UNIVERSITY OF VAASA
FACULTY OF INFORMATION TECHNOLOGY
DEPARTMENT OF COMPUTER SCIENCE

Mohammed Al-Rawi

SCHEDULING AND ADMISSION CONTROL IN OPPORTUNISTIC WIRELESSS SYSTEMS

Licentiate Thesis

VAASA
2008

UNIVERISTY OF VAASA
Faculty of Information Technology

Author:	Mohammed Al-Rawi
Name of Thesis:	Scheduling and Admission Control in Opportunistic Systems
Supervisor:	Prof. Riku Jäntti
Degree:	Licentiate of Technology
Department:	Computer Science
Subject:	Information Technology
Year:	2008

Abstract

The new millennium has been labeled as the century of the personal communications revolution or more specifically, the digital wireless communications revolution. The introduction of new multimedia services has created higher loads on available radio resources. These services can be presented in different levels of quality of service. Namely, the task of the radio resource manager is to provide these levels. Radio resources are scarce and need to be shared by many users. The sharing has to be carried out in an efficient way avoiding as much as possible any waste of resources.

The main contribution focus of this work is on radio resource management in opportunistic systems. In opportunistic communications dynamic rate and power allocation may be performed over the dimensions of time, frequency and space in a wireless system. In this work a number of these allocation schemes are proposed.

A downlink scheduler is introduced in this work that controls the activity of the users. The scheduler is a simple integral controller that controls the activity of users, increasing or decreasing it depending on the degree of proximity to a requested quality of service level. The scheduler is designed to be a best effort scheduler; that is, in the event the requested quality of service (QoS) cannot be attained, users are always guaranteed the basic QoS level provided by a proportional fair scheduler. In a proportional fair scheduler, the user with the best rate quality factor is selected.

The rate quality here is the instantaneous achievable rate divided by the average throughput

Uplink scheduling is more challenging than its downlink counterpart due to signalling restrictions and additional constraints on resource allocations. For instance, in long term evolution systems, single carrier FDMA is to be utilized which requires the frequency domain resource allocation to be done in such a way that a user could only be allocated subsequent bands. We suggest for the uplink a scheduler that follows a heuristic approach in its decision. The scheduler is mainly based on the gradient algorithm that maximizes the gradient of a certain utility. The utility could be a function of any QoS. In addition, an optimal uplink scheduler for the same system is presented. This optimal scheduler is valid in theory only, nevertheless, it provides a considerable benchmark for evaluation of performance for the heuristic scheduler as well as other algorithms of the same system.

A study is also made for the feedback information in a multi-carrier system. In a multi-carrier system, reporting the channel state information (CSI) of every subcarrier will result in huge overhead and consequent waste in bandwidth. In this work the subcarriers are grouped into subbands which are in turn grouped into blocks and a study is made to find the minimum amount of information for the adaptive modulation and coding (AMC) of the blocks.

The thesis also deals with admission control and proposes an opportunistic admission controller. The controller gradually integrates a new user requesting admission into the system. The system is probed to examine the effect of the new user on existing connections. The user is finally fully admitted if by the end of the probing, the quality of service (QoS) of existing connections did not drop below a certain threshold.

It is imperative to mention that the research work of this thesis is mainly focused on non-real time applications.

Preface

This thesis consists of research work that has been conducted during the first years of my doctoral studies at the University of Vaasa. The work mainly is concerned with radio resource management methods in wireless opportunistic networks and more specifically, cellular networks.

The work presented in this thesis is composed as a monograph based on published and submitted conference papers and journals.

The work has been done under the supervision of Professor Riku Jäntti to whom I owe a great deal of gratitude for his overwhelming support. His way of thinking and his systematic approach in research has been an inspiration to me. His door open to my many questions giving me the most efficient answers and suggestions. For all that I am truly grateful and thankful. I would also like to express my thanks to the members of the faculty of technology at the University of Vaasa for all the assistance they've provided me. Special thanks to Petri Ingström for all his help.

Abbreviations

3G	Third Generation
4G	Fourth Generation
ALP	Active Link Protection
AMC	Adaptive Modulation and Coding
ARQ	Automatic Retransmission request
BS	Base Station
CAC	Call Admission Control
CDFT	Channel Dependent Frequency and Time domain
CDMA	Code Division Multiple Access
CDT	Channel Dependent Time domain
CIR	Carrier to Interference Ratio
CN	Circular Normal
CSI	Channel State Information
D-FDMA	Distributed-Frequency Division Multiple Access
DS-CDMA	Direct Spread-Code Division Multiple Access
EVDO	Evolution-Voice and Data Optimized
FDMA	Frequency Division Multiple Access
FFT	Fast Fourier Transform
GIR	Gain to Interference Ratio
HARQ	Hybrid Automatic Retransmission reQuest
HDR	High Data Rate
HLGA	Heuristic Gradient Algorithm

HSDPA	High Speed Downlink Packet Access
IFFT	Inverse Fast Fourier Transform
IS-95	Interim Standard 95
L-FDMA	Localized- Frequency Division Multiple Access
LGA	Localized Gradient Algorithm
LOS	Line of Sight
LTE	Long Term Evolution
MAC	Medium Access Control
MC-CDMA	Multi-Carrier Code Division Multiple Access
NAC	Negative Acknowledgment
NLGA	Non Localized Gradient Algorithm
OFCDM	Orthogonal Frequency and Code Division Multiplexing
OFDM	Orthogonal Frequency Division Multiplexing
OFDMA	Orthogonal Frequency Division Multiple Access
PAPR	Peak to Average Power Ratio
PDF	Probability Density Function
PF	Proportional Fair
QCA	Quality Control Algorithm
QoS	Quality of Service
RB	Resource Block
RCA	Rate Control Algorithm
RF	Radio Frequency
RLS	Recursive Least Square
RNC	Radio Network Controller
RR	Round Robin
RRM	Radio Resource Management
SC-FDMA	Single Carrier-Frequency Division Multiple Access
SIDR	Signal-to-Interference Density Ratio
SINR	Signal-to-Interference+Noise Ratio

SNR	Signal to Noise Ratio
TDM	Time Division Multiplexing
TDMA	Time Division Multiple Access
TTI	Transmission Time Interval
UE	User Equipment
UMTS	Universal Mobile Telecommunications System
WCDMA	Wideband Code Division Multiple Access

Notations

Models and Assumptions

α	Attenuation factor
β_i	Gain in the ith multi path
γ_{ij}	SINR (SINR from transmitter j at receiver i)
ϕ_m	Phase of the incoming wave.
λ	Wavelength
θ	Orthogonality factor
φ_m	Angle between the incoming wave and mobile direction.
τ	Time delay
ξ	Shadowing factor
ω_D	Maximum Doppler angular frequency
a	Amplitude
B	Bandwidth
d_{ij}	Distance between i and j
c	Rayleigh distributed variable
C	Capacity
E	Electric field
G_{ij}	Link gain between i and j
f	Frequency
f_m	Doppler spread
h	Channel transfer function
K	Rician factor

N_i	Noise power at i
P_i	Transmission power of i
t	Time
x	Rayleigh fading component
y	LOS component

Downlink Scheduling

β	Positive integration gain
γ	Fixed positive value
η	CIR
∇	Gradient operator
π_k	Probability that the set k was used at a particular time
Ψ_{VV}	Covariance matrix for the state noise
Ψ_{ee}	Covariance matrix for the measurement noise
Ψ_{Ve}	Cross-Covariance matrix for the state and measurement noises
μ	Data rate
$\mathcal{A}$	Set of active users
B	Gain vector
C	Vector that maps the states to the measurement
e	Filter measurement noise
k	Set of users
K	Filter gain
$\mathcal{K}_i$	Selection of active sets that contain user i
$\mathcal{N}$	Set of admitted users
P	Correlation matrix
q	Activity probability
Q	Activity probability vector
$\bar{x}_i$	Mean throughput of user i
s_i	QoS of user i

$\bar{s}_i$	Mean QoS of user i
s_i^{req}	Requested QoS for user i
V	Filter state noise
X	State vector for the Kalman filter

Uplink Scheduling

η	CIR
$\gamma_{i,n}^{eff}$	Effective SNR of user i on RB n
μ	Data rate
∇	Gradient operator
θ	Angle of antenna beam
τ	Fixed predefined time
$\mathcal{A}$	Set of active users
C	Total number of RBs
$C_{i,n}$	Capacity of RB n for user i
F_i	Set of RBs that could be allocated to user i
$G_{i,k}$	Path gain for subcarrier k of user i
J_i	Set of RBs assigned to user i
k	Set of users
K	Set of all feasible RB assignments
$\mathcal{K}_i$	Selection of active sets that contain user i
L_i	Number of RBs allocated to user i
M	Number of carriers in one RB
N	Number of users
$\mathcal{N}$	Set of admitted users
$P_{i,k}$	Transmision power allocated to subcarrier k of user i
P_{max}	Maximum transmission power
P_n	Noise power
μ_i	Service rate of user i

S	Antenna gain
$U(\bar{X})$	Utility function of $\bar{X}$
$\bar{x}_i$	Average throughput of user i
$\bar{X}$	Vector of average throughputs
y	Selection variable

Feedback

Δf_c	Coherence bandwidth
γ	Mean SINR
μ	Service rate
ξ	Fast fading element
B	Number of available feedback bits
C_i	Capacity of subband i
D	Retransmission Delay
$E_1(x) = \int_1^\infty t^{-1} e^{-xt}$	Exponential integral
f_c	Carrier frequency
F	Cumulative distribution function
h	Channel transfer function
k	Time-slot
L	Total number of blocks of subbands
M	Total number of subcarriers
N	Total number of subbands
Q	Total number of quantization levels
T	Throughput
W	Total bandwidth
Z	Channel state information available to the scheduler

Admission Control

α	Fixed (small) parameter
ϵ	Estimation error
$\chi\{L\}$	Indicator function of an event L
λ	Forgetting factor
Ψ_{VV}	Covariance matrix for the state noise
Ψ_{ee}	Covariance matrix for the measurement noise
Ψ_{Ve}	Cross-Covariance matrix for the state and measurement noises
$\mathcal{A}$	Set of active users
C	Vector that maps the states to the measurement
e	Filter measurement noise
E_b/I_0	Bit energy to interference ratio
k	Time-slot
K	Filter gain
p	Back-off probability
P	Correlation matrix
μ_i	Service rate of user i
$\bar{x}_i$	Mean throughput
$\tilde{x}_i$	Expected average throughput
$\bar{x}_i(Z_N)$	Mean throughput of user i with N users in the system
V	Filter state noise
w	Filter coefficients
Y	State vector for the Kalman filter
Z_N	Channel statistics matrix with N users in the system

List of Tables

Contents

Chapter 1

Introduction

Radio resource management is a key part of wireless cellular networks. With the introduction of new generations in cellular technologies, the demand for efficient resource management schemes has increased. The growth of multimedia services has increased the complexity of balancing the operator-customer equation. The operator in this equation wants to maximize revenue by utilizing its limited resources to the fullest to accommodate as much users as possible. The problem of radio resource management (RRM) is to allocate bandwidth, transmitter power and transmitter time to the users so that certain quality of service (QoS) targets are met while the system resource utilization is maximized. In a single channel, the perceived QoS is proportional to the received signal-to-interference+noise ratio (SINR). The higher the SINR the higher order modulation and lower coding rate that can be utilized for a fixed frame error level. In order to maximize the subchannel SINR, signal strength should be maximized while the interference part minimized. In case of non-real time data, this could be achieved by scheduling the data packet transmissions. For real-time data, the SINR could be efficiently controlled by controlling the transmitter power. As the number of users increases, the risk of quality of service degradation increases. Users expect to receive the same quality of service if not better without paying more. The operator will have to provide a balance through efficient radio resource management. The operator needs to guarantee reasonable QoS levels in terms of probabilities of call blocking (a user being denied a new connection), call dropping (a user losing an ongoing connection), maximum packet delay, delay jitter and packet dropping. The operator understands that failure to deliver these guarantees results in client dissatisfaction and consequently changing to another operator.

Radio resource management is a vast field that attracts a great deal of research. In this thesis, the objective is to analyze and develop new radio resource management algorithms for opportunistic systems where RRM functionalities exploit the channel variations of users. Scheduling is one of the key RRM functionalities in opportunistic systems. Different favorable outcomes can be obtained with opportunistic scheduling such as increasing the system's capacity or providing some degree of fairness in resource allocation to different users in a system or providing certain QoS levels. In order to provide QoS guarantees, the scheduling entity must be combined with an admission control scheme. Using opportunistic admission control provides some flexibility in granting resources to new users seeking admittance to the system unlike traditional admission control algorithms that are more strict in their decisions. Due to the significance of these two functionalities in opportunistic systems, the author considered them in his research work manifested in this thesis.

The main focus in this thesis is on non-real time traffic applications in two standards which are; high speed downlink packet access (HSDPA) and long term evolution (LTE). The wideband code division multiple access (WCDMA) system's adaptability enabled a new and significant evolutionary steps in packet data access. The High Speed Packet Access Solution (Parkvall et al. 2001) is one of those steps which comprises of: HSDPA and high speed uplink packet access (HSUPA). HSDPA is a key feature included in Release 5 specifications. Features of HSDPA include:

- Adaptive modulation and coding.
- A fast scheduling function, which is controlled in the base station (BS), rather than by the radio network controller (RNC).
- Fast retransmissions with soft combining and incremental redundancy.

- Peak data rates of HSDPA reach up to 10 Mbps.

HSUPA is a release 6 feature in 3GPP specifications. The main aim of HSUPA is to increase the uplink data transfer speed in the UMTS environment and it offers data speeds of up to 5.8 Mbps. HSUPA achieves its high performance through more efficient uplink scheduling in the base station and faster retransmission control. HSUPA is expected to use an uplink enhanced dedicated channel (E-DCH) on which it will employ link adaptation methods similar to those employed by HSDPA, namely:

- Shorter Transmission Time Interval enabling faster link adaptation;
- HARQ (hybrid ARQ) with incremental redundancy making retransmissions more effective.

LTE is another evolutionary step of 3G that is paving the way toward 4G mobile communications technology. LTE will stretch the performance of 3G technology, thereby meeting user expectations in a 10-year perspective and beyond. The fundamental targets of this evolution - to further reduce user and operator costs and to improve service provisioning - will be met through improved coverage and system capacity as well as increased data rates and reduced latency (Ekström et al. 2006). Targets of LTE include:

- Download rates of 100Mbps, and upload rates of 50Mbps for every 20MHz of spectrum.
- At least 200 active users in every 5MHz cell. (ie 200 active phone calls).
- Sub-5ms latency for small IP packets.
- Increased spectrum flexibility, with spectrum slices as small as 1.25MHz (and as large as 20MHz) supported (W-CDMA requires 5MHz slices, leading to some problems with roll-outs of the technology in countries where 5MHz is a commonly allocated amount of spectrum, and is frequently already in use with

legacy standards such as 2G GSM and cdmaOne.) Limiting sizes to 5MHz also limited the amount of bandwidth per handset

- Optimal cell size of 5km, 30km sizes with reasonable performance, and up to 100km cell sizes supported with acceptable performance
- Co-existence with legacy standards (users can transparently start a call or transfer of data in an area using an LTE standard, and, should coverage be unavailable, continue the operation without any action on their part using GSM/GPRS or W-CDMA-based UMTS)
- A large amount of the work is aimed at simplifying the architecture of the system, as it transitions from the existing UMTS circuit + packet switching combined network, to an all-IP system.

1.1 Scheduling

In a wireless system, resources such as power, time and frequency are scarce commodities that need to be divided wisely among users. The scheduler could be based on different purposes such as providing fairness in resource allocation or maximizing some utility function. Schedulers could be divided into two categories; channel independent and channel dependent (or opportunistic). Channel independent scheduling is also called blind scheduling due to the fact that it does not need any information about channel conditions to perform scheduling. An example of a blind scheduler is the Round-Robin (RR) scheduler. Round-robin is one of the simplest scheduling algorithms for processes in an operating system, which assigns time slices to each process in equal portions and in order, handling all processes without priority. Round-robin scheduling is both simple and easy to implement, and starvation-free. In channel dependent schedulers, the scheduler forms the decision based on feedback information from the terminal (for downlink) and from the BS (for uplink). The advantage of channel dependent scheduling is the exploitation of channel fluctuations. That is, by assigning resources to a user who benefits the most from using them.

Several different scheduling rules have been introduced in the literature. The max-CIR rule always selects the user having the highest CIR (Tse and Hanly 1998). This rule maximizes system throughput, but leads to very unfair division of resources as only users close to the base station have the chance to transmit. A very good trade-off between fairness and throughput can be obtained with the proportional fair (PF) scheduler (Tse 1997, Jalali et al. 2000), which utilizes the instantaneously achievable service rate divided by the average throughput as a decision variable. Such a scheduling rule leads to resource fairness: All users asymptotically get equal access to the channel. Their throughput, however, depends on their positions. The gain of such multi-user diversity scheduling was found to be equal to the gain of selection diversity of a multipath channel (Berggren and Jäntti 2004). Such a gain can also be exploited in the multi-channel case although the problem becomes more complex since the SINR of the channels will also depend on the power allocation as well. If there are strict QoS constraints, they need to be enforced in some manner. Many modifications of the original PF rule have been suggested to control the quality of service level perceived by users. See e.g. (Shakkottai and Stolyar 2000), (Shakkottai and Stolyar 2001) and (Chang and Han 2002). The gradient algorithm (Stolyar 2005) is a natural generalization of the PF algorithm in that it applies to any concave utility function and to systems where multiple users can be served at a time. The gradient algorithm chooses a (possibly nonunique) decision that maximizes the scalar product of the gradient of a concave utility function with a certain service rate vector. Other fairness principles include for example; the min-max fairness scheduler (Bonald and Proutiere 2003) The notion of min-max fairness can be defined in the following way: no flow can increase its allocation without reducing the allocation of another flow with less or equal demand. Under min-max fairness, given no additional resources, an unsatisfied flow cannot increase its allocation by merely demanding more. The Max-CIR, PF and min-max rules are related to each other by the fact that they can be derived from the gradient rule as shown below.

$$i^*(t) = \arg_i \max \nabla U(\bar{x}_i(t))\mu_i(t) \tag{1.1}$$

Where ∇ is the gradient operator ($\nabla U(\bar{x}_i(t)) = \partial U/\partial \bar{x}_i(t)$). The variable U denotes a required utility that is a function of $\bar{x}_i(t)$ which in turn is the mean throughput of user i at time t. The variable $\mu_i(t)$ denotes the instantaneous service rate user i would obtain if selected at time t. The mean throughput can be computed as follows.

$$\bar{x}_i(t) = E\left\{\mu_i(t)\chi\{i = i^*\}(t)\right\} \tag{1.2}$$

Where $E\{.\}$ is the expectation operator. The operator $\chi\{L\}$ is an indicator function of an event L: $\chi\{L\} = 1$ if the event L occurs and zero otherwise.

If $log(\bar{x}_i(t))$ is used as the utility function then (1.1) will become.

$$i^*(t) = \arg_i \max\left\{\left(\frac{1}{\bar{x}_i(t)}\right)\mu_i(t)\right\} \tag{1.3}$$

Different powers of the gradient will result in the different scheduling rules.

$$i^*(t) = \arg_i \max\left\{\left(\frac{1}{\bar{x}_i(t)}\right)^\alpha \mu_i(t)\right\} \tag{1.4}$$

When $\alpha = 0$ we will have the Max-CIR rule, $\alpha = 1$ gives the PF rule and $\alpha = \infty$ will give the min-max rule.

