

The radio network and the RU acquisition schemes under investigation are studied in a Manhattan like environment. The early work of this study was performed within the European FRAMES project. Thus, for our purposes, we adopted the Pedestrian Deployment model of [24] with a few modifications.

The evaluation environment consists of ten by ten blocks. The blocks are $d_{\text{block}} = 200$ m wide and separated by $d_{\text{street}} = 30$ m wide streets. Base stations are regularly spaced every fourth block, Figure A.2. BSs are placed along the streets halfway between intersections, thus increasing the isolation between BSs compared to an 'in-intersections' deployment. With this configuration there are $N_C = 55$ BSs and cells.

Traffic and Mobility

In the literature, the number of active mobile stations in a cell is commonly modelled with a Poisson distribution. This is generally considered to be a good model for voice traffic.

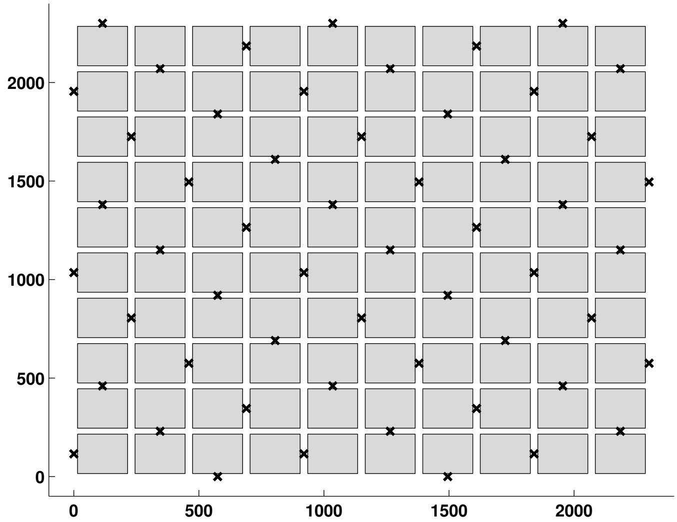


Figure A.2: Ten-by-ten-block Manhattan environment with base stations, crosses, halfway between intersections.

For data traffic, measurements have shown that actual data exchange is more accurately described by other models [3, 71, 19]. However, since our focus is on radio resource allocation and the implications of asymmetric traffic in cellular networks, not on traffic models nor protocols, we opt for the simple model following.

A fixed number of mobile stations, N, are uniformly distributed along the streets in the Manhattan like environment. Thus, the number of mobile stations in a cell is binomially distributed. Since all of the modelled users are active and the number of cells is relatively large, this binomial distribution approximates a Poisson distribution. We define the relative load of the system, $\bar{\omega}_c$, as the average number of users in the system divided by the number of cells and the number of channels available, i.e. users/channel/cell:

$$\bar{\omega}_c = \frac{N}{N_{\mathrm{RU}} N_{\mathrm{C}}}. \tag{A.2}$$

A measure of asymmetry can be defined in many ways. We chose a linear definition where asymmetry, A, is simply the fraction of the traffic that appears in the downlink.

$$A = \frac{\omega_{\mathrm{DL}}}{\omega}, \tag{A.3}$$

where ω_{DL} is the offered downlink traffic and ω is the total offered traffic. Hence, A will take values in the range $[0, 1]$, where 0 corresponds to only uplink traffic and 1 to pure downlink traffic. Further, asymmetry is modelled by assuming that terminals are active either in uplink or in downlink, not both, with probabilities p and $1 - p$ respectively.

This model should be reasonable for performance evaluations on a system level where the behaviour of individual users is not of major concern. It is not, however, believed to yield useful results on a per user basis. The model gives a high level of control over the offered traffic asymmetry. Thus, the impact of different asymmetries can easily be studied.

The Pedestrian Deployment model of [24] also describes mobility, but since the snapshot type of study performed here will hide the mobility aspects anyway, mobility is not implemented here.