Mainly schedulers are divided into two categories:

1. **Memoryless schedulers**: In these schedulers, the current scheduling decision is independent of past scheduling decisions. One example of a memoryless scheduler is the Max-CIR scheduler. In general channel-aware memoryless schedulers can be written in the form

$$i^*(t) = \arg\max_i f_i(\mu_i(t)) \tag{1.5}$$

   where $f_i(.)$ is a function independent of time t.

2. **Schedulers with memory**: A main limitation with memoryless schedulers is that fairness can only be ensured over long time windows compared to the coherence time of the fading. In order to control delay and ensure fairness

over smaller time horizons, memory has to be introduced in the scheduler. By introducing memory, the priority of users that have not been served for a long time can be raised. An example of a scheduler with memory is the PF scheduler.

A challenging task in scheduling is to meet the QoS demands of multiple users while maintaining high system throughput. In wireless communication systems where a common medium is shared, a good scheduling policy should provide a satisfactory tradeoff between (i) maximizing capacity, (ii) achieving fairness, and (iii) satisfying rate or delay constraints of users.

1.2 Admission control

The objective of call admission control (CAC) is to provide QoS guarantees for individual connections while efficiently utilizing network resources. Specifically, a CAC algorithm makes the following decision: Given a call arriving to a network, can it be admitted by the network, with its requested QoS satisfied and without violating the QoS guarantees made to the existing connections? The decision is based on the availability of network resources as well as traffic specifications and QoS requirements of the users. If the decision is positive, necessary network resources need to be reserved to support the QoS. Hence, CAC is closely related to channel allocation, base station assignment, scheduling, power control, and bandwidth reservation. Therefore, before a user can be admitted, the admission controller estimates the impact of admitting that user on the QoS of the existing connections since there will be an additional link competing for resources (Wu 2005). The new user is rejected if it was found that it will jeopardize the QoS of the current users, otherwise accepted. From a psychological point of view, it is easier for a user to be denied admittance than rather being admitted and later dropped during a call.

The admission control functionality is located in the radio network controller (RNC). CAC algorithms may differ in their admission criteria; they may be centralized or distributed; they may use global (all-cell) or local (single-cell) information about resource availability and interference levels to make admission decisions. The design of distributed CAC for cellular networks is not an easy task since intra-cell and inter-cell interference should be taken into account. The associated intra-cell and inter-cell resource allocation will therefore be complicated due to the interference. A typical admission criterion is SIR. For example, Liu and Zarki (1994) employ SIR to define a measure called residual capacity, and use it as the admission criterion: if the residual capacity is positive, accept the new call; otherwise, reject it. Evans and Everitt (1999) use the concept of effective bandwidth to measure whether the signal to interference density ratio (SIDR) can be satisfied for each class with certain probability. If the total effective bandwidth including that for the new call, is less than the available bandwidth, the new call will be accepted; otherwise, it will be rejected. Traditional admission control algorithms make acceptance decisions for new and handoff calls to satisfy certain QoS constraints such as the dropping probability of handoff calls and the blocking probability of new calls being lower than a pre-specified threshold. A base station may support only a limited number of connections (channel assigned) simultaneously due to bandwidth limitations. Handoff occurs when a mobile user with an ongoing connection leaves the current cell and enters into another cell. Thus an ongoing, incoming connection may be dropped during a handoff if there is insufficient bandwidth in the new cell to support it. We can reduce the handoff call drop probability by rejecting new connection requests. Reducing the handoff call drop probability could result in an increase in the new call blocking probability. As a result, there is a tradeoff between the handoff and new call blocking probabilities. These control algorithms usually enforce hard admission decisions. Opportunistic admission control algorithms on the other hand provide softer decisions due to the property of adapting to the variation of the channel condition of the users, permitting more flexibility in the admission decision. There has been little study on admission control in opportunistic multi-user communications.

1.3 Objectives

The objective of this work is to analyze and develop new radio resource management algorithms for opportunistic wireless systems. The algorithms should be distributed and lend themselves for practical implementation. The performance of the proposed algorithms are evaluated theoretically and by simulation. The main goal is to find algorithms that can provide or guarantee quality of services for users in cellular systems and the mechanisms to maintain the quality of these services.

1.4 Contributions

The results from this thesis shed more light on radio resource management in opportunistic systems which are an integral part in beyond third generation wireless communication systems. The contributions of this work are included in Chapters 4-7 of this thesis. The chapters were constructed based on submitted and published conference and journal papers. The following is an overview of the contributions.

Chapter 4

M. Al-Rawi and R. Jäntti, "Opportunistic Best-effort Scheduling for QoS-aware Flows," in *in Proc. 17th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '06)*, pp. 1-5, 2006.

A best-effort scheduler is suggested for High Speed Downlink Packet Access (HSPDA) systems that schedules resources in a way that achieves certain target QoS levels requested by users. Scheduling is based on adjusting the activity of a user such that the farther it is from the target the higher its activity becomes and vice versa. The objective of the algorithm will be to find the activity probabilities that achieve and maintain that target QoS level. The scheduler is based on a best-effort paradigm in a sense that when a requested target cannot be reached for instance in the case of congestion or limited resources in the system, then the

algorithm would fall back to some fair scheduler allocating the resources in a fair manner.
In this work the author provided the theoretical analysis for the convergence of the algorithm. The author also provided the numerical analysis and results.

Chapter 5

M. Al-Rawi, R. Jäntti, J. Torsner and M. Sågfors, "Opportunistic Uplink Scheduling for 3G LTE Systems," *in Proc. 4th IEEE Innovations in Information Technology (Innovations07)*, 2007.

A heuristic scheduler is suggested that is consistent with the resource allocation constraints of the uplink channel for 3G long term evolution (LTE) systems. Additionally, the scheduler will associate retransmissions with the scheduling decision. A heuristic scheduler is a suitable choice since it is computationally feasible and is able to find a practical solution to the resource allocation problem. An integer programming problem that provides a theoretical optimal solution is also formulated. The performance of the heuristic scheduler is compared with this solution. Results suggest that the heuristic scheduler performs well when compared to the optimal solution.
In this work the author proposed and introduced the HLGA algorithm in addition to providing all the simulations.

Chapter 6

R. Jäntti and M. Al-Rawi, "On the block-wise feedback of channel adaptive multicarrier systems," *in Proc. 65th IEEE Vehicular Technology Conference, (VTC2007-Spring)* , pp. 2946 - 2950 , 2007.

Channel state information is an important part of the adaptation process in orthogonal frequency division modulation (OFDM) where the modulation and coding schemes are adapted according to the channel state of each subcarrier. However, reporting the CSI for each subcarrier can result in considerable overhead. Besides

coding and modulation, another scheme that affects the overall throughput of the system is the utilized automatic repeat request (ARQ) process. If hybrid ARQ (HARQ), such as chase combining, is utilized, all the transmitted bit energy can be harnessed by the receiver by combining the erroneously received code block with the consecutive copies transmitted by the ARQ process. The feedback information can be reduced by grouping the subcarriers into subbands with one feedback per subband. The subbands can further be grouped into one or several blocks with a joint CSI feedback per block. The subbands still have their own ARQ process to cope with the possible mismatch between the joint CSI and the actual subbands state. However, this raised the question of what is the best basis on which the joint CSI should be determined since there are subbands with different channel states in a block. In this work the impact of different decision variables for the feedback information is studied. The use of rank ordering to find the subband state that maximizes the total throughput is proposed. The results suggest that the use of the maximum subband CSI as the decision variable is desirable in case of fast mobility and low SINR while the minimum value subband provides better performance in a high SINR region. However, in practical ranges of SINR and mobility speed, some value in between should be chosen. It is also shown that using block feedback achieves higher throughput than individual subband feedback when the number of feedback bits is small.
The author's contribution in this paper was to provide the necessary simulations for the work.

Chapter 7

R. Jäntti and M. Al-Rawi, "An admission Control Scheme for Opportunistic Scheduling" *in Proc. 2nd IEEE International Symposium on Wireless Communication Systems (ISWCS2005)* pp. 298-302, 2005. [1]

A new Call Admission Control (CAC) algorithm is suggested for High Speed Downlink Packet Access UMTS systems. In this algorithm the user requesting

[1] Extended into a journal paper that is currently under review.

admission is gradually admitted to the system. This is done by setting a back-off factor limiting the new user's throughput. The back-off factor is gradually decreased resulting in an increase in the new user's throughput until it is fully integrated into the system. The factor works on protecting the already existing links in a way that their rates do not drop below a guaranteed Quality of Service (QoS) level.
In this work the author constructed the admission control simulator providing the numerical results. The author revised the equations in the theory as well as setting up the RLS admission controller to compare with the proposed algorithm.

1.5 Organization of the Thesis

The thesis focuses on two systems in particular, HSDPA and LTE systems. Chapters 4 and 7 cover HSDPA systems while 5 and 6 cover the LTE part. In Chapter 3 we introduce the models and assumptions used throughout the thesis. We mainly describe in this chapter the channel model utilized in the simulator. Chapters 4 and 5 contain the scheduling part of this thesis. First we propose a downlink scheduler. Later, we introduce an uplink scheduler. Chapter 6 focuses on feedback information needed to support adaptive modulation and coding in opportunistic systems. In chapter 7 we study the admission control part. Finally, in chapter 8 the conclusion of this work is drawn and ideas for future work are suggested.

Chapter 2

Opportunistic Systems

In a multi-user environment it is highly probable that at least one link has high quality at any given point in time. Taking advantage of this opportunity leads to what is often called multiuser diversity. The notion of multiuser diversity is from (Knopp and Humblet 1995), where they mentioned that the best strategy is to always transmit to the user with the best channel for the uplink. Tse (1997) provided a similar result for the downlink where he analyzed the problem of communication over a set of parallel Gaussian broadcast channels, each with a different set of noise powers for the users. He showed that capacity can be achieved by optimal power allocation over the channels, and obtained an explicit characterization of the optimal power allocations and the resulting capacity region. Bender *et al.* (2000) examined practical aspects of downlink multi-user diversity in the context of the IS-95 CDMA standard. Viswanath, Tse and Laroia (2002) examined this problem for the downlink and presented a method of opportunistic beamforming via phase randomization.

In a multi-carrier system, channel variations for different frequency bands could also be exploited. In a single-user OFDM system, the transmit power for each subcarrier can be adapted to maximize data rate using the water-filling algorithm (Willink and Wittke 1997). In multiuser environments, the situation becomes more complicated, as each user will likely have a different multipath fading profile, due to the users not being in the same location. Thus, it is likely that while one subcarrier may be in deep fade for a particular user, it may be in a good condition for another user due to temporal and spatial diversity in user locations. Thus, this effect can be exploited to further enhance system performance. By dynamically allocating different subcarriers and transmit power to users, this scheme can enhance system performance beyond a fixed-power, fixed-subcarrier scheme. There are a number of papers that discuss waterfilling in a multiuser environment such as (Münz, Pfletschinger and Speidel 2002) which presents an algorithm that determines the subcarrier allocation for a multiple access OFDM system. In their algorithm, once the subcarrier allocation is established, the bit and power allocation for each user is determined with a single-user bit loading algorithm. Kobayashi and Caire (2007) proposed a an iterative waterfilling algorithm based on dual composition. Two decompositions are considered, one in subcarrier domain and another in both subcarrier and user domain.

2.1 Scheduling

To implement the idea of opportunistic scheduling, two issues need to be addressed: fairness and users' service requirements. In reality, channel statistics of different users are not identical and, therefore, a scheme designed only to maximize the overall throughput could be very biased, especially where there are users with widely disparate distances from the base station. For example, allowing only users close to the base station to transmit may result in very high throughput, but sacrifice the transmission of other users. Also, a scheduling strategy should not be concerned only with maximizing long-term average throughputs because, in practice, applications may have different utilities and service constraints. For instance, for real-time applications, the major concern is latency: if the channel variations are too slow, a user may have to wait for a long time before it gets the chance to transmit. When designing a scheduling algorithm, the challenge is to address these issues while at the same time exploiting the multi-user diversity gain inherent in a system. Improving the efficiency of spectrum utilization is important, especially to provide high-rate-data services. However, the potential to exploit higher data throughputs in an opportunistic way, introduces the tradeoff problem between wireless resource efficiency and levels of satisfaction among users. The cellular system itself also has to satisfy certain requirements in order to extract the multi-user diversity benefits. The base station has to have access to channel quality measurements: in the downlink, each receiver needs to track its own channel SINR and feed back this information to the base station. The base station has to be able to schedule transmissions among the users on a short timescale as well as to adapt users' data rates to the instantaneous channel quality. These features are already present in the designs of many 3.5G high data-rate (HDR) systems. This is the reason why opportunistic scheduling has received lots of attention recently (Vukadinovic and Drogou 2006).

2.1.1 Examples of Opportunistic Schedulers

We say that a user is *active* if it at time instant t contained data in its transmission buffer $b_i(t)$. That is, if $b_i(t) > 0$; otherwise a user is said to be *idle*. Let $\mathcal{A}(t)$ denote the set of active users at time t. In this section, we will present an example of an unfair opportunistic scheduler such as max-CIR and another of a fair scheduler such as PF, giving a simple insight of the mechanism for opportunistic scheduling.

1. Max-CIR scheduler: Selects the user $i^*(t)$ with the best channel state at time t.

$$i^*(t) = \operatorname{argmax}\left\{\xi_i(t), \quad i \in \mathcal{A}(t)\right\} \tag{2.1}$$

   where $\xi_i(t)$ denotes the channel condition of user i at time t.

2. PF scheduler: Selects the user with the highest service rate divided by the average throughput the user obtained up to the time of scheduling.

$$i^*(t) = \operatorname{argmax}\left\{\frac{\mu_i(t)}{\bar{x}_i(t)}, \quad i \in \mathcal{A}(t)\right\} \tag{2.2}$$

   where $\mu_i(t)$ is the instantaneous service rate of user i at time t if it would've been selected to transmit. $\bar{x}_i(t)$ denotes the average throuput.

2.1.2 Utility based scheduling

Many opportunistic scheduling algorithms can be viewed as "gradient-based" algorithms, which select the transmission rate vector that maximizes the projection onto the gradient of the system's total utility (Agrawal and Subramanian 2002). The utility is a function of each user's throughput and is used to quantify fairness and other QoS considerations. Several such gradient-based policies have been studied for TDM systems, such as the the "proportional fair rule" Jalali *et al.* (2000), Viswanath, Tse, Laroia (2002) and Kushner and Whiting (2002), first proposed for CDMA 1xEVDO, which is based on a logarithmic utility function. In (Agrawal *et al.*, 2001), a larger class of utility functions is considered that allows efficiency and fairness to be traded-off. Generalized cμ-policies (Van Mieghem 1995), (Mandelbaum and Stoylar 2002) and (Liu, *et al.* 2006), such as a Max Weight policy (Stolyar 2001) and (Tassiulas and Ephremides 1993) can also be viewed as a type of gradient-based policy, where the utility is a function of a user's queue-size or delay.

The gradient algorithm can be applied to any concave utility function $U(\bar{X})$ and to systems where multiple users can be served at a time. Stolyar (2005) has proved the asymptotical optimality of this algorithm for multiuser throughput allocation. Users $n = 1, \cdots, N$ are served by a switch in discrete time $t = 0, 1, 2, \cdots$ Switch state $m = (m(t), t = 0, 1, 2, \cdots)$ is a random ergodic process. In each state m, the switch can choose a scheduling decision k from a set $K(m)$. Each decision k has the associated service rate vector $\mu^{(m)}(k) = (\mu_1^{(m)}(k), \cdots, \mu_N^{(m)}(k))$. This vector represents the service rate at a specified "time-slot" if decision k is chosen. The gradient algorithm is defined as follows. If at time t the switch is in state m, the algorithm chooses a (possibly non-unique) decision

$$k(t) \in \arg \max_{k \in K(m)} \nabla U(\bar{X}(t))^T \mu^m(k) \tag{2.3}$$

Where U is a strictly concave smooth utility function. $\bar{X}(t)$ is a vector representing exponentially smoothed average service rates $\bar{x}_i$. Typically the utility function has the aggregate form $U(\bar{X}) = \sum_i u_i(\bar{x}_i)$. In (Stolyar 2005), it has been shown that

(5.4) converges to the optimal solution of $\max_{\bar{X}} U(\bar{X})$ as $t \to \infty$.

Andrews *et al.* (2005) consider a concave utility maximization problem with minimum and maximum rate ($\bar{X}_i^{min}$ and $\bar{X}_i^{max}$) constraints. They propose a solution to the problem, i.e. scheduling algorithm by modifying the token counter. In the paper, two specific forms of the scheduling algorithm are shown to guarantee $\bar{X}_i^{min}$ and $\bar{X}_i^{max}$.

Patil and Veciana (2007) propose a scheduling scheme that combines a policy to decide which users will be active with a mechanism to select the user to serve during a time-slot. Activity control is done in a heuristic manner. Users are divided into two categories: Real-time users and Best Effort users. Each real-time user is assigned k tokens (slots) within a frame. If a user has used up all its tokens, it will be removed from the real-time users active set.

Long and Feng (2007) present a rate-guaranteed opportunistic scheduling scheme. The authors consider the transmission rate (throughput) as the fairness criteria, i.e. on the average the expected throughput of user i should be a fraction $\frac{\phi_i}{\sum_j \phi_j}$ of the whole system throughput, where ϕ_i is a positive constant (acting as a queueing weight) for flow i. Their design goal is to achieve system throughput maximization with the aid of time-varying channel condition knowledge, subject to the throughput fairness constraint. At the beginning of each time-slot, the scheduler chooses a user to transmit according to its maximum possible transmission rate μ_i^k, which is determined by the user's channel state. After the selected user receives data in this time-slot, the system throughput is increased by the amount of data transmitted in this time-slot. Their design goal is a scheduler which can maximize system throughput while performing as a guaranteed rate node by exploiting known channel states, and provide some performance bound with a low computation complexity for wireless networks.

Assaad and Zeghlache (2006) propose an opportunistic scheduler that allows to transmit streaming traffic over HSDPA without losing much cell capacity. The proposed scheduler consists of modifying the priority as in (2.4).

$$\mu_i^*(t) = \text{argmax}_{\mu_{\mathrm{i}}(\mathrm{t})} \left\{ -\log(\delta) \frac{\mu_i(t)}{\bar{Z}_i} \left(1 - \frac{\frac{\mu_i(t)}{G\mu_i(t)}}{\sum_{j=1}^{N} \frac{\mu_j(t)}{G\mu_j(t)}} \right) \right\} \tag{2.4}$$

Where $\mu_i^*(t)$ is the transmission rate for user i in the current time-slot t, μ_i is the achieved bit rate, $G\mu_i$ is the required bit rate and N is the number of simultaneous streaming users in the cell. δ denotes the system data unit error ratio at the RNC level. The algorithm allocates the channel to the user having a compromise between the actual channel conditions (represented by the bit rate), the mean statistical channel conditions ($\bar{Z}$) and the achieved bit rate according to the required bit rate. When all the users have the same achieved rate and the same required bit rate, the channel is allocated to the user having the $\max(\mu_i/\bar{Z}_i)$ which allows to take advantage of the instantaneous peaks in the received signal i.e. to keep track of the fast fading peaks in the radio channel. When all the users have the same channel conditions, the TTI is then allocated to the user having the most need in bit rate (i.e. highest required bit rate or lowest achieved bit rate) according to the need in bit rate of the other users.

For the uplink, we consider LTE systems only. The number of papers about opportunistic uplink scheduling for LTE has been small so far. The most noted work in this area is that of Lim *et al.* In (Lim et al. 2006) the authors suggests assigning resource blocks to users who obtain the highest marginal utility. SC-FDMA is utilized for the uplink in LTE. To preserve the single carrier property, it is required for the localized mode of SC-FDMA that the resource units assigned to one user to be contiguous. The algorithm suggested in (Lim et al. 2006) can be interpreted as a special case of the gradient scheduling rule discussed in (Stolyar 2005).

Another work, is the thesis work (Jersenius 2007) that provides a number of basic[2] allocation rules. The work suggests a channel dependent time domain scheduler assigns all RBs to the user who has the largest average gain to interference ratio (GIR) in every transmission time interval (TTI). The work also suggests a time-

[2]Due to the contiguity constraint that makes fair resource allocation difficult.

frequency scheduler that assigns groups of multiple consecutive RBs to users with the highest average GIR over the RBs of a group in every TTI. The number of the groups can be equal to the number of active users as long as the number of active users does not exceed the total number of resource blocks.

2.2 Admission Control

The admission controller in an opportunistic system bases its decision on the channel behavior of the ongoing calls. The opportunistic scheduler would provide different levels of QoS to the users depending on their channel conditions. However, the controller will impose the minimum acceptable QoS level for all users in the system. For this, the controller needs to estimate the impact the new user will have on the system. This estimation is possible in opportunistic systems given that the channel states $\xi_i(t)$ are stationary ergodic processes that can be determined at a specific time t by their cumulative density function $f_{\xi_i}(\xi_i)$, which is independent of t. For example, in (Berggren and Jäntti 2004) the asymptotic throughput of a user with N users in the system obtained with the relatively best scheduler was found to be

$$\bar{x}_i = \frac{W}{\Gamma_i}\frac{\bar{\xi}_i}{N}\sum_{k=1}^{N}\left(\begin{array}{c} N \\ k \end{array}\right)(-1)^{k+1}\frac{y_1}{\bar{\xi}_i}\left(e^{-y_1 k/\bar{\xi}_i} + \sum_{m=2}^{M}\mu^{m-2}(\rho-1)e^{-\rho^{m-1}y_1 k/\bar{\xi}_i}\right) \tag{2.5}$$

where W is the bandwidth, Γ_i is the bit-energy-to-interference ratio for user i, ξ_i is the mean channel state, $\frac{y_1}{\xi_i}$ denotes a relative threshold and ρ is the number of available rates in the system. With this information at the admission controller's disposal, it could estimate the impact of the new user and form the decision on whether to accept the new user or reject it.

In (Lee and Chong 2006), the authors propose a measurement-based admission control algorithm combined with a utility based opportunistic scheduling algorithm. When a new call arrives, it is admitted and served by using a predefined utility function for admission trial. If the average throughput of the new arrival after a certain trial period satisfies its minimum requirement, then the new arrival is admitted, otherwise it is blocked. A user k arrives to the system. $U_k(\bar{x}_k)$ denotes a utility function that is different for real-time users and best effort users, $\nabla U_k(\bar{x}_k)$ is the gradient of that function. $\bar{x}_k$ represents the average throughput of user k, β_k is the mean rate of user k. If $U_k(\bar{x}_k)$ is selected such that $\beta_k \nabla U_k(\bar{x}_k) \leq \beta_i \nabla U_i(\bar{x}_i)$ for $\bar{x}_i \leq m_i$ (m_i is the minimum demand rate for user i) for every existing QoS user i then serving user k according to (2.6) does not violate the minimum guarantee of existing QoS users.

$$j^* \in \mathrm{argmax}_j \nabla U_j(\bar{X}_j(t))\mu_j(t) \tag{2.6}$$

With $\mu_j(t)$ as the service rate of user j at time t.

In (Hu and Zhang 2002) the authors propose a smooth admission control scheme. The basic idea of the controller is to increase gradually the amount of time allocated to the new users of a trial period. Specifically, they first propose an adaptive resource allocation algorithm-QoS driven weight adaptation for weighted proportional fair opportunistic scheduling. Building on this algorithm, they allocate more time resources to the new users by adaptively increasing their weights, while ensuring the QoS of other active users. Based on the observed average throughput, an admission decision is made within a time-out window: the system admits an incoming user if its throughput is above the threshold; otherwise, the user drops out and requests access again after a back-off time.

Chapter 3

Models and assumptions

The end to end communication link consists of several layers with different functionalities. Each layer applies a number of tasks to the outgoing and incoming data. The focus will be on the medium access control (MAC) sub-layer. The main job of the MAC protocol is to regulate the usage of the medium, and this is done through a channel access mechanism. A channel access mechanism is a way to divide the main resource between nodes by regulating its usage. The access mechanism tells each terminal when it can transmit and when it is expected to receive data. The channel access mechanism is the core of the MAC protocol.

Wireless transmission is characterized by the generation, in the transmitter of an electric signal representing desired information, the propagation of corresponding radio waves through space and a receiver that estimates the transmitted information from the recovered electric signal. There are four basic modes of propagation:

- Free-space or line of sight (LOS): as the name implies, corresponds to a clear transmission between the transmitter and receiver.

- Reflection: which happens as a result of the bouncing of waves from surrounding objects such as buildings and passing vehicles.

- Diffraction: that results from the bending of waves.

- Scattering: which occurs when waves are forced to deviate from a straight trajectory by one or more localized non-uniformities in the medium through which they pass.

Other factors that limit communication quality are noise and interference. Interference stems from the fact that the frequency spectrum is a scarce resource that has to be divided in an efficient way among many users. However, different circumstances may lead to users interfering with the transmission of each other. This results in the degradation of the signal quality at the receiver and the loss of information. The channel quality is measured from the signal-to-interference+noise ratio SINR which as calculated from the following equation.

$$\gamma_{ij} \quad = \quad \frac{G_{ij}.P_i}{\sum_{b \neq i} P_b \theta_{i,b} G_{ib} + N_i} \tag{3.1}$$

Where γ_{ij}, denotes the SINR of a signal being transmitted by j and received by i. G_{ij} represents the link gain between i and j. P_j is the transmission power of j. N_i denotes the the (thermal) noise power at receiver i. θ_{ib} denotes the normalized squared cross correlations between waveforms. The work presented in this thesis is based on snapshot assumptions in which link gains are fixed. The schemes in this work are modeled for a **single-cell system**. Channel quality normally defines how many information bits could the channel convey i.e. channel capacity. Channel capacity is proportionally related to channel condition. In 1948, Claude Shannon an American electrical engineer and mathematician, found that there is a bound on the maximum amount of error-free information bits that can be transmitted over a communication link with a specified bandwidth in the presence of the noise interference. This bound is defined in Shannon's formula (Shannon 1948). A transmission link with a bandwidth B and SINR γ will have the following limit.

$$C = B \log_2 (1 + \gamma) \ \text{(bps)} \tag{3.2}$$

3.1 Fading

Mobile communication systems are mostly used in and around centers of population. As a result, the communication is mostly achieved via scattering of electromagnetic waves from surfaces or diffraction over and around buildings. These multiple propagation paths have both slow and fast aspects (Haykin and Moher 2005):

1. Slow fading arises from the fact that most of the reflectors and diffracting objects along the transmission path are distant from the terminal. The motion of the terminal relative to these distant objects is small; consequently, the corresponding propagation changes are slow. The slow fading process is also referred to as shadowing or lognormal fading.

2. Fast fading is the rapid variation of signal levels when the user terminal moves short distances. Fast fading is due to reflections of local objects and the motion of the terminal relative to those objects. The received signal will thus be the sum of a number of signals reflected from local surfaces. These signals sum in a constructive or destructive manner, depending on their relative phase relationships. The resulting phase relationships are dependent on relative path lengths to the local objects and they can change significantly over short distances. In particular, the phase relationships depend on the speed of motion and the frequency of transmission.

As the carrier frequency of a signal is varied, the magnitude of the change in amplitude will vary. The coherence bandwidth measures the minimum separation in frequency after which two signals will experience uncorrelated fading.

- In **flat fading**, the coherence bandwidth of the channel is larger than the bandwidth of the signal. Therefore, all frequency components of the signal will experience the same magnitude of fading.

- In **frequency-selective fading**, the coherence bandwidth of the channel is smaller than the bandwidth of the signal. Different frequency components of the signal therefore experience decorrelated fading.

In a frequency-selective fading channel, since different frequency components of the signal are affected independently, it is highly unlikely that all parts of the signal will be simultaneously affected by a deep fade. Certain modulation schemes such as OFDM is well-suited for employing frequency diversity to provide robustness to fading. OFDM divides the wideband signal into many slowly modulated narrowband subcarriers, each exposed to flat fading rather than frequency selective fading. This can be combated by means of error coding and sometimes simple equalization and adaptive bit loading. Inter-symbol interference is avoided by introducing a guard interval between the symbols.

Frequency-selective fading channels are also dispersive, in that the signal energy associated with each symbol is spread out in time. This causes transmitted symbols that are adjacent in time to interfere with each other. Equalizers are often deployed in such channels to compensate for the effects of the intersymbol interference.

3.1.1 Multipath component model

For a mobile traveling through an interference pattern, the following continuous-time fading model is adopted.

$$h(t) = \underbrace{\frac{1}{\sqrt{K+1}} \lim_{M\to\infty} \frac{1}{\sqrt{M}} \sum_{m=1}^{M} a_m e^{j(\omega_D cos(\theta_m)t+\phi_m)}}_{\bar{x}(t)} + \underbrace{\sqrt{\frac{K}{K+1}} e^{j(\omega_D cos(\theta_0)t+\phi_0)}}_{y(t)} \tag{3.3}$$

x(t) represents what is known as the Rayleigh fading component (Hummels and Ratcliffe 1981). y(t) is the LOS component. $\omega_D = 2\pi v/\lambda$ is the maximum Doppler angular frequency which is determined by the velocity v and wavelength λ. M is the number of independent scatterers, a_m is the amplitude of path m, θ_m is the angle between the incoming wave and mobile direction and ϕ_m is the phase of the incoming wave.

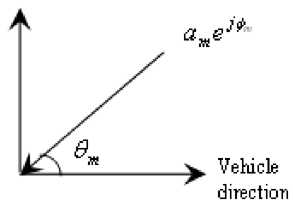

Figure 3.1: Contribution of one wave

K is the Rician factor and is defined by:

$$K = \frac{\text{power in the dominant path}}{\text{power in the scattered path}} = \frac{E|y(t)|^2}{E|x(t)|^2} \tag{3.4}$$

If $K > 0$ then there exists a dominant path in (3.3) which is $y(t)$ and the result will be a Rician fading model. If $K = 0$ the dominant path term will disappear and the result will be a Rayleigh fading model.

Thus, assuming only Rayleigh fading, the link gain G_{ij} in equations (3.1) can be written as:

$$G_{ij} = d_{ij}^{-\alpha}.10^{\xi/10}.\sum_{l=1}^{L}|c_l|^2 \tag{3.5}$$

where d_{ij} is the distance between i and j. α is an attenuation factor which ranges between 2 and 4. ξ is a normal random variable representing the slow fading component. The variable c is a Rayleigh distributed random variable. L denotes the number of diversity paths. The term $\sum_{l=1}^{L}|c_l|^2$ represents the fast fading component.

3.1.2 Input output models for wireless channels

Channel impulse response

Let $\beta_j(t)$ denote the gain in the j^{th} multi path. Let $\tau_i(t)$ denote the total propagation delay. The total electric field expression for an input $\cos(2\pi f t)$ is

$$E_r(t) = \sum_{j=1}^{k} \Re[\beta_j(t) \exp{(i2\pi f[t - \tau_j(t)])}] \tag{3.6}$$

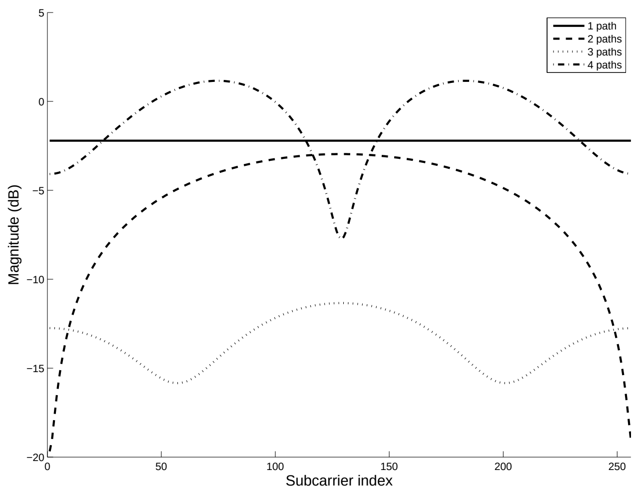

Figure 3.2: Magnitude plot for various number of multipath components

Where $\Re$ denotes the real part of the argument. Eq. (3.6) is viewed as giving the response to sinusoids at arbitrary frequencies. Define the time varying system function $h(f, t)$ as

$$h(f, t) = \sum_{j=1} \beta_j(t) \exp(-i2\pi f t) \tag{3.7}$$

Rayleigh fading distribution

When there are a large number of scatterers in the channel that contribute to the signal propagation, application of the central limit theorem leads to a Gaussian process model for the channel impulse response. If the process is zero mean , then the envelope (i.e. square of the real and imaginary part) has a Rayleigh distribution and the phase is uniformly distributed in the interval $(0, 2\pi)$. The probability density function (PDF) of the received signal power $|h|$ is given by

$$p_{|h|}(r) = \frac{2r}{\Omega} e^{\frac{-r^2}{\Omega}}, r \geq 0 \tag{3.8}$$

Where Ω is the average power received by a terminal. The variable r denotes the received signal. Hence, the envelope power $|h|^2$ will follow an exponential distribution. In our work we consider Jakes model for Rayleigh fading. Jakes model

(Jakes 2001) is based on the summing of several sinusoids. The normalized autocorrelation function of a Rayleigh faded channel with motion at a constant velocity is a zeroth-order Bessel function of the first kind;

$$\phi_H(\Delta t) = J_0(2\pi f_m \Delta t) \tag{3.9}$$

at delay Δt when the maximum Doppler shift is f_m. The Doppler shift is a measure of time variation in the channel; the larger the value, the more rapidly the channel changes in time. Its reciprocal $T_{coh} = \frac{1}{f_m}$ is called the *coherence time*. Each channel remains strongly correlated during this time. Analogously, the *coherence bandwidth* B_{coh} is a statistical measure of the range of frequencies over which the channel can be considered "flat", i.e., having approximately equal gain and linear phase. In other words, coherence bandwidth is the range of frequencies over which any two frequency components have a strong correlation. $B_{coh} = \frac{1}{T_m}$, where T_m is the largest delay produced by the channel.

3.1.3 The Frequency Selective Channel Model

A wideband fading channel can be modelled as a sum of several differently delayed, independent Rayleigh fading processes. The corresponding channel impulse response is described as

$$h(t,\tau) = \sum_{p=1}^{P} a_p \cdot R_p(t) \cdot \delta(\tau - \tau_p), \tag{3.10}$$

where a_p is the normalised amplitude such that $\sum_{p=1}^{P} a_p^2 = 1.0$, $R_p(t)$ is the Rayleigh fading process with $E[R_p^2] = 1.0$ and τ_p delay of the p-th path.

Fig. 3.2 represents the effect of selective fading in a multi carrier system. we could see that as the number of components increase, the frequency selectivity also increases. This shows that multipath makes the channel frequency selective.

For our multi-carrier model we consider a typical urban area propagation model with a 6 tap setting specified in 3GPP release 7. A description of this model could be seen in Figures 3.3 and 3.4.

3.2 Access technologies

The main access technologies considered in this thesis are time/frequency/code division multiple access technologies (TDMA/FDMA/CDMA). Most of the considered models in this work combine two or more of these technologies as well as extensions of particular technologies such as OFDMA and SC-FDMA. In TDMA, time is the resource that is shared among the users. Time is divided into time-slots known as TTIs. Time division multiple access (TDMA) is a channel access method for shared medium (usually radio) networks. It allows several users to share the same frequency channel by dividing the signal into different time-slots. Users transmit in rapid succession or by selection, each using his own time-slot. This allows multiple stations to share the same transmission medium (e.g. radio frequency channel) while using only the part of its bandwidth they require. Using TDMA allows the exploitation of the channel variations the users experience making it possible to use adaptive modulation and coding (AMC), improving this way transmission efficiency. In Frequency Division Multiple Access (FDMA) the given Radio Frequency (RF) bandwidth is divided into adjacent frequency segments. Each segment is provided with bandwidth to enable an associated communications signal to pass through a transmission environment with an acceptable level of interference from communications signals in adjacent frequency segments. The bandwidth is divided between users who transmit simultaneously. In practice, the available bandwidth is subdivided into a large number of narrow band channels. These bands in turn are assigned to different users. Orthogonal Frequency Division Multiple Access (OFDMA) is one extension of FDMA which uses a large number of closely-spaced orthogonal sub-carriers. Each sub-carrier is modulated with a conventional modulation scheme (such as quadrature amplitude modulation) at a low symbol rate, maintaining data rates similar to conventional single-carrier modulation schemes in the same bandwidth. In practice, OFDM signals are generated using the Fast Fourier transform algorithm. The

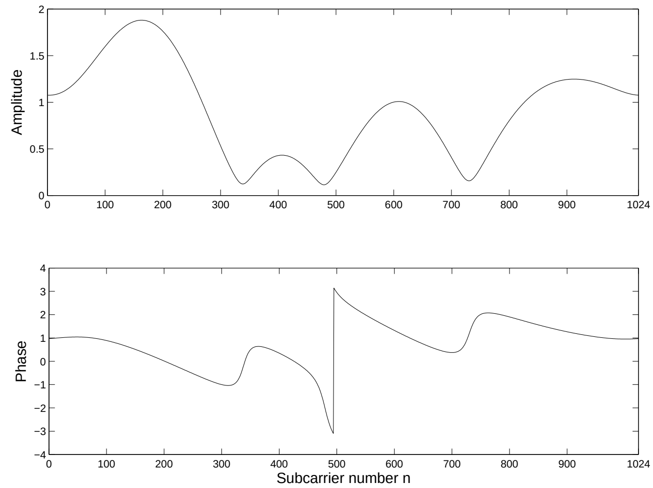

Figure 3.3: Selective frequency channel response

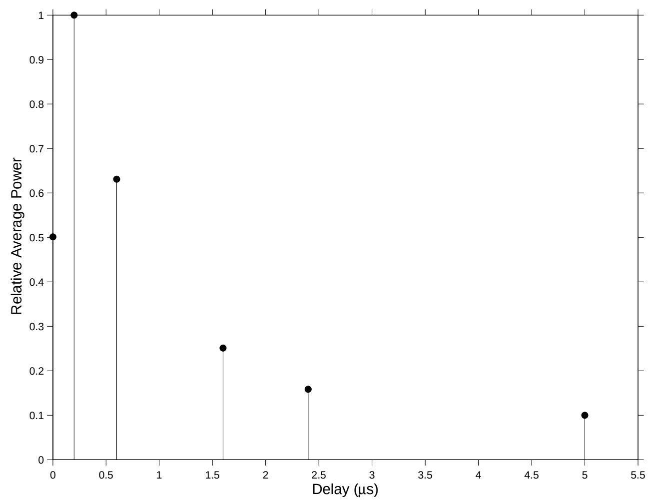

Figure 3.4: Relative powers of the delay profile

other extension; Single Carrier Frequency Multiple Access (SC-FDMA) utilizes single carrier modulation at the transmitter and frequency domain equalization at the receiver is a technique that has similar performance and essentially the same overall structure as those of an OFDMA system. One prominent advantage over OFDMA is that the SC-FDMA signal has lower peak to average power ratio due to the single carrier property. SC-FDMA has drawn great attention as an attractive alternative to OFDMA, especially in the uplink communications where lower PAPR greatly benefits the mobile terminal in terms of transmit power efficiency. Code division Multiple Access (CDMA) describes a communication channel access principle that employs spread-spectrum technology and a special coding scheme (where each transmitter is assigned a code). CDMA is a form of "spread-spectrum" signaling, since the modulated coded signal has a much higher bandwidth than the data being communicated. In this thesis we considered HSDPA and LTE technologies. HSDPA as mentioned in Chapter 1 is an evolution step of WCDMA where a time dimension is added to the access function. Adding the time property exposed the possibility of utilizing the opportunistic features of channels. In LTE time access is added to the frequency access function.