A.3 Propagation Models

Propagation losses are modelled by means of a two-slope recursive path loss model [10] for street level propagation and the COST-Walfisch-Ikegami model [2] for over-rooftop propagation. For each link, whichever yields the highest gain is used.

Street Level Model

The street level path loss, in decibels, is given by [10, 24]:

$$L_{\text{street}} = 20\log_{10}\left[\frac{4\pi d_n}{\lambda} D\left(\sum_{j=1}^{n} s_{j-1}\right)\right], \tag{A.4}$$

where d_n is an 'illusory' distance between the transmitter and receiver while n is the number of street segments between the two. The segments are numbered from 0 to $n-1$ and have lengths s_0 to s_{n-1}. The illusory distance is basically a sum of these street segments with turns accounted for through magnification of distances. The illusory distance is obtained by recursive application of the expression:

$$\begin{cases} k_j = k_{j-1} + d_{j-1}q_{j-1} \\ d_j = k_j s_{j-1} + d_{j-1} \end{cases}, \tag{A.5}$$

with the initial values

$$k_0 = 1 \quad \text{and} \quad d_0 = 0. \tag{A.6}$$

q_j in Equation A.5 controls the 'turn' loss and is a function of the angle Θ_j between street segments s_j and s_{j-1}:

$$q_j\left(\Theta_j\right) = 0.5\frac{\|\Theta_j\|}{\pi/2}. \tag{A.7}$$

Street segments, angles and node indices are conceptually depicted in Figure A.3. We simplify the model by assuming that the angles between street segments are always $90°$; i.e., $\Theta_j = 90°, \forall j$.

$D(x)$, in Equation A.4, is a function adding a distance dependent two-slope behaviour according to:

$$D(x) = \begin{cases} \frac{x}{x_{\text{br}}} & \text{if } x > x_{\text{br}}, \\ 1 & \text{otherwise.} \end{cases}, \tag{A.8}$$

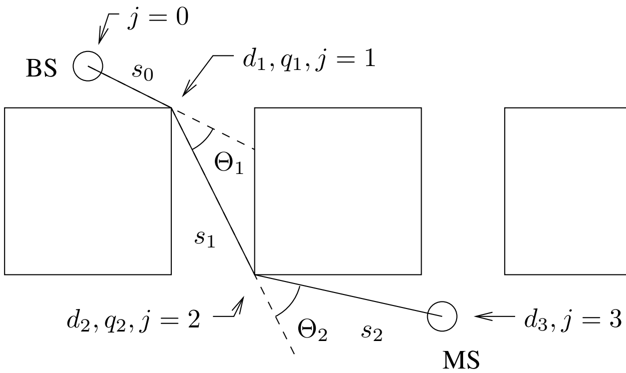


Figure A.3: Illustration of street segments s_j, turning angles Θ_j and 'turn' loss parameters q_j used in the recursive path loss model.

where x is the actual distance travelled from the transmitter to the receiver and x_{br} is the break point where attenuation characteristics changes. The incremental Line Of Sight (LOS) path loss increases from 20 dB per decade before the break point to 40 dB per decade after the break point. The break point x_{br} is set to 300 m.

Roof Top Model

For propagation over roof tops the COST-Walfisch-Ikegami model [2, 24] is used. For a frequency of 2000 MHz, an average building height of 16.5 m and BS and MS antennas below rooftops at 11.5 m and 1.5 m above the ground, respectively, it becomes

$$L_{\mathrm{roof}} = 24 + 45 \log_{10}(d + 20), \tag{A.9}$$

where d is the straight line horizontal distance between the transmitter and the receiver.