3.3 Retransmissions

Perfect channel estimation requires timely knowledge of the channel state. In practice, we would need to select the rate based on possibly outdated and imperfect channel state information (CSI). If the selected rate exceeds the instantaneous channel capacity $\mu_i(t) > C_i(t)$, then the transmitted data cannot be decoded at the receiver. In that case a request for retransmission is made. In our model, we consider synchronous non-adaptive hybrid automatic retransmission requests (HARQ). In this protocol, retransmissions will occur at a predefined (normally fixed) time after the previous (re)transmission using exactly the same modulation and coding rate even though the channel might have changed. The benefit with synchronous non-adaptive HARQ is that the control signaling can be minimized since the HARQ process ID does not need to be signaled explicitly and a received HARQ negative acknowledgment (NACK) can be used as an implicit grant for a HARQ retransmission. In chase combining HARQ (Sanayei et al 1985), the receiver coherently combines the original code block and the retransmitted block. All the transmitted bit energy can be harnessed by the receiver by combining the erroneously received code block with the consecutive copies transmitted by the ARQ process. The method also includes a loss that is added to the cumulated E_b/N_0, this indicates that a perfect gain is not achieved and a certain loss is produced. Transmission is successful when

$$\sum_{m=1}^{M} \left(\frac{E_b}{N_0}\right)_m \geq \left(\frac{E_b}{N_0}\right)_0 * L, \tag{3.11}$$

where $m = 1, 2, \cdots, M$ is the transmission number with M being the maximum allowed number of retransmissions. L denotes the combining loss. $(\frac{E_b}{N_0})_m$ is the bit-to-energy ratio at the time of transmission. $(\frac{E_b}{N_0})_0$ is the bit-to-energy ratio at the time of scheduling. If the receiver is able to decode the code block after combining the original packet and the retransmitted replicas of that packet, the actual rate would become $\frac{\mu(t)}{m}$. Decoding fails if this rate still exceeds the capacity of the channel.

Chapter 4

Downlink Scheduling

The resource allocation problem can be solved as an optimization problem having the QoS demands as constraints or solve it using some control engineering methods. In (Liu et al., 2001a) a simple I-control (stochastic approximation) method was introduced to control the QoS. If a data rate constraint of a user was not met in a time window, a weight could be added to make its selection more probable. If the rate allocated to the user was higher than target, the weight could be decreased. This approach converts the scheduling problem into a control problem. In general the problem of scheduling packets over a fading channel could be viewed as stochastic optimal control problem. In (Liu et al. 2001b), a method for controlling the resource allocation for the different users was suggested. The proposed scheme added one control parameter to the scheduling metric that was changed based on the observed channel access time in some time window. In this thesis, opportunistic scheduling is proposed for flows that have target QoS levels, such as mean data rates or mean packet delays. This corresponds e.g. to the case where a leaky bucket traffic shaping filter is applied at the edge router of the radio access network. The proposed scheduler resembles the scheduler suggested in (Liu et al., 2001a), but instead of modifying the scheduling metric, we control the active set of users, i.e. the number of active users at a given moment of time. The advantage of our approach is that it can be utilized with any scheduling rule. In addition, it avoids the difficulties in shaping the decision variable distributions by simply averaging over the performance achievable with different number of users. Our method resembles the one suggested for multiple channels in (Zhang et al. 2005). In their work they utilized stochastic approximation algorithm to guarantee certain quality of service level in terms of minimum data rate. We note that the stochastic approximation algorithm can result to either very slow convergence or very high variance of the control parameters. In our work we utilize the Kalman filter due to the noisy estimates of the mean QoS that result from averaging over finite time windows.

4.1 System model

Consider a downlink of DS-CDMA system. Let $\zeta_i(t)$ denote the received carrier-to-interference ratio of a user i at time instant t and $\bar{\zeta}_i$ denote its expected value. Assuming that the channel is stationary, more specifically, it is assumed that the fast fading process is stationary and ergodic so that time averages approach the mean variable as the number of samples increase. Later this assumption is relaxed and we consider the case where the channel statistics are slowly changing.

The mobile is assumed to estimate the channel based on a pilot signal transmitted by the base station. Based on the channel measurement the mobile then determines the maximum data rate that it could achieve under the current channel condition $\mu_i(t) = f(\zeta_i(t))$ and reports that back to the base station which then utilizes this information for making the scheduling decision. The PF scheduler (2.2) is considered as the base for the scheduler.

4.2 Quality control

Assume that user i requests some mean QoS s_i^{req}. The objective of the Quality control apparatus is to find the activity probabilities $q_i = Pr\{i \in \mathcal{A}(t)\}$ so that $\bar{s}_i = s_i^{req}$. In case the requests are unachievable, we want to allocate the resources in a fair manner. That is; in the congested case, we want the quality control scheme to fall-back to the proportional fair scheduler.

Assume that time is divided into scheduling time intervals or frames that consist of several time slots. During a frame n the probabilities $q_i^{(n)}$ are kept fixed and in each slot the set of active users is randomly determined. The achieved quality level in the frame $s_i^{(n)}$ is then observed based on which control actions are taken. The quality control problem can be solved using a simple integral controller:

$$q_i^{(n+1)} = \min\left\{1, \max\left\{0, q_i^{(n)} + \beta_i\left(\bar{s}_i^{req} - \bar{s}_i^{(n)}\right)\right\}\right\} \tag{4.1}$$

where β_i is some positive integration gain, for now one. The algorithm will be called the Quality Control Algorithm (QCA).

We note that in the original PF scheme $q_i = 1$ for all i. That is, if a user has data in its buffer it will belong to the active set.

4.2.1 Convergence analysis

Let $Q = (q_1, q_2, \cdots, q_N)'$ denote the activity probability vector that contains the activity probabilities of users $1, 2, \cdots, N$ and $B = (\beta_1, \beta_2, \cdots \beta_N)'$ denote the gain vector. Define a mapping:

$$\mathcal{T}_i(Q, B) = q_i + \beta_i(\bar{s}_i^{req} - \bar{s}_i(q_i)) \tag{4.2}$$

let us define a set of feasible probabilities:
$\mathcal{Q} = \{Q : 0 \leq Q \leq 1\}$.

Proposition 1. *Suppose that $\mathcal{Q}$ is convex. If $\bar{s} : \Re^n \mapsto \Re^n$ is continuously differentiable and there exists a scalar $\alpha \in [0, 1)$ such that*

$$\|I - \beta_i(\nabla_i \bar{s}_i(Q))'\|_\infty + \Sigma_{j\neq i}\|\beta_i(\nabla_j \bar{s}_i(Q))'\|_\infty \leq \alpha, \quad \forall i \tag{4.3}$$

then the mapping $\mathcal{T} : \mathcal{Q} \mapsto \Re^n$ defined by $\mathcal{T}_i(Q, B) = q_i + \beta_i\left(\bar{s}_i^{req} - \bar{s}_i(q_i)\right)$ is a *contraction with respect to the maximum norm* $\|.\|_\infty$

Proof. We fix some i and $Q_1, Q_2 \in \mathcal{Q}$, and we define a function $g_i : [0,1] \mapsto \Re^{n_i}$ by:

$$g_i(t) = tq_{i1} + (1-t)q_{i2} - \beta_i \bar{s}_i(tQ_1 + (1-t)Q_2) + \beta_i \bar{s}_i^{req} \tag{4.4}$$

Notice that g_i is continuously differentiable. Let dg_i/dt be the n_i-dimensional vector consisting of the derivatives of the components of g_i. We then have

$$\|\mathcal{T}_i(Q_1) - \mathcal{T}_i(Q_2)\| = \|g_i(1) - g_i(0)\|$$
$$= \|\int_0^1 \frac{dg_i(t)}{dt} dt\|_\infty \leq \int_0^1 \|\frac{dg_i}{dt}(t)\|_\infty dt \leq \max_{t\in[0,1]} \|\frac{dg_i}{dt}(t)\|_\infty$$

It therefore, suffices to bound the norm of dg_i/dt. The chain rule yields

$$\begin{aligned}
&\|\frac{dg_i}{dt}(t)\|_\infty \\
&= \|q_{i1} - q_{i2} - \beta_i(\nabla\bar{s}_i(tQ_1 + (1-t)Q_2))'(q_{i1} - q_{i2})\|_\infty \\
&= \|[I - \beta_i(\nabla_i\bar{s}_i(tQ_1 + (1-t)Q_2))'](q_{i1} - q_{i2}) \\
&\quad -\Sigma_{j\neq i}\beta_i(\nabla_j\bar{s}_i(tQ_1 + (1-t)Q_2))'(q_{j1} - q_{j2})\|_\infty \\
&\leq \|I - \beta_i(\nabla_i\bar{s}_i(tQ_1 + (1-t)Q_2))'\|_\infty.\|(q_{i1} - q_{i2})\|_\infty \\
&\quad +\Sigma_{j\neq i}\beta_i(\nabla_j\bar{s}_i(tQ_1 + (1-t)Q_2))'\|_\infty.\|(q_{j1} - q_{j2})\|_\infty \\
&\leq \alpha \max_j \|q_{j1} - q_{j2}\|_\infty = \alpha\|Q_1 - Q_2\|
\end{aligned}$$

which establishes the contraction property. ■

Lemma 1. *Assume the following:*

(a) The set $\mathcal{Q}$ *is convex, and the function* $\bar{s} : \Re^n \mapsto \Re^n$ *is continuously differentiable.*

(b) There exists a positive constant κ *such that*

$$\nabla_i\bar{s}_i(Q) \leq \kappa, \qquad \forall Q \in \mathcal{Q},\ \forall i$$

(c) There exists some $\gamma > 0$ *such that*

$$\Sigma_{j\neq i}|\nabla_j\bar{s}_i(Q)| \leq \nabla_i\bar{s}_i(Q) - \gamma, \qquad \forall Q \in \mathcal{Q},\ \forall i$$

Then the mapping $\mathcal{T} : \mathcal{Q} \mapsto \Re^n$ defined by $\mathcal{T}_i(Q,B) = q_i + \beta_i\left(\bar{s}_i^{req} - \bar{s}_i(q_i)\right)$ is a contraction with respect to the maximum norm, provided that $0 < \beta_i \leq \frac{1}{\gamma}$.

Proof. Under the assumption $0 < \beta_i \leq \frac{1}{\gamma}$, we have

$$\begin{aligned} & |1 - \beta_i \nabla_i \bar{s}_{q_i}| + \beta_i \Sigma_{j \neq i} |\nabla_j \bar{s}_i(q_i)| \\ &= 1 - \beta_i (\nabla_i \bar{s}_{q_i} - \Sigma_{j \neq i} |\nabla_j \bar{s}_i(q_i)|) \\ &\leq 1 - \beta_i \gamma < 1 \end{aligned} \tag{4.5}$$

which shows that inequality (4.3) holds. The result follows from Prop. (1). ■

4.2.2 Examples

Rate control

It is useful to request specific rates in the case the receiver's processing rate is less than the transmitter's service rate. This kind of rate control will help match the transmitter's service rate to that of the receiver's in an attempt to avoid congestion and consequent overflow of the receiver's buffer. In this case the QoS metric $\bar{s}_i$ is equal to the mean data rate $\bar{x}_i$. The algorithm in this case will be referred to as the Rate Control Algorithm (RCA). Let $\mathcal{N}$ denote the set of admitted users and let N denote the cardinality of that set. Due to the different activity of the users, at a given instant of time t, the set of active users can be an arbitrary subset of the users $A(t) \subset \mathcal{N}$. Let us order all the $S = \sum_k \binom{N}{k}$ possible subsets of $\mathcal{N}$: $\{\mathcal{A}_k, k = 1, 2, \cdots S\}$. Let A_k denote the cardinality of subset k. Let $\mathcal{A}_S = \mathcal{N}$ denote the active set during which all users are active and $\mathcal{A}_l = \emptyset$ denote the nonactive set.

Now assume that all users individually make the decision whether to be active or idle at a given instant of time. Let q_i denote the probability that the user i decides to be active. Let π_k denote the probability that the set k was used at a particular time. It follows that

$$\pi_k(Q) = \prod_{m \in \mathcal{A}_k} q_m \prod_{n \in \mathcal{N} \setminus \mathcal{A}_k} (1 - q_n) \tag{4.6}$$

Let $\bar{x}_i(\mathcal{A}_k)$ denote the expected data rate of a user i when the active set of users was $\mathcal{A}_k$. Furthermore, let $\mathcal{K}_i$ denote the selection of active sets that contain user i. That is $i \in \mathcal{A}_k$ if and only if $k \in \mathcal{K}_i$.

With the help of the above notation we can now express the expected data rate of user i as follows

$$\bar{x}_i = \sum_{k \in \mathcal{K}_i} \bar{x}_i(\mathcal{A}_k)\pi_k(Q) \tag{4.7}$$

For the sake of comparison, we note that in the original PF scheme $q_i = 1$ for all i. That is, if a user has data in its buffer it will belong to the active set. Consequently the expected rate of the user corresponds to $\bar{x}_i(\mathcal{A}_S)$.

In addition, we note that in most scheduling rules, especially in the proportional fair case $\bar{x}_i(\mathcal{A}_l) \geq \bar{x}_i(\mathcal{A}_k)$ if $A_k \geq A_l$. That is, the less there are users, the higher the chance that user i gets selected and the higher its data rate will be.

Let $\pi_{k,i} = \frac{\pi_k}{q_i}$ for all $k \in \mathcal{K}_i$. It follows that

$$\pi_{k,i} = \prod_{\substack{m \in \mathcal{A}_k \\ m \neq i}} q_m \prod_{n \in \mathcal{N} \setminus \mathcal{A}_k} (1 - q_n) \quad k \in \mathcal{K}_i \tag{4.8}$$

which is independent of q_i.

The mapping in eq. (4.2) will now become:

$$\mathcal{T}_i(Q;B) = q_i + \beta_i \left(\bar{x}_i^{req} - q_i \sum_{k \in \mathcal{K}_i} \bar{x}_i(\mathcal{A}_k)\pi_{k,i} \right) \tag{4.9}$$

$\mathcal{T}(Q) = (\mathcal{T}_1(Q), \mathcal{T}_2(Q), \cdots, \mathcal{T}_N(Q))'$ denote the rate control mapping. We note that

$$q_i^{(n+1)} = \min\left\{1, \max\left\{0, \mathcal{T}_i(Q^{(n)}, B^{(n)})\right\}\right\} \tag{4.10}$$

Furthermore, let us define a set of feasible probabilities $\mathcal{Q} = \{Q : 0 \leq Q \leq 1\}$. Now we are ready to consider the convergence properties of the algorithm:

Lemma 2. *If* $0 < \beta_i \leq \frac{1}{max_{k \in \mathcal{K}_i} \bar{x}_i(\mathcal{A}_k)}$, *then the mapping* $\mathcal{T}(Q, B)$ *is a contraction mapping for all* $Q \in \mathcal{Q}$.

Proof. The proof for this lemma will follow from proposition (1) and lemma (2) with $\gamma = max_{k \in \mathcal{K}_i} \bar{x}_i(\mathcal{A}_k) > 0$ which satisfies condition (c) in lemma (1). ■

Proposition 2. *If $0 < \beta_i \leq \frac{1}{max_{k \in \mathcal{K}_i} \bar{x}_i(\mathcal{A}_k)}$, then RCA converges to a unique fixed point $0 \leq Q^* < 1$ where all the users are supported with the rates that they requested, if such a point exists.*

Proof. The fixed point satisfies the following for all $0 < B < \infty$

$$q_i^* = \mathcal{T}_i(Q^*, B) \tag{4.11}$$

$$q_i^* = q_i^* + \beta_i \left(\bar{x}_i^{req} - q_i^* \sum_{k \in \mathcal{K}_i} \bar{x}_i(\mathcal{A}_k) \pi_{k,i}^* \right) \tag{4.12}$$

By subtracting q_i^* from both sides and dividing by β_i, the above can be written as

$$\bar{x}_i^{req} = q_i^* \sum_{k \in \mathcal{K}_i} \bar{x}_i(\mathcal{A}_k) \pi_{k,i}^* \tag{4.13}$$

The right hand side of the above equation is equal to the asymptotic value of $\bar{x}_i(n)$ as $n \to \infty$. Thus we can conclude that if $0 \leq Q^* < 1$, then $\bar{x}_i^{req}$ is achieved for all i.

The uniqueness of the fixed point follows from the fact that $\mathcal{T}(Q, B)$ is a contraction mapping. ■

Proposition 3. *If the system is congested such that not all requested rates can be supported and $0 \leq \beta_i \leq \frac{1}{max_{k \in \mathcal{K}_i} \bar{x}_i(\mathcal{A}_k)}$, then RCA converges to a unique fixed point $0 < Q^* \leq 1$, where at least for one user $q_i = 1$. The rate of the non-supported users is at least as high as the rate achievable by using PF scheduling.*

Proof. The assumption is that at least one user has $\bar{x}_i^{req}$ larger than what can be supported by the system. That is there does not exist $0 \leq Q^* < 1$ such that $\bar{x}_i^{(n)} \to \bar{x}_i^{req}$. In such a case, the error term in RCA is all the time positive $\bar{x}_i^{req} - \bar{x}_i^{(n)} \geq 0$ and thus the resulting sequence of probabilities $\{q_i^{(n)}\}$ is monotonously increasing. The min-operator in RCA will cause the probability to saturate to the value $q_i^* = 1$ which corresponds to normal PF operation.

The fixed point data rate of RCA is

$$\bar{x}_i^* = q_i^* \sum_{k \in \mathcal{K}_i} \bar{x}_i(\mathcal{A}_k)\pi_{k,i}^* \tag{4.14}$$

If $q_i^* = 1$, then $\mathcal{K}_i = \{1, 2, \cdots S\}$. That is, i belongs to all active sets. Recall that

$$\pi_{k,i}^* = \prod_{\substack{m \in \mathcal{A}_k \\ m \neq i}} q_m^* \prod_{n \in \mathcal{N} \setminus \mathcal{A}_k} (1 - q_n)^* \quad k \in \mathcal{K}_i \tag{4.15}$$

If $q_i = 1$ for all i, then $\pi_{S,i}^* = 1$ and $\pi_{k,i}^* = 0$ for $k \neq S$. In which case $\bar{x}_i^* = \bar{x}_i(\mathcal{A}_S)$ which is equal to the expected rate of the PF scheme.

Our assumption on the magnitude of the rates was that $\bar{x}_i(\mathcal{A}_l) \geq \bar{x}_i(\mathcal{A}_k)$ if $A_k \leq A_l$. Hence, $\bar{x}_i(\mathcal{A}_S) \leq \bar{x}_i(\mathcal{A}_k)$ for all $k \in \mathcal{K}_i$. Now in case there exists at least one user j for which $0 \leq q_j^* < 1$, then there exists at least one active set $\mathcal{A}_m$ such that $A_m < A_S$ and $\bar{x}_i(\mathcal{A}_S) \leq \bar{x}_i(\mathcal{A}_m)$ and $\pi_{i,m}^* > 0$. Hence the resulting expected rate will be a convex combination of at least two rates $\bar{x}_i(\mathcal{A}_S)$ and $\bar{x}_i(\mathcal{A}_m)$. It thus follows that

$$\bar{x}_i^* = q_i^* \sum_{k \in \mathcal{K}_i} \bar{x}_i(\mathcal{A}_k)\pi_{k,i}^* > \bar{x}_i(\mathcal{A}_S) \quad | \quad \exists j \quad s.t. \quad \bar{x}_j^* < \bar{x}_j(\mathcal{A}_S) \tag{4.16}$$

That is, the expected rate of RCA is larger than the rate of PF. ■

Packet delay control

Packet Delay was also considered as a QoS that could be provided to requesting users. In that case, access to the active set would be controlled by the expected delay experienced by the user. The probability that a user gets a given time slot just depends on the number of users in the active set and the channel statistics of the individual user. Assuming that the fading is fast enough, so that service times of successive packets are independent of each other.

4.3 Imperfect estimates

In practice, the mean value $\bar{s}_i(t)$ is not available. What can be estimated instead is the average $\bar{s}_i(t)$ over some time window. $\bar{s}_i(t)$ can be treated as a noisy estimate for $\bar{s}_i(t)$. In order to cope with imperfect estimates, Kalman state estimation (Kalman 1992) is made. A Kalman Filter is essentially a set of mathematical equations that implements a predictor-corrector type estimator that is optimal in the sense that it minimizes the estimated error covariance - when some presumed conditions are met. The strength of the formulation is that it allows the tracked parameter (in our case the average rate or delay) to change slowly in time regardless of the channel's statistics. A natural candidate for the state vector $X^{(n)}$ would be a vector describing all the $|\mathcal{K}_i| = 2^{N-1}$ possible mean quality parameters $\bar{s}_i(\mathcal{A}_k)$, $k \in \mathcal{K}_i$. This approach, however, suffers from the curse of dimensionality - even with relatively small value of N, the number of states grows to be far too large for real-time operation.

To limit the parameter dimension, it is assumed that the decision variable utilized in the opportunistic scheduler is independent of the moments of the channel state process. For instance, in the proportional fair scheduler, the decision variable asymptotically becomes $\frac{\mu_i(t)}{\bar{x}_i}$, which in case of Rayleigh fading is independent of the mean $\bar{x}_i$. Let $\bar{s}_{n,i} = \bar{s}_i(\mathcal{A}_k)$ for all $A_k = n$, $k \in \mathcal{K}_i$. Furthermore, define $\bar{\pi}_{n,i} = \sum_{k \in \{k \in \mathcal{K}_i : A_k = n\}} \pi_{k,i}$. The mean quality parameter equation can be written as

$$\bar{s}_i = \sum_{n=1}^{N} \bar{\pi}_{n,i} \bar{s}_{n,i} \tag{4.17}$$

which consists of N unknown variables $\bar{s}_{n,i}$.

Let $X_i^{(n)} = (\bar{s}_{1,i}^{(n)}, \bar{s}_{2,i}^{(n)} \cdots \bar{s}_{N,i}^{(n)})'$ be a column vector of the state variables and $C_i^{(n)} = (\bar{\pi}_{1,i}^{(n)}, \bar{\pi}_{2,i}^{(n)} \cdots \bar{\pi}_{N,i}^{(n)})$ be a row vector that maps the state to the measurement value. Let $V_i^{(n)}$ denote white noise process that describes the slowly changing nature of the state parameters. The covariance of the state noise is $\mathrm{E}\{V_i^{(n)}(V_i^{(n)})'\} = \Psi_{VV} > 0$. Note that the state noise is likely to be very highly correlated, so Ψ_{VV} is typically not sparse, nor diagonal. In addition, we have the measurement noise $e_i^{(n)}$ that has

Table 4.1: System parameters for downlink scheduling in HSDPA

Parameter	Value
Carrier frequency	2 GHz
Spreading factor	16
Number of multicodes	10
TTI duration	2 ms
Fading model	One path Rayleigh (Jake's model)
No. of associated DPCH	3
Radio propagation	Site to site distance 500 m
	Min. mobile speed 3 km/h
	Max. mobile speed 10 km/h
BS Tx power	17 W
Hybrid ARQ	Chase combining
Target mean rates	16, 32 and 64 kbps
Target mean packet delays	40, 60 and 80 ms
$\Psi_{ee}, \Psi_{Ve}, \Psi_{VV}$	$10 \times 10^3, 0, 10$ respectively

a covariance of $\Psi_{ee} = \mathrm{E}\{(e_i^{(n)})^2\} = \sigma^2 > 0$. If the channel is varying rapidly that may have negative impact on the estimation accuracy as well. To model this, we assume that the cross covariance $\Psi_{Ve} = \mathrm{E}\{V_i^{(n)} e_i^{(n)}\} \geq 0$. The process dynamics can be expressed as

$$X_i^{(n+1)} = X_i^{(n)} + V_i^{(n)} \tag{4.18}$$

$$\bar{s}_i^{(n)} = \bar{s}_i^{(n)} + e_i^{(n)} \tag{4.19}$$

The Kalman state estimator can be written as

$$\bar{X}_i^{(n+1)} = \bar{X}_i^{(n)} + K_i^{(n)} \left(\bar{s}_i^{(n)} - C_i^{(n)} \bar{X}_i^{(n)} \right) \tag{4.20}$$

$$K_i^{(n)} = (P_i^{(n)} (C_i^{(n)})' + \Psi_{Ve})((C_i^{(n)})' P_i^{(n)} C_i^{(n)} + \Psi_{ee})^{-1} \tag{4.21}$$

$$P_i^{(n+1)} = P_i^{(n)} + \Psi_{VV} - K_i^{(n)} (C_i^{(n)} P^{(n)} (C_i^{(n)})' + \Psi_{ee})(K_i^{(n)})' \tag{4.22}$$

4.4 Numerical example

For this example, the simulator creates an environment of mobile users in a single cell with parameters listed in Table 4.1. The simulator generates three users having different packet sizes. It is assumed that the transmission buffers are always full i.e. there is always data to transmit to each user. The initial values for the activity probabilities were randomly selected. In the simulations two QoS measures are considered: mean rates and mean packet delays. For the former the targets were set to 16, 32 and 64 kbps and for the latter the targets were 40, 60 and 80 ms. Fig. 4.1 illustrates the average and Kalman estimate mean values for data rates and packet delays respectively. It is assumed that all users were admitted to the system at the same time and their initial rates and delays were zero. This explains the transient state each user's average suffers at the beginning of the simulation. As the system stabilizes, the averages tend to keep oscillating at a constant rate, this is due to the finite frame size effect explained in section 7.4. The average values were obtained using the stochastic approximation method. Traffic was considered dynamic and packets were generated according to a log normal distribution. The Kalman filter was later applied to these average values yielding the mean values. Adaptive modulation and coding was used to guarantee high throughputs depending on the channel condition. For the AMC we utilize the quantization values enlisted in Table 7.2.

Fig. 4.2 illustrates the activity probabilities that achieve the target values. We could see from the figures how the probabilities converge to fixed points confirming the convergence analysis outlined in section 4.2.1 The activity probabilities for the rates appear to be decreasing while for the packet delay they're increasing. This is mainly due to the fact that at the beginning packet delays are relatively small due to the small queue sizes that steadily increase. While for date rates, all users will have high activity probabilities to enable them of transmitting and gradually form the desired mean rates.

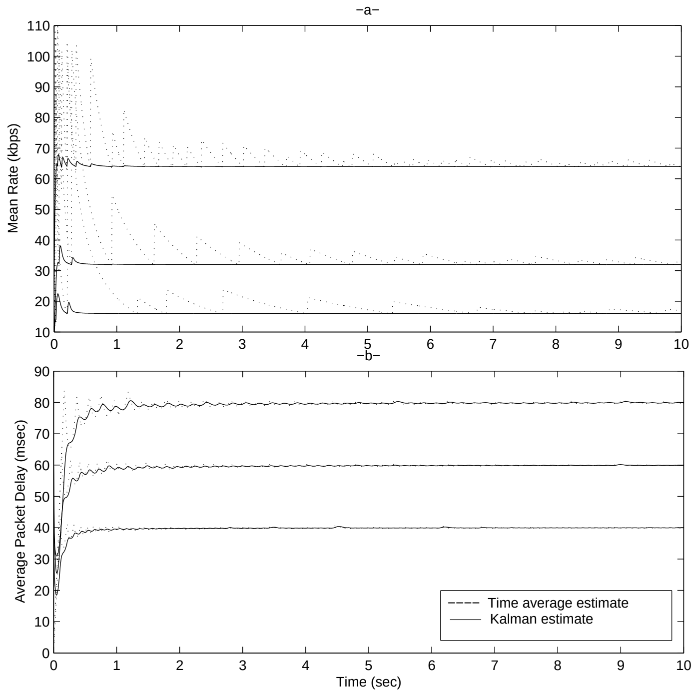

Figure 4.1: a) Time average based estimate & Kalman filter based estimate rates b) Time average based estimate & Kalman filter based estimate delays

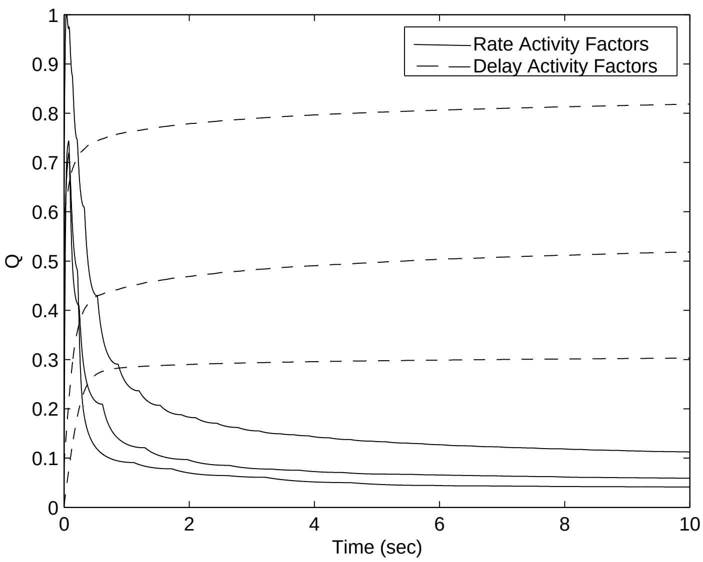

Figure 4.2: Activity probilities for mean rates & delays

4.5 Concluding remarks

In this chapter, an opportunistic scheduler was proposed that would provide users with requested target QoS levels while maintaining fairness between the users in the system. The idea of the algorithm relies on the tuning of a user's activity in a way that it reaches and maintains the designated target. The scheduler has the special property that it can fall back on some fair scheduler when the system becomes congested and target QoS levels cannot be delivered. The proposed scheduler can be utilized to support different QoS classes.

Chapter 5

Uplink Scheduling

The air interface access technologies considered for LTE are OFDMA for the downlink and SC-FDMA for the uplink. SC-FDMA is based on transmitting frequency chunks consisting of multiple subcarriers at a time. These chunks are usually referred to as resource blocks (RB). SC-FDMA has two modes: localized-FDMA (L-FDMA) where users are assigned RBs of adjacent subcarriers and the other mode is distributed-FDMA (D-FDMA) where the subcarriers of one RB are distributed over the entire frequency band but with an equal distance of each other. D-FDMA has the advantage of being robust against frequency selective fading because its information is spread across the entire signal band. Therefore, it offers the advantage of frequency diversity. On the other hand L-FDMA can potentially achieve multi-user diversity in the presence of frequency selective fading if it assigns each user to subcarriers in a portion of the signal band where that user has favorable transmission characteristic (high channel gain). Multi-user diversity relies on independent fading among dispersed transmitters.

Our work follows the approach taken in (Lim et al. 2006). That is, schedulers that seek to maximize a throughput based aggregate utility are considered. However, the difference is in the way RBs are allocated to the users and that the effects of imperfect channel state information and hybrid automatic repeat request (HARQ) on the scheduling decisions are taken into account. In this work, only the L-FDMA case is considered.

5.1 System model

The model consists of a single cell that contains one base station communicating simultaneously with N mobile user terminals. The bandwidth W consists of M subcarriers that are grouped into $L = \frac{W}{\Delta f_c}$ RBs where Δf_c denotes the coherence bandwidth of the channel. Each RB will contain M/L consecutive subcarriers. The channel is assumed to be slowly fading such that the channel state stays essentially constant during one TTI. That is, the coherence time of the channel is assumed to be longer than the duration of the TTI and thus the channel exhibits block fading characteristics. The RBs fade independently, but the fading seen by individual subcarriers in a RB is approximately the same since the subcarrier spacing is small compared to the coherence bandwidth of the channel. Assuming that the channel is subjected to Rayleigh fading. Assuming that the amount of power available for every user will be constant and is represented by the maximum transmission power P_{max}. The transmission time interval for the uplink is 1 ms and consists of two 0.5 ms subframes. The capacity of RB n for user i at TTI k is given by Shannon's formula.

$$C_{i,n} = B_n \log_2 \left(1 + \gamma_{i,n}^{(eff)}\right) \tag{5.1}$$

where B_n is the bandwidth of RB n, $\gamma_{i,n}^{(eff)}(t)$ denotes the effective (SNR) of user i on RB n and is computed as follows

$$\gamma_{i,n}^{(eff)} = \left(\frac{1}{\frac{1}{M} \sum_{k=1}^{M} \frac{\gamma_{i,k}}{\gamma_{i,k}+1}} - 1 \right)^{-1} \tag{5.2}$$

$$\gamma_{i,k} = \frac{P_{i,k}.G_{i,k}.S(\theta_0)}{P_n}, \quad P_{i,k} = \frac{P_{max}}{L_i.M} \tag{5.3}$$

where $P_{i,k}$ is the amount of power allocated to subcarrier k for user i, $G_{i,k}$ is the path gain for that subcarrier, $S(\theta_0)$ is a θ_0 degrees directional antenna gain, P_n is the noise power and P_{max} is the maximum transmission power. Assuming all users transmit at maximum power. L_i is the number of RBs assigned to user i. M is the number of subcarriers in one RB and is the same for all RBs since it is assumed they all have the same size.

5.1.1 Sounding

Since the channel-dependent scheduling can only be applied to low-speed UEs, usually the localized RB is assigned to transmit the channel-dependent scheduling traffic. In the case of localized data transmission the reference signal is also localized which means that the reference signal occupies the same spectrum as data transmission in two short blocks. Only one sounding pilot is required for each UE in each RB. Therefore in each localized RB, multiple uplink sounding channels can be supported for UEs that are not transmitting in the current RB and the current subframe. On the other hand a sounding pilot should be transmitted in every RB in order for the base station to sound the channel over the whole transmission bandwidth for each UE. Although this limits the number of sounding UEs in each sub-frame, more uplink sounding channels can be obtained by TDM because each low-speed UE can perform uplink channel sounding over multiple sub-frames (3GPP RP-060907, 2006). Fig. 5.1 gives an example of the FDM multiplexing scheme of the UE dedicated pilots and the sounding pilots. In this example 12 uplink sounding channels can be supported which provide the whole-band channel information for 12 channel-dependent scheduling UEs in each subframe.

5.2 Localized Gradient Algorithm LGA

The gradient algorithm is considered as the scheduling metric for the scheme. Referring to Stolyar's framework presented in Section 2.1.2,

$$k(t) = \arg \max_{k \in K(m)} \nabla U(\bar{X}(t))^T . \mu^m(k)$$

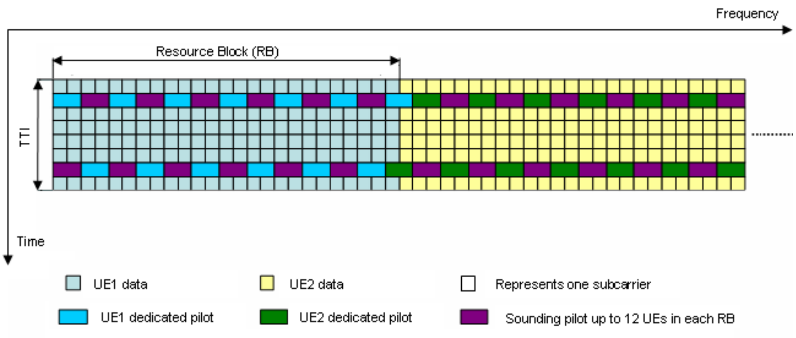

Figure 5.1: Sounding for the uplink channel in LTE

where $K(m)$ in this case will denote all feasible RB assignments that can be made with state m at time instant t. The set is confined by the channel capacity as well as the constraints that we have on the allocation of the RBs. In what follows, an integer programming assignment problem is formulated for solving $k(t)$ under the constraint that all RBs assigned to a single UE must be consecutive in the frequency domain. The integer programming solution is then used as a reference to validate the performance of the suggested heuristics to be introduced in the next section.

Let $y_{i,n}$ denote a selection variable: $y_{i,n} = 1$ if RB n is assigned to user i; otherwise $y_{i,n} = 0$. It is assumed that a mobile divides its available power evenly among the assigned RBs. Based on the channel sounding, the scheduler forms an estimate of the rate $\mu_{i,n}$ that user i expects to obtain if RB n is assigned to it. Given the estimated throughput $\bar{x}_i$, the scheduler needs to solve the following assignment problem.

$$y(t) = \arg\max_{y} \sum_{i=1}^{N} \sum_{n=1}^{C} U_i(\bar{x}_i)\mu_{i,n} y_{in} \tag{5.4}$$

subject to

$$\begin{aligned}
& y_{i,n} \in \{0, 1\} \\
& \sum_{i=1}^{N} y_{i,n} \leq 1, \quad i = 1, 2, \cdots, N \\
& y_{i,n} - y_{i,(n+1)} + y_{i,m} \leq 1, \quad m = n+2, n+3, \cdots, C
\end{aligned}$$

Where N is the total number of users. C is the total number of RBs. We could see that the first inequality limits the RB to one user only. The second inequality enforces the requirement for consecutive blocks. If $y_{i,n} = 1$ and $y_{i,(n+1)} = 0$, then $y_{i,m} \leq 0$ for $m > n+1$. If on the other hand both $y_{in} = 1$ and $y_{i,(n+1)} = 1$, the inequality requires that $y_{i,m} \leq 1$. If $y_{i,n} = 0$ then the inequality states that $y_{i,m} \leq -\big(1-(1-y_{i,(n+1)})\big)$. That is, the inequality becomes redundant.