Eventually, for each link, the model that yields the lowest loss is selected. Thus, the basic path loss, L_b, becomes:

$$L_b = \min(L_{\mathrm{street}}, L_{\mathrm{roof}}). \tag{A.10}$$

Fading

In addition to the basic path loss, the signal also experiences large-scale and small-scale fading. Large-scale fading is mainly due to shadowing from large objects near or in the signal path and, hence, is also known as shadow fading. Small-scale fading is commonly referred to as multipath fading or fast fading and is caused by variations in path length differences in a multipath environment. Thus, large signal variations occur even for small displacements, in the order of half a wave length, of either of the communicating entities or any reflector in the path between them.

In this thesis, the large-scale fading is modelled with a log-normally distributed variable with a log-mean of 0 dB and a log-standard deviation of 8 dB. No spatial nor temporal correlation is modelled. Small-scale fading is not modelled, but can often be overcome with transmitter power control or interleaving and error control coding. Thus, the over all path loss modelled, becomes:

$$L = B \cdot L_b = B \cdot \min\left(L_{\text{street}}, L_{\text{roof}}\right), \tag{A.11}$$

where

$$10 \log_{10}(B) \in \mathrm{N}(0, 8). \tag{A.12}$$

In this work, it is assumed that, in the Manhattan like environment, the antenna heights of base stations and mobile stations have the same order of magnitude. Therefore we approximate inter-mobile and inter-base station link gains with the same models with antenna heights adjusted according to the type of station.

A.4 Performance Measures

The algorithms studied in this thesis are evaluated in terms of:

a) the user blocking rate ν_{blk} , i.e. the rate at which terminals are unable to find a free time-slot,

b) the user outage rate ν_{out}, i.e. the rate at which users that are assigned channels fail to reach the required SIR,

c) the Grade of Annoyance (GoA), a weighted user assignment failure rate,

$$\text{GoA} = \nu_{\text{blk}} + 10\left(1 - \nu_{\text{blk}}\right)\nu_{\text{out}} \tag{A.13}$$

[1] and

d) the system capacity, C.

An assignment failure occurs when a user either does not get a channel or gets a channel with too low a *Signal to Interference Ratio* (SIR). The rates above are studied as functions of the relative traffic load $\bar{\omega}_c$ and separately for uplinks and downlinks; i.e., there is an uplink blocking rate and a downlink blocking rate etc.. Occasionally we will also refer to a *composite* blocking rate, outage rate or GoA. In the composite version, no distinction is made between uplinks and downlinks; any link is just a link.

Since both uplink and downlink are usually required for a useful connection, the capacity, C, of the system is taken to be the load at which the Grade of Annoyance becomes 5 percent for *either* uplink or downlink, whichever occurs first. This is a stronger condition than requiring the composite GoA not to exceed 5 percent. The condition is loosened somewhat, however, by the fact that we don't require paired links; i.e., simultaneously working uplink from and downlink to the same MS, see Section A.2. Though capacity may be overestimated, the impact should be similar for all of the studied resource allocation algorithms.

[1] In paper T1 we used the GoA measure $\text{GoA} = \nu_{\text{blk}} + \nu_{\text{out}}$.

Table A.1: Summary of system parameters.

Parameter	**Symbol**	**Value**
Number of resource units	$N_{\text{RU}} = N_{\text{TS}}$	16
CRP target	P_{tg}	-88 dBm
Noise	$\mathcal{N}$	-108 dBm
Maximum transmit power	P_{max}	32 dBm
SIR threshold	γ_{th}	7 dB
SIR target for feasibility check	γ_{tg}	8 dB
System size		10 by 10 blocks
Block size	d_{block}	200 m
Street width	d_{street}	30 m
Number of BSs	N_C	55
Centre frequency		2000 MHz
Building height		16.5 m
BS antenna height		11.5 m
MS antenna height		1.5 m
LOS path loss:		
dual slope breakpoint	x_{br}	at 300m
path loss before breakpoint		20 dB per decade
path loss after breakpoint		40 dB per decade
Street corner turn angle	Θ_j	90°
Shadow fading	B	log-normal with 8 dB standard deviation
Fast fading		not modelled

A.5 Summary of System Parameters

The system parameters are summarised in Table A.1