The gradient scheduler discussed above is optimal for perfect channel state information. In case of measurement delays and estimation errors, the selection rule occasionally picks rates that do not match the channel capacity. It is assumed that synchronous non-adaptive HARQ mentioned in Section 3.3 is utilized to take care of the errors. Now the scheduler has to reserve those RBs to the UE that has scheduled retransmissions. To take this into account in our integer programming problem, we need to add a constraint

$$y_{i,n} = 1, \quad \text{if user } i \text{ has an ARQ process on RB } n \tag{5.5}$$

It is worth noting, that the integer programming approach presented here does not provide the optimal solution in case of imperfect channel estimates. However, it is expected still to provide a close to optimal solution to be used as a reference. To validate this claim, we note that the performance loss due to retransmissions is low as will be shown in Section 5.4.

5.3 Heuristic Localized Gradient Algorithm HLGA

The localized gradient algorithm described in the previous section requires that the scheduler solves an integer programming problem for every TTI. As the number of users and available resource blocks grow, the computational complexity and time required to solve the problem soon becomes infeasible. Hence, there is a need for simpler algorithms that can find good enough solutions very fast. In this section a scheduling algorithm is suggested that would follow a simple heuristics in allocating the resource blocks to the users while maintaining the required allocation constraint and taking retransmission requests into consideration.

Let J_i denote the set of RBs assigned to user i. F_i denotes the set of RBs that could be allocated to user i, (i.e. the RBs that do not violate the localization constraint if assigned to user i). Initialize by defining J_i and F_i for all i and t.

$$\begin{aligned} J_i^{(0)} &= \{\phi\} \\ F_i^{(0)} &= \{RB_1, RB_2, \cdots, C\} \end{aligned}$$

Where C is the total number of RBs in the frequency band.

Step 1: Iterate by finding the user-RB pair that has the maximum value

$$(i^*, j^*) = \arg \max_{(i,j), j \in F_i^{(k)}, i} \nabla U_i(\bar{x}_i(t)) \mu_{i,j}(t)$$

Step 2: Assign RB j^* to user i^* and update F_i.

$$\begin{aligned} J_{i^*}^{(k+1)} &= J_{i^*}^{(k)} \cup j^* \\ F_i^{(k+1)} &= F_i^{(k)} \setminus \mathcal{N}^{(k)}, \quad i \neq i^* \end{aligned}$$

$\mathcal{N}^{(k)} = \{n : n \geq J_{i^*}^{(k+1)}\}$ for users who have been assigned RBs located before $J_{i^*}^{(k+1)}$ and $\mathcal{N}^{(k)} = \{n : n \leq J_{i^*}^{(k+1)}\}$ for users with RBs located after $J_{i^*}^{(k+1)}$.

Step 3: If user i^* is assigned an RB that is not consecutive to the previously assigned RB(s) then all RBs in between in between will be allocated to that user since assigning any of these RBs to any other user will breach the localization for user i^*.

$$\begin{aligned} J_{i^*}^{(k+1)} &= J_{i^*}^{(k)} \cup \tilde{J}_{i^*}^{(k)} \\ \tilde{J}_{i^*}^{(k)} &= \begin{cases} \{j : J_{i^*}^{(k)} < j \leq j^*\}, & j^* > J_{i^*}^{(k)} \\ \{j : J_{i^*}^{(k)} > j \geq j^*\}, & j^* < J_{i^*}^{(k)} \end{cases} \end{aligned}$$

Update F_i in the same way as in step 3.

Step 4 Repeat the previous steps until all RBs are assigned.

Step 5 If a user has failed transmissions on certain RBs, then these RBs plus any blocks located in between two non-consecutive ARQs will be reserved for retransmission.

$$J_r(t+\tau) \quad = \quad \{RB_{ARQ}^{(1)}, \cdots, RB_{ARQ}^{(a)}\}, \quad r \in R$$

Where $RB_{ARQ}^{(1)}$ represents the block with the lowest order that has an ARQ process and $RB_{ARQ}^{(a)}$ is the ARQ block with the highest order. R is the set of users that have ARQ processes. τ is a fixed predefined time. Iterating for F_i with $F_i(t+\tau)^{(0)} = \{RB_1, RB_2, \cdots, C\}$, we have

$$F_i^{(y+1)}(t+\tau) \quad = \quad F_i^{(y)}(t+\tau) \setminus J_r(t+\tau), \quad i \neq r$$

Numerical Example

For a better understanding of the heuristics this simple example is presented. Assume a system with 3 users and 6 RBs. Assume perfect channel estimation. The selection metric forms an $i \times j$ matrix that has the following values for time-slot t.

$$\nabla U_i(\bar{x}_i(t))\mu_{i,j}(t) =$$

$$\begin{pmatrix} 0.26 & 1.65 & 0.10 & 1.60 & 0.85 & 0.88 \\ 0.82 & 0.50 & 0.30 & 0.90 & 0.63 & 0.87 \\ 0.41 & 0.39 & 0.47 & 0.62 & 0.89 & 0.59 \end{pmatrix}$$

The scheduler will start by allocating RB2 to UE1 since it forms the highest value in the matrix 1.65 and naturally any RB that is allocated to a user will be excluded for all other users. The next highest value is 1.60 with UE1 on RB4. UE1 is allocated RBs 4 and 3 due to the fact that RB3 will fall between two RBs that belong to the same user (UE1) and to maintain localization it cannot be allocated to any other user. Next is the value 0.89 with UE3 on RB5 leading to the exclusion of RB1 from the set of possible RBs for UE3. Following that is 0.87 with UE2 on RB6 excluding RB1 for UE2. Finally RB1 is allocated to UE1 since there is no possibility to grant it to any other user due to localization. The RB allocation will have the final form:

RB index	1	2	3	4	5	6
UE	1	1	1	1	3	2

5.4 Numerical results

A computer simulator is used to create a single-cell environment. The simulator generates N users with locations uniformly distributed over the cell area. Assuming pedestrian profiles for the users, hence channel conditions are slowly changing. Different users experience different channel conditions that vary depending on their distance from the base-station and speed. The speeds of the mobile users were independent random variables uniformly distributed between 3 km/h and 10 km/h. All users are assumed to have full buffers with different packet sizes. System parameters are shown in Table 5.1. The proportional fair rule is used as the metric for selecting the RBs in every TTI for LGA and HLGA. Retransmissions are included in the scheduling process and are prioritized. The scheme is compared against the solution provided by the LGA as well as the solution from a blind scheduler that assigns all RBs to one user at a time in a round-robin (RR) fashion. Fig. 5.2 shows the cell throughput for cases of perfect and imperfect channel estimation. The figure is normalized to the performance of a restriction-free, retransmission-free case where there is no constraint in the block allocation and channel estimation is assumed to be perfect. The scheduling used in this reference case is simply the original non-localized version of the gradient algorithm, non localized gradient algorithm (NLGA). We notice that there is only a 4% gap between the LGA optimal solution and the NLGA optimal solution. The gap represents the impact of the localization requirement. This implies that the localization constraint has a low impairment on the performance of the LGA. It could also be seen in the figure that the HLGA provides a close to optimal performance when compared to the NLGA and LGA optimal solutions.

Table 5.1: System parameters for uplink scheduling in LTE

Parameter	Value
Carrier frequency	2 GHz
RB bandwidth	375 kHz
Total number of blocks	10 (25 subcarrier/RB)
TTI duration	1 ms
Fading model	One path Rayleigh
No. of terminals	5 (Single cell)
Site to site distance	1000 m
Number of Tx antennas	1
Max. Tx power	21 dBm
Noise power	-108.5 dBm

Table 5.2: Cell throughput values in Mbps for perfect and imperfect channel estimation

	NLGA	LGA	HLGA	RR
Perfect	26.78	25.64	23.05	12.69
Imperfect	-	24.08	21.29	9.83

For the imperfect channel information case we see that the LGA and HLGA still perform well with retransmissions now associated in the block scheduling decision. The gap between the LGA solution which is now a sub-optimal solution and the NLGA optimal solution grows to 10%. Thus, the impact of the retransmissions on the performance of the LGA was only 5%. The HLGA also maintains a good position with a drop in performance of only 7%. Table 5.2 shows the exact throughput values for both cases.

5.5 Concluding Remarks

The constraint in resource allocation for the uplink channel in LTE systems makes channel dependent scheduling a challenging task due to the fact that some resources have to be allocated to satisfy the constraint rather than the channel condition. Finding the exact optimal solution is effort requiring. Therefore, a heuristic approach to the scheduling problem would be a suitable choice for finding a practical solution to the allocation optimization problem. For this purpose, the HLGA was suggested

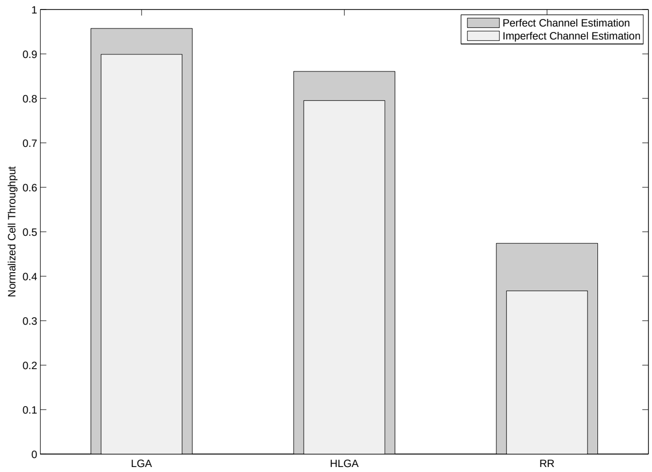

Figure 5.2: Normalized cell throughput with perfect and imperfect channel estimation

and provided a good performance when compared to the optimal solution provided by the LGA. The LGA on the other hand, was considered realizable only in theory due to its complexity. The algorithm could therefore present a good benchmark to measure the performance of other channel dependent scheduling rules for localized SC-FDMA.

Chapter 6

Feedback in Multi-Carrier systems

Feedback information is an essential part of channel adaptive radio resource management. In this chapter feedback in multi-carrier systems is investigated looking through different decisions in the selection of the type of feedback information. In OFDMA (Fazel and Fettewis 1997) networks, the bandwidth is divided into many narrowband subchannels. The task of the resource scheduler is to divide the transmitter power among the different channels and the channels among the different users. Different frequency bands experience different fading, so the power allocation can be opportunistic by allocating more power to good subchannels. This technique is known as water filling. In OFDMA and MC-CDMA (multi-carrier code division multiple access) the transmitter utilizes inverse fast Fourier transform (IFFT) followed by digital to analogue conversion. Since the different subchannels are formed using digital signal processing, it is possible to dynamically control the utilized spectrum. If the channel is static (e.g. in digital subscribers lines (DSL)) or slowly time varying, the receiver can provide the transmitter with detailed CSI using a robust feedback channel. Based on the CSI, more sophisticated adaptive transmission techniques have the possibility to dynamically modify the parameters of the modulator in order to improve performance (Keller and Hanzo 2000). Thanks to the characteristic of multi-carrier modulation, it is also possible to dynamically change the transmitting power and bit rate of each subchannel according to channel selectivity variations (adaptive bit loading). Recently, some studies regarding the application of adaptive bit loading algorithms to wireless channels appeared (Czylwik 1996), (Perre et al. 1998), (Barreto and Furrer 2001) and (Dardari 2004). Adaptive OFDM has been considered for the 3G LTE (3GPP RP-040461). To reduce the amount of overhead resulting from reporting the CSI of each subcarrier, Cherriman *et al.* (2002) suggested to group the subcarriers into subbands and report one CSI for each subband. In this work we further group the subbands into blocks and form a joint CSI based on the state of a particular subband that would maximize the total throughput. Rank ordering is used to find this subband. Rank ordering statistics is deemed very useful in computing the expected throughput for a block of subbands with different state values.

6.1 System model

Consider the downlink of a multi-carrier system with M subcarriers and total bandwidth W. Let Δf_c denote the coherence bandwidth of the channel. We assume that the subcarriers are grouped into $N = \frac{W}{\Delta f_c}$ subbands having $\frac{M}{N}$ subcarriers each. The base station allocates power equally among the different subcarriers and subbands. The channel is assumed to be slow fading such that the channel state stays essentially constant during one TTI. That is, the coherence time of the channel is assumed to be longer than the duration of the TTI and thus the channel exhibits block fading characteristics. The subbands fade independently, but the fading seen by individual subcarriers in a subband is approximately the same, since the subcarrier spacing is small compared to the coherence bandwidth of the channel. Assume that the channel is subject to Rayleigh fading. In case CSI is known at the transmitter, the capacity of subband i at TTI k is given by

$$C_i(k) = \frac{W}{N} \log_2 \left(1 + \gamma |h_i(k)|^2\right) \tag{6.1}$$

where γ denotes the mean SINR over time and frequency and $h_i(k)$ is the channel's transfer function. $h_i(k)$ is a circular symmetric normal (CN) distributed random variable with zero mean and unit variance.

Assume that the number of transport formats (modulation and coding schemes) is large so that the corresponding data rates can be approximated with a real number μ. In case of perfect CSI, the rate assigned to subband i is matched to the channel state $\mu_i(k) = C_i(k)$. This would require timely knowledge of the channel state of all the subbands at the transmitter. In practice, we would need to select the rate based on possibly outdated and partial CSI. If the selected rate exceeds the instantaneous subband capacity $\mu_i(k) > C_i(k)$, then the transmitted data cannot be decoded at the receiver. In that case, the code block needs to be retransmitted. In case, chase combining HARQ (Sanayei et al. 1985) is utilized, the receiver coherently combines the original code block and the retransmitted block. Assume that the transport format was selected to match rate $\mu = \frac{W}{N}\log_2(1+\gamma Z)$ at TTI k_0. The variable Z denotes the channel state information available to the scheduler. If $\mu(k_0) > C(k_0)$, retransmission occurs. If the receiver is able to decode the code block by combining the original and the retransmitted packet, the actual rate would become $\frac{\mu(k_0)}{2}$. Decoding fails if this rate still exceeds the capacity of the channel. The probability that the packet has to be retransmitted at least d times is given by

$$\begin{aligned}\Pr\{D_i \geq d\} = \ \Pr \ & \left\{\log_2\left(1+\gamma\sum_{k=k_0}^{k_0+d}|h_i(k)|^2\right)\right. \\ & \leq \log_2(1+\gamma Z)\} \end{aligned} \tag{6.2}$$

where k_0 refers to the TTI when the code block was first transmitted.

We note that $\xi_i(k) = |h_i(k)|^2$ is an exponentially distributed random variable with a probability distribution function

$$f_\xi(x) = e^{-x} \tag{6.3}$$

and cumulative probability density function

$$F_\xi(x) = \Pr\{\xi_i(k) < x\} = 1 - e^{-x} \tag{6.4}$$

Now (6.2) can be rewritten as

$$\Pr\{D_i \leq d\} = \Pr\left\{\sum_{k=k_0}^{k_0+d} \xi_i(k) \geq Z\right\}, \tag{6.5}$$

If the coherence time of the channel is small, then $\{\xi_i(k),\, k = k_0, k_0+1, k_0+2, \cdots\}$ become independent and identically distributed (i.i.d) random variables and the sum $\sum_{k=k_0}^{k_0+d} \xi_i(k)$ becomes Erlang-$(d+1)$ distributed. It follows that the delay (in terms of number of retransmissions) becomes Poisson distributed conditioned on the CSI value Z.

$$\Pr\{D_i \leq d\} = \sum_{k=0}^{d} \frac{(Z)^k}{k!} e^{-Z} \tag{6.6}$$

Hence we get

$$\Pr\{D_i = d\} = \frac{(Z)^d}{d!} e^{-Z}, \quad d = 1, 2, 3, \cdots \tag{6.7}$$

The throughput T_i is thus proportional to $\frac{\mu(Z)}{D_i+1}$. Still conditioning on Z, we can find the expected throughput as follows

$$\begin{aligned} \mathrm{E}\{T_i|Z\} &= \sum_{d=0}^{\infty} \frac{\mu(Z)}{d+1} \frac{Z^d}{d!} e^{-Z} && (6.8)\\ &= \frac{\mu(Z)}{Z}(1 - e^{-Z}) && (6.9)\end{aligned}$$

Consider a blind system, in which no CSI is utilized. In that case, Z can be considered as a constant. In the low SINR region ($\gamma \to 0$), we have $\mu(Z) = \frac{W}{N}(\log_2(1+\gamma Z) \approx \frac{W}{N}\gamma Z$. Hence we have $\mathrm{E}\{T_i|Z\} \propto (1 - e^{-Z})$ which suggests that Z should be large to maximize the throughput. That is, the system should rely on the ARQ process to achieve high throughput.

If the coherence time of the channel is very long, the channel gain could be assumed to be constant, hence

$$\Pr\{D_i \leq d\} = \Pr\left\{\frac{Z}{\xi_i} < d\right\} \tag{6.10}$$

$$= \Pr\left\{\xi_i > \frac{Z}{d}\right\} = 1 - F_\xi\left(\frac{Z}{d}\right) \tag{6.11}$$

This distribution $F_\xi\left(\frac{Z}{d}\right)$ is known as the *inverted exponential distribution* and it belongs to the class of heavy tailed distributions. The probability density function of d can be written as

$$f_D(y) = \frac{1}{y^2} e^{-\frac{Z}{y}} \tag{6.12}$$

The throughput conditioned on Z becomes

$$\mathrm{E}\{T_i|Z\} = \int_0^\infty \frac{\mu(Z)}{y+1} \frac{1}{y^2} e^{-\frac{Z}{y}} dy \tag{6.13}$$

$$= \frac{\mu(Z)}{Z}\left(1 - Ze^Z E_1(Z)\right) \tag{6.14}$$

where $E_1(x) = \int_1^\infty t^{-1}e^{-xt}dt$ denotes the exponential integral. For large values of Z, $E_1(x) \sim \frac{e^{-Z}}{Z}$ and we get $\mathrm{E}\{T_i|Z\} \to 0$. Thus in the case of very slow fading, if a fading dip occurs, it lasts for a long time and the number of retransmissions can become very large. On the other hand, if Z is very small, then retransmissions can be avoided. The drawback is that also $\mu(Z)$ is going to be very small as well. Compared to the very fast fading case discussed earlier, the conclusion here is the contrary. In very slow fading, one should try to avoid retransmission as the retransmission delays are expected to be large.

Now assume that the number of available feedback bits B for a number of subbands N is small. If $N \leq B$, we still could use individual feedback for each subband with $\frac{B}{N}$ bits per subband. However, if $N > B$, this is not possible, since there is less than 1 feedback bit per subband. If B is large, say 8 to 16, then joint feedback information could be approximated with a real number.

One way to compose the joint feedback information for a block of subbands is to use the average value $Z = \frac{1}{N}\sum_{j=1}\xi_j$. This variable is correlated with ξ_i, but for large N we can approximate the throughput simply by noting that the law of large numbers dictates that Z approaches the mean value (in this case 1 since $\xi_j = |h_j|^2$ and $|h_j|$ is a circular symmetric normal distributed random variable with zero mean and unit variance), so $Z \to 1$. Hence, we have $T_i \sim \mu(1)(1-e^{-1}) \approx 0.6321 r(1)$ in the fast fading case and $T_i \sim \mu(1)\left(1 - e^1 E_1(1)\right) \approx 0.4037 r(1)$ in slow fading. Another way to compose the joint feedback would be rank ordering described in section 6.2.

6.2 Feedback based on rank ordering

Let $Z_{N,n}$ denote the n^{th} smallest subband state value in a block containing N subbands $\{\xi_i(k_0), i = 1, 2, \cdots, N\}$; where $\xi_i(k_0)$ denotes the state value of subband i at time k_0. If $Z_{N,n}$ is utilized to select the utilized data rate, then $n-1$ subbands have to use retransmission while the rest can transmit the packet directly. The probability density function of $Z_{N,n}$ in the case of exponentially distributed random variables is given by

$$f_{Z_{N,n}}(x) = \frac{N!}{(N-n)!} \sum_{k=0}^{n-1} \frac{1}{(n-1-k)!k!} e^{-(k+n)x} \tag{6.15}$$

For the selected $Z = Z_{N,n}$ channel feedback information, the expected throughput becomes

$$\begin{aligned}
\mathrm{E}\{T_i\} &= \int_0^\infty \frac{\mu(Z)}{Z} F_\xi(Z) f_{Z_{N,n}}(x)dx & (6.16)\\
&= \int_0^\infty \frac{\mu(Z)}{Z} (1 - (1 - F_\xi(Z))) f_{Z_{N,n}}(x)dx \\
&= \int_0^\infty \frac{\mu(Z)}{Z} (f_{Z_{N,n}}(x) - \frac{N-n+1}{N+1} f_{Z_{N+1,n}}(x))dx \\
&= \int_0^\infty \frac{\mu(Z)}{Z} (f_{Z_{N,n}}(x) - (1 - \frac{n}{N+1}) f_{Z_{N+1,n}}(x))dx
\end{aligned}$$

(6.17)

Table 6.1: System parameters for a multi-carrier system

Parameter	Value
Carrier frequency	2 GHz
Total Bandwidth	100 MHz
Coherence Bandwidth	333 kHz
Spreading factor	16
Number of multicodes	10
TTI duration	2 ms
Fading model	One path Rayleigh (Jake's model)
Radio propagation	Path loss component 3.52
	Std. of shadow fading 8 dB
BS Tx power	16 W
	80% of total cell transmit power
Hybrid ARQ	Chase combining
Total no. of subcarriers	512
OFDM symbol period	4μs
Subcarrier spacing	10 kHz
Subband spacing	320 kHz
Subcarriers/Subband	32

Consider a low SINR region ($\gamma \rightarrow 0$). In that case, we have $\mu(\gamma Z) \propto Z$. Now the term $\frac{\mu(Z)}{Z}$ becomes a constant, and we can derive a closed form solution for the throughput

$$\mathrm{E}\{T_i\} \propto \frac{n}{N+1} \tag{6.18}$$

Hence, throughput is maximized with $n = N$ which corresponds to SINR of the best subband. That is, $Z = \max_i\{\xi_i\}$ maximizes the throughput. In a high SINR domain, the term $\frac{\mu(\gamma)}{Z}$ cannot be ignored and (6.16) must be solved numerically with the help of a computer simulator as illustrated in section 7.6.

6.3 Numerical analysis

In this section further analysis will be carried out via simulations.

6.3.1 Rank ordering based decision variable

Considering a single user case. The bandwidth allocated to the user is W which is divided using the inverse fast Fourier transform into multiple orthogonal subcarriers

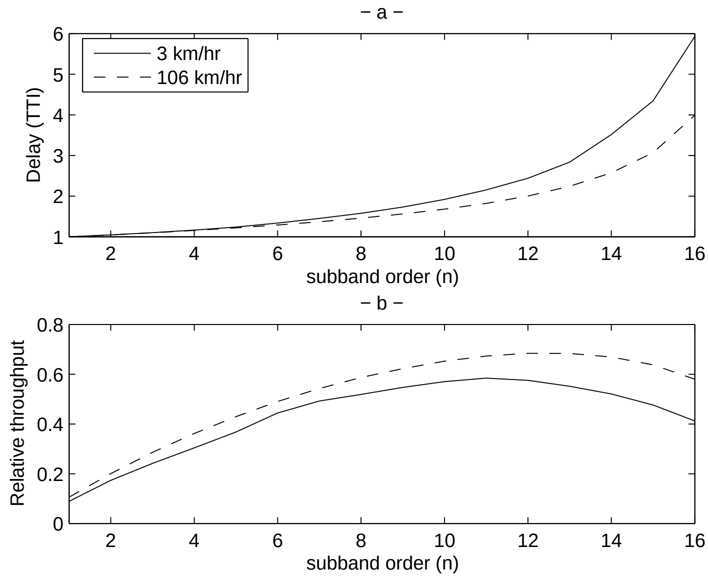

Figure 6.1: a) Delay and b) relative throughput as functions of the decision variable for different speeds (mean SINR=3 dB)

with equal spacing. The subcarriers are grouped into subbands with a bandwidth less than the coherence bandwidth. Table 6.1 shows the parameters used in the simulations. The system employs orthogonal frequency and code division multiplexing (OFCDM). The system resembles WCDMA-HSDPA but with OFDM in the radio interface (3GPP RP-25.892 2004). Adaptive Modulation and Coding (AMC) is used in a transmission time interval based on the CSI report. The TTI is short enough that we can assume the channel is constant during that time to perform the necessary rate-SINR mapping.

The simulations will characterize parameters that affect the selection decision for the feedback information such as mobile speed and mean SINR. Fig. 6.1 represents the delay (i.e. average number of transmissions) per block and the relative throughput (relative to the throughput when full channel knowledge is known i.e. the transmitter has knowledge of the state of all the subbands) for different decision variables. The variable n represents the rank order of a subband i.e. $n = 1$ is the subband with the lowest channel state and $n = 16$ represents the highest. We note that the retransmission delay associated with high speeds is relatively small due to the fact that the channel coherence time for high speed mobiles is much shorter than low speed mobiles. This leads to the possibility of using the state of higher order subbands for the joint CSI. This is illustrated in Fig. 6.1-b where it could be seen that the relative throughput increases at orders higher than in the case of low mobility.

6.3.2 Comparison of different decision variables

In this section different decision variables for the joint feedback information of one block are compared with each other. In section 6.1, we saw that one way to compose the common feedback is to use the average value $Z = \frac{1}{N}\sum_{j=1}\xi_j$. In this section we will see the impact of different decision variables on the expected throughput and determine the best decisions in different SINR regions for low and high mobility cases. Figures 6.2 and 6.3 show the total throughput with different decision variables for speeds of 3 km/hr and 106 km/hr respectively. The decision variables considered were the minimum, median and maximum subbands as well as the block average. The throughput obtained with the optimal subband (the subband order that maximizes the throughput) was also included for the sake of comparison. In a low mobility case, Fig. 6.2, the median is a good choice in most regions except for high SINR regions where the minimum outperforms all other decisions. On the otherhand we can see from Fig. 6.3 that in a high mobility case, it is more convenient to depend on the maximum subband as the feedback decision in low SINR regions. In practical regions, the median and average become better choices giving almost similar results. At extremely high SINRs the minimum as in low mobility becomes the best decision. However, overall the median value based feedback provides adequate throughput.

6.3.3 Multiple Blocks

In this section we will study the effect of having multiple CSI feedback channels. We group the subbands into multiple blocks instead of only one block considered

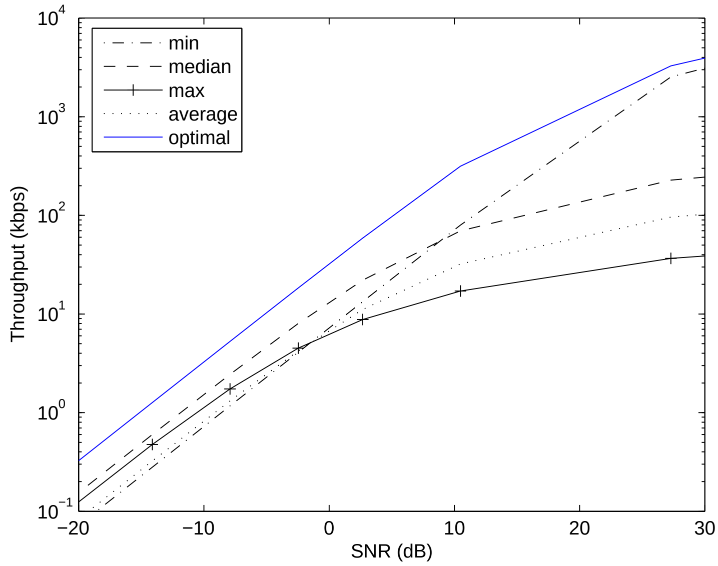

Figure 6.2: Total throughput for different decision variables (speed=3 km/hr)

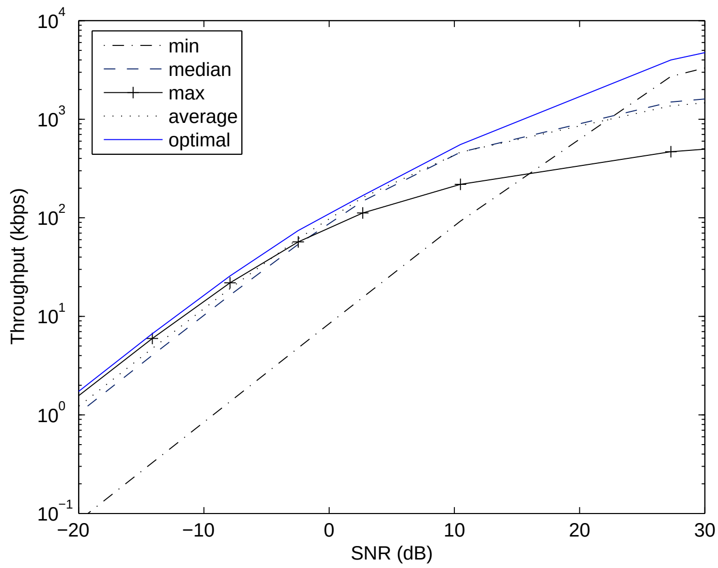

Figure 6.3: Total throughput for different decision variables (speed=106 km/hr)

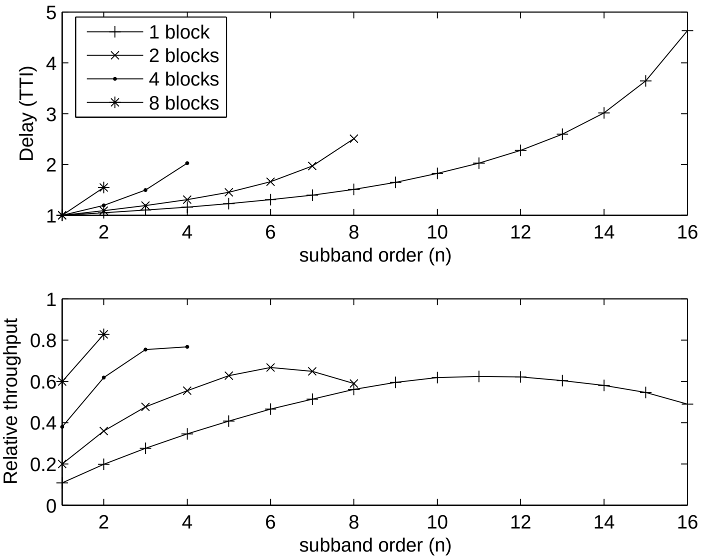

Figure 6.4: Relative throughput for different block sizes

earlier. We are interested in seeing the impact of using multiple blocks each having it's own ARQ process on the total throughput. We assume that each block will have the same number of feedback bits for its feedback channel. We also assume that the number of quantization levels for the AMC obtained with this number is large (e.g. 16 bits → 65536 levels).

Figure 6.4 shows that aggregating the throughput of multiple blocks of smaller sizes results in a higher relative throughput. This is mainly due to the fact that increasing the number of blocks increases the diversity gain.

6.3.4 Effect of number of feedback bits

This section studies the effect of the number of feedback bits allocated to the subbands and compare the throughput obtained with joint feedback and that with indiviual feedback. Let's assume that the total number of feedback bits allocated to a number of subbands N is B. There are two scenarios for utilizing these bits. 1) Each subband could individually have its own feedback channel with a capacity of B/N bits giving $2^{B/N}$ quantization levels. 2) Group the subbands into L blocks of N/L subbands with one feedback channel per block. The number of feedback

Table 6.2: Different bit allocation ($B = 32, N = 8$)

Number of blocks	Feedback bits per block	Quantization levels	Total throughput (Mbps)
1	32	4.2950e+009	1.3007
2	16	65536	1.7373
4	8	256	1.8500
8	4	16	2.2049

bits for this channel will be the sum of the feedback bits of the individual subbands within the block yielding $B/N * N/L = B/L$ feedback bits and consequently $2^{B/L}$ levels ($2^{B/L} > 2^{B/N}$). Rank ordering is then performed for the block to find the subband state that maximizes the total throughput and report that state back to the transmitter using the B/L bits. For small bit feedback channels we use a coarse quantization method (Kim 2006) that finds the optimal quantization levels based on the mean SNR assuming that there is no delay in the feedback. That is, the quantization levels $\{\gamma_l, l = 0, 1, 2, \cdots Q - 1\}$, where l is the level number and Q is the total number of quantization levels are determined by solving the following optimization problem

$$\max_{\gamma_l} \sum_{l=0}^{Q-1} \left(F_\xi \left(\frac{\gamma_{l+1}}{\gamma} \right) - F_\xi \left(\frac{\gamma_l}{\gamma} \right) \right) \log_2(1 + \gamma_l) \tag{6.19}$$

s.t. $\gamma_l \geq 0$, $\gamma_{-1} = 0$ and $\gamma_Q = \infty$.

Two cases are considered. The first is for a large number of bits per subband ($\frac{B}{N}$) and the second is with a smaller one. $B = 32$ and 8 respectively, $N = 8$ subbands, $L = 1, 2, 4$ and 8 blocks. Tables 6.2 and 6.3 illustrate the total throughput with different feedback channel capacities. Results in Table 6.2 indicate that the throughput loss from individual feedbacks using coarse quantization (last row in the table) is less severe than the loss due to joint feedback. However, joint feedback is still needed in case the number of bits is equal or less than the number of subbands as shown in Table 6.3.

Table 6.3: Different bit allocation ($B = 8, N = 8$)

Number of blocks	Feedback bits per block	Quantization levels	Total throughput (Mbps)
1	8	256	1.4748
2	4	16	1.5769
4	2	4	1.7569
8	1	2	1.4113

6.4 Concluding remarks

Rank ordering is a useful method for partial CSI systems when chase combining ARQ is utilized. The n^{th} order statistic directly implies that $\frac{n-1}{N}$ fraction of the time retransmissions are needed. In case of linear SINR-rate mapping, using the highest received SINR as a feedback would be optimal. However, in case the SINR-rate mapping has a logarithmic shape, some $n < N$ order statistic would be needed to maximize the channel throughput. The results indicate that average value and median based channel feedback works relatively well for all values of SINR. It was found that dividing the subbands into multiple small sized blocks would further increase the aggregate throughput. Rank ordering is a good choice when the number of feedback bits allocated to a single subband is small. It was found that grouping the subbands and using the sum of their feedback channel bits for the joint feedback outperforms the case when each subband used an individual feedback channel when the number of total bits was less or equal to the number of subbands.

Chapter 7

Admission Control

The call admission control problem can be formulated as an optimization problem, i.e., maximize the network efficiency/utility/revenue subject to the QoS constraints of connections. The QoS constraints could be signal-to-interference ratio (SIR), the ratio of bit energy to interference density E_b/I_0, bit error rate (BER), call dropping probability, or connection-level QoS (such as a data rate, delay bound, and delay-bound violation probability triplet). For example, a CAC problem can be maximizing the number of users admitted or minimizing the blocking probability.

The performance of opportunistic schedulers depends on the number of users. Although it was shown in (Berggren and Jäntti 2004) that the gain of using channel adaptive scheduling compared to round robin scheduling increases as the number of users increases, the throughput per user decreases rapidly as the load increases. Under ideal conditions (i.e. all users have the same channel distribution and channel statistics are assumed to be stationary, for example all the users have Rayleigh or Ricean distributed channels) the rate obtained using PF scheduling is independent of the rates of the other users and only depends on the mean channel gain of the user and the total number of users. However, this functional relation can vary from one user to another and depends on parameters such as quantization levels of the rates and measurement delays, see (Berggren and Jäntti 2004). In addition, impairments such as retransmissions can cause dependencies among the users' rates if priority is given to retransmitted packets. This means that a simple number based admission control that makes the admission decision directly based on the number of users might not perform adequately.

The importance of combining opportunistic scheduling with admission control has been recognized in many papers. However, surprisingly little results have been published on the matter so far. In (Liu et al. 2001a) the system was probed using a simple iterative admission control scheme in which some weight in the decision rule was changed to find the feasible region. If the minimum requested rates for the active users were inside the feasible region then the new user was admitted, otherwise rejected. However, during the probing process, the QoS of the active users could not be guaranteed. In this chapter, a new admission control scheme is proposed, that allows the operator to limit the maximum impact that the new user can cause to the active users in terms of throughput loss. The scheme resembles the sliding window based call admission control scheme suggested in (Zhao and Zhang 2002) and can be interpreted as a modification of the active link protection scheme (Bambos, Chen and Pottie 2000) to the multiuser diversity channel.

7.1 System model

Consider the downlink of a DS-CDMA time-slotted system where time is the resource to be shared among all users. It is first assumed that the channel is stationary, more specifically it is assumed that the fast fading process is stationary and ergodic so that the time average approaches the mean value as the number of samples increases. The fading distributions of the users could be different but their channel should be stationary. Later this assumption is relaxed in Section 7.4 where parameter estimation is discussed and consider the case, where the channel statistics are slowly changing. The mobile is assumed to estimate the channel based on a pilot signal transmitted by the base-station. Based on the channel measurement the mobile then determines the maximum service rate $\mu_i(k)$ that it could achieve under the current channel condition in time-slot k, and reports that back to the base-station which then utilizes this information for the scheduling decision.

Let $\mathcal{A}(k)$ denote the set of active users having pending data in the transmission buffer at the base-station. Consider an opportunistic scheduler that allocates the channel to the users based on the maximum instantaneous service rate $\mu_i(k), i \in \mathcal{A}(k)$ and the estimate of the long term average throughput $\bar{x}_i(k)$ using a simple selection rule of the form:

$$i^*(k) \in \text{argmax}_{i \in \mathcal{A}(k)} f(\mu_i(k), \bar{x}_i(k)) \tag{7.1}$$

Where $f(\mu_i(k), \bar{x}_i(k))$ is some non-decreasing function of $\mu_i(k)$ and $\bar{x}_i(k)$ is updated as follows:

$$\bar{x}_i(k+1) = (1-\alpha)\bar{x}_i(k) + \alpha\mu_i(k)\chi(i = i^*(k)) \tag{7.2}$$

Where $\alpha > 0$ is a fixed (small) parameter. The operator $\chi\{.\}$ is the indicator function of an event. We see from this definition that, $\bar{x}_i(k)$ represents an exponentially smoothed average throughput. The initial value of the estimator is $\bar{x}_i(0) = \mu_i(0)\chi(i = i^*(0))$. Taking the expectation of eq. (7.2) yields:

$$E\{\bar{x}_i(k+1)\} = (1-\alpha)E\{\bar{x}_i(k)\} + \alpha E\{\mu_i(k)\chi(i = i^*(k))\} \tag{7.3}$$

The scheduler is very general and contains all the memoryless scheduling rules suggested by Park *et al.* (2003) and gradient scheduling rules suggested by Shakkottai and Stolyar (2000) as special cases. Thus, we define two types of schedulers:

7.1.1 Memoryless schedulers

A memoryless policy is a stationary policy whose decision does not depend on time-slot k but rather depends on a performance vector in that time-slot. Consider the saturated case in which all the active users have their transmission buffers full of data all the time. Assume the channel and data rate processes are wide-sense stationary and ergodic. It follows that the rate is also stationary as long as the selection process (7.1) which is a function of $\mu_i(k)$ and $\bar{x}_i(k)$ is stationary. Since $\mu_i(k)$ is assumed to be stationary, consequently the selection process is stationary as long as the estimate $\bar{x}_i(k)$ is stationary. Assuming that $\bar{x}_i(k) = \bar{x}_i, \forall k$ then $\tilde{x}_i(k) = \bar{x}_i, \forall k$. This allows the implementation of the admission control scheme in a slot-by-slot fashion.

7.1.2 Schedulers with memory

Schedulers with memory are dynamic schedulers whose policy depend on time-slot k. In our case the memory of the scheduler contains information of past throughput. If we choose,

$$f(\mu_i(k), \bar{x}_i(k)) = \nabla U_i(\bar{x}_i(k))\mu_i(k) \tag{7.4}$$

then scheduling rule (7.1) becomes equivalent to the gradient rule which is known to maximize the aggregate utility $\sum_i u_i(\bar{x}_i(k))$. The choice $U_i(\bar{x}_i(k)) = \log(\bar{x}_i(k))$ leads to the proportional fair scheduler. Let us define $\bar{x}_i(k) = E\{\mu_i(k)\chi(i = i^*(k))\}$ and $\tilde{x}_i(k) = E\{\bar{x}_i(k)\}$. For a memory scheduler, the rate in eq. (7.3) can be written as:

$$\begin{aligned}
\tilde{x}_i(1) &= (1-\alpha)\bar{x}_i(0) + \alpha\bar{x}_i(0) = \bar{x}_i(0) \\
\tilde{x}_i(2) &= (1-\alpha)\tilde{x}_i(1) + \alpha\bar{x}_i(1) = (1-\alpha)\bar{x}_i(0) + \alpha\bar{x}_i(1) \\
\tilde{x}_i(3) &= (1-\alpha)\tilde{x}_i(2) + \alpha\bar{x}_i(2) = (1-\alpha)^2\bar{x}_i(0) + (1-\alpha)\bar{x}_i(1) + \alpha\bar{x}_i(2) \\
&\vdots \\
\tilde{x}_i(k) &= \sum_{m=0}^{k-1}(1-\alpha)^m\alpha\bar{x}_i(k-1-m) \qquad (7.5)
\end{aligned}$$

If the initial value of the estimator is a steady state value, then $\tilde{x}_i(k) = \bar{x}_i, \forall k$ and we can apply the admission scheme in a slot-by-slot basis. However, if the initial rate was not a steady state value then for an ergodic channel $\tilde{x}(k)$ asymptotically will converge to $\bar{x}_i$ as k grows within a frame. In this case the admission control scheme could be carried out in a frame-by-frame basis.

7.2 Single-user iterative admission control

In this section the admission control scheme is described with its application to the two types of schedulers.

7.2.1 Memoryless schedulers

Assume that a new user tries to join the system at time-slot k_0. At time-slot $k > k_0$, the new user is excluded from the active set $\mathcal{A}$ with a back-off probability p and included with probability $1-p$. Usually a user is included in the active set $\mathcal{A}$ if there is data to transmit to that user. However, for a new user with this back-off probability, it will be excluded from the active set $\mathcal{A}$ even if there is data (which in the beginning will consist of probing packets only) for it in the transmission buffer. Once the new user is fully admitted, it will start receiving the original data designated for it. Let $\bar{x}_i(Z_N)$ denote the mean (expected) rate of user i with N users in the system and their channel statistics are defined by the matrix Z_N, where Z_N is defined as a matrix consisting of vectors describing the sufficient statistics of the channel. The number of initial users is N and user $N+1$ is the user trying to get admitted. Prior to the arrival of user $N+1$, the mean rates for the initial users were determined from channel statistics Z_N and given by $\tilde{x}_i(k) = \bar{x}_i = \bar{x}_i(Z_N), \forall k \leq k_0$. Based on our analysis of the mean rates, it is possible to observe the impact of the new user on the active users in the following manner:

$$\begin{aligned} \tilde{x}_i(k) &= p\bar{x}_i(Z_N) + (1-p)\bar{x}_i(Z_{N+1}) \quad i = 1, 2, \cdots N, \quad k \geq k_0 && (7.6) \\ \tilde{x}_{N+1}(k) &= (1-p)\bar{x}_{N+1}(Z_{N+1}) \quad k \geq k_0 && (7.7) \end{aligned}$$

And

$$\begin{aligned} \tilde{x}_i(k) &= \bar{x}_i(Z_N) \quad i = 1, 2, \cdots N, \quad k < k_0 && (7.8) \\ \tilde{x}_{N+1}(k) &= 0 \quad k < k_0 && (7.9) \end{aligned}$$

Hence,

$$\tilde{x}_i(k_0) - \tilde{x}_i(k_0 - 1) = -(1-p)(\bar{x}_i(Z_N) - \bar{x}_i(Z_{N+1})) \quad i = 1, 2, \cdots N \tag{7.10}$$

$$\tilde{x}_{N+1}(k_0) - \tilde{x}_{N+1}(k_0 - 1) = (1-p)\bar{x}_{N+1}(Z_{N+1}) \tag{7.11}$$

Thus, in a slot the impact of the new user is limited to the fraction $1-p$. This behavior resembles the active link protection (ALP) scheme suggested by Bambos e*t al.* (2000) for power controlled systems. Therefore, the scheme will be referred to as the ALP-CAC.

7.2.2 Schedulers with memory

The mean rates obtained with these schedulers experience transient states unlike in the memoryless schedulers. This property makes the admission control more difficult. Let us observe the system after m slots. Assuming m is large, so that the estimator will converge.

$$\tilde{x}_i(k_0 + m) \to p\bar{x}_i(Z_N) + (1-p)\bar{x}_i(Z_{N+1}) \tag{7.12}$$

Similarly,

$$\tilde{x}_i(k_0 + m) - \tilde{x}_i(k_0) = -(1-p)(\bar{x}_i(Z_N) - \bar{x}_i(Z_{N+1})) \geq -(1-p)\bar{x}_i(Z_N) \tag{7.13}$$

The rate has dropped no less than by a factor $1-p$. We decrease the back-off probability p exponentially and observe $\bar{x}_i(k)$. We will divide the frame into a frame of m slots and observe $\bar{x}_i(k)$ in the end of each frame. Hence, we change the back-off factor at a rate of $1/m$ slots. The length of the frame should ensure the convergence of the rate estimator. Due to the exponential decrease of the back-off probability at frame n, the utilized back-off probability becomes $p_n = \prod_{k=1}^{n} p = p^n$. The size of p limits the impact of the new user, hence we could reject the new user if $p^n \bar{x}_i(k_n) < \bar{x}_{min}$, where k_n denotes the last slot in frame n. The drop in rate from frame to frame will remain by a $(1-p)$ fraction as shown in the following analysis: Let us observe the mean rates of the users over the whole frame. During a frame, p_n is kept constant. At the first frame we use p, at the second we use p^2 and so on.

$$\tilde{x}_i(k_1) = p\bar{x}_i(Z_N) + (1-p)\bar{x}_i(Z_{N+1}) \tag{7.14}$$

$$\tilde{x}_i(k_2) = p^2\bar{x}_i(Z_N) + (1-p^2)\bar{x}_i(Z_{N+1}) \tag{7.15}$$

substituting for $\bar{x}_i(Z_N)$ from (7.14) into (7.15) we are able to write the rate obtained in a frame with the help of the rate obtained in the previous one:

$$\begin{aligned}
\tilde{x}_i(k_2) &= p\tilde{x}_i(k_1) + (1-p)\bar{x}_i(Z_{N+1}) > p\tilde{x}_i(k_1) \\
&\vdots \\
\tilde{x}_i(k_n) &= p\tilde{x}_i(k_{n-1}) + (1-p)\bar{x}_i(Z_{N+1}) > p\tilde{x}_i(k_{n-1})
\end{aligned} \tag{7.16}$$

Hence, in a frame the rate does not drop more than by a factor $(1-p)$.

$$\tilde{x}_i(k_n) > p\tilde{x}_i(k_{n-1}) \tag{7.17}$$

Similarly for the new user we could write

$$\tilde{x}_{N+1}(k_1) = (1-p)\bar{x}_{N+1}(Z_{N+1}) \tag{7.18}$$

And

$$\tilde{x}_{N+1}(k_n) = \tilde{x}_{N+1}(k_{n-1}) + p^{n-1}(1-p)\bar{x}_{N+1}(Z_{N+1}) > \tilde{x}_{N+1}(k_{n-1}) \tag{7.19}$$

The rate of the new user is in turn monotonically increasing.

$$\tilde{x}_{N+1}(k_n) > \tilde{x}_{N+1}(k_{n-1}) \tag{7.20}$$

In practice, we assume the length of the frame is long enough compared to the number of users such that the time average rate observed in the frame is approximately the same as the mean rate of the user.

7.2.3 Iterative admission control

According to the analysis of the previous section we can implement an iterative control scheme that is applicable to both memoryless schedulers and schedulers with memory. The outline of the scheme is as follows:
The Iterative CAC: *Iteratively decrease the back-off probability until some of the rates of the active users deteriorate below some minimum tolerable value. That is, a new packet call is rejected if* $\tilde{x}_i(k_n) < \frac{x_{min}}{p_n}$ *for any* $i = 1, 2, ..., N$ *at some iteration* n. *Otherwise* $p_n \rightarrow 0$ *and the user is admitted to the network.*

We can generalize the admission control scheme by allowing the decrease rate of the back-off probability to change from one iteration to another i.e. not necessarily an exponential decrease. Let the back-off probability at iteration n be p_n. The rate equation for the active users becomes

$$\tilde{x}_i(k_n) = p_n \tilde{x}_i(k_{n-1}) + (1 - p_n)\bar{x}_i(Z_{N+1}) \quad i = 1, 2, \cdots N \tag{7.21}$$

With $p_n < p_{n-1}$. From a protective point of view we could select the initial value p_0 to be $0.99 \leq p_0 < 1.0$. Later we could select

$$p_n = \frac{\mu_{min}}{\tilde{x}_i(k_{n-1})} \quad n > 0 \tag{7.22}$$

If $p_n = 1$, the new packet call is rejected; otherwise we keep iterating. As long as $0 < p_n < 1$ for all n, $p_n \rightarrow 0$.

7.2.4 Non-stationarity

In this section a discussion of a slot-by-slot approach is made for schedulers with memory. In the earlier analysis it was conditioned that the implementation of the admission controller should be made on a frame-by-frame basis for this type of schedulers. Consider a selection rule of the form

$$i^*(k) = \text{argmax}_{i \in \mathcal{A}(k)} \{\nabla U(\bar{x}_i(k))\mu_i(k)\} \tag{7.23}$$

Where $\mu_i(k)$ is the actual service rate process (equivalently the instantaneous channel state), $\bar{x}_i(k)$ denotes the estimated throughput and $\mathcal{A}(k)$ denotes the set of active users at time-slot k. Assume that the actual data rate process is wide sense stationary and ergodic. The rate estimator is assumed to be unbiased. We assume that U is a non-increasing function of $\bar{x}_i(k)$ and a non-decreasing function of $\mu_i(k)$. Define

$$\begin{aligned} \chi_i(\mu_i(k), \bar{x}_i(k), \mathcal{A}(k)) &= 1 \quad i = \text{argmax}_{i \in \mathcal{A}(k)} \{\nabla U(\bar{x}_i(k))\mu_i(k)\} \\ &= 0 \quad \textit{otherwise} \end{aligned} \tag{7.24}$$

At time-slot k, the mean rate is given by

$$\bar{x}_i(k) = E\{\mu_i(k)\chi_i(\mu_i(k), \bar{x}_i(k), \mathcal{A}(k))\} \tag{7.25}$$

Let us consider the saturated condition in which $\mathcal{A}(k) = \mathcal{A} = \{1, 2, ..., N\}\ \forall k$. It has been shown by (Stoylar (2005)) that the scheduling rule (7.23) converges $\bar{x}_i(k) \to \bar{x}_i$ which maximizes the utility function $U(\bar{x}_i(k))$. In the steady state, the scheduling rule becomes

$$i^*(k) = \text{argmax}_{i \in \mathcal{A}(k)} \{\nabla U(\bar{x}_i(k))\mu_i(k)\} \tag{7.26}$$

Lemma 1. *For $\mathcal{A}_N \subseteq \mathcal{A}_{N+1}$, we have*

$$\Pr\{i^*(k) = \text{argmax}_{i \in \mathcal{A}_N} \{\nabla U(\bar{x}_i(k))\mu_i(k)\}\} \geq$$
$$\Pr\{i^*(k) = \text{argmax}_{i \in \mathcal{A}_{N+1}} \{\nabla U(\bar{x}_i(k))\mu_i(k)\}\}$$

and

$$E\{\mu_i(k)\chi_i(\mu_i(k),\bar{x}_i(k),\mathcal{A}_N)\} \geq E\{\mu_i(k)\chi_i(\mu_i(k),\bar{x}_i(k),\mathcal{A}_{N+1})\}$$

Proof.

Assume at time-slot $k = 0$, the system with N users was in steady state. Assume further that we use an admission control rule, in which a new user trying to get access to the system is excluded from the active set at time-slot k with probability p^k. Define

$$\bar{x}_i[\mathcal{A}_N;\mathcal{A}_{N+1}] = E\{\mu_i(k)\chi_i(\mu_i(k),\bar{x}_i(\mathcal{A}_{N+1}),\mathcal{A}_N)\} \tag{7.27}$$

and

$$\bar{x}_{L,i}(k) = p^k\bar{x}_i[\mathcal{A}_N;\mathcal{A}_{N+1}] + (1-p^k)\bar{x}_i[\mathcal{A}_{N+1};\mathcal{A}_{N+1}] \tag{7.28}$$

$$\bar{x}_{U,i}(k) = p^k\bar{x}_i[\mathcal{A}_{N+1};\mathcal{A}_N] + (1-p^k)\bar{x}_i[\mathcal{A}_N;\mathcal{A}_N] \tag{7.29}$$

Where $\bar{x}_{L,i}(k)$ and $\bar{x}_{U,i}(k)$ are lower and upper bounds of $\tilde{x}_i(k)$ and

$$\lim_{k\to\infty}\tilde{x}_i(k) = \lim_{k\to\infty}\bar{x}_{L,i}(k) = \bar{x}_i(Z_{N+1}) \tag{7.30}$$

Showing that the expected rate with the new user added is equivalent to the lower bound rate as $k \to \infty$. Similar to our earlier analysis, we can write the rate obtained in slot k with the help of the rate obtained in slot $k-1$ as follows

$$\bar{x}_{L,i}(k) = p\bar{x}_{L,i}(k-1) + (1-p)\bar{x}_i[\mathcal{A}_{N+1};\mathcal{A}_{N+1}] \tag{7.31}$$

$$\bar{x}_{U,i}(k) = p\bar{x}_{U,i}(k-1) + (1-p)\bar{x}_i[\mathcal{A}_N;\mathcal{A}_N] \tag{7.32}$$

Based on (7.30) we can write

$$\tilde{x}_i(k) \geq p\bar{x}_{L,i}(k-1) + (1-p)\bar{x}_i[\mathcal{A}_{N+1},\mathcal{A}_{N+1}] \tag{7.33}$$

$$\tilde{x}_i(k) - \bar{x}_{L,i}(k-1) \geq -(1-p)[\bar{x}_{L,i}(k-1) - \bar{x}_i[\mathcal{A}_{N+1},\mathcal{A}_{N+1}]] \tag{7.34}$$

■

Hence, we have been able to write the expected rate in a slot-by-slot basis with the help of the steady lower bound mean rate $\bar{x}_{L,i}$ that asymptotically converges with the expected rate $\tilde{x}_i$

7.3 Multi-user iterative admission control

Let us limit the number of new users to M. The back-off probability is now set as $p^{\frac{1}{M}}$. This choice will guarantee that the rate of the old users cannot drop more than the fraction $(1-p)$ in a frame. Let $\mathcal{Z}_{N+m}$ denote the set of channel statistics matrices of length $N+m$ that can be formed from the vector of length $N+M$ by removing some of the elements $N+1, N+2, \cdots, N+M$. With the help of this notation the mean rate obtained in frame n can be written as follows

$$\tilde{x}_i^{(n)} = \sum_{k=0}^{M} p^{\frac{nk}{M}} \left(1 - p^{\frac{n}{M}}\right)^{M-k} \sum_{Z \in \mathcal{Z}_{N+M-k}} \bar{x}_i(Z) \tag{7.35}$$

Notice that when $M = 1$ (7.35) is reduced to (7.6). For $0 < p < 1$, we have $(1 - p^{\frac{n+1}{M}})^m \geq (1 - p^{\frac{n}{M}})^m$ for all $m = 0, 1, 2, \cdots$. Hence, we have

$$\begin{aligned} \tilde{x}_i^{(n+1)} = p\tilde{x}_i^{(n)} + \sum_{k=0}^{M} p^{\frac{(n+1)k}{M}} (p^{\frac{k}{M}} (1 - p^{\frac{(n+1)}{M}})^{M-k} \\ -p(1 - p^{\frac{n}{M}})^{M-k}) \sum_{Z \in \mathcal{Z}_{N+M-k}} \bar{x}_i(Z) \geq 0 \end{aligned} \tag{7.36}$$

That is, the rate of an active user cannot drop more than a factor of $(1-p)$ from frame to frame even if M new users are trying to get access to the system. Thus, we can directly use the iterative CAC described in section 7.2. Let us define two column vectors

$$R_i(n) = \left(\mu_i^{(n)}, \mu_i^{(n+1)}, \cdots, \mu_i^{(n+M)}\right)'$$

and

$$Q_i = \left(\bar{x}_i(Z_{N+M}), \sum_{Z \in \mathcal{Z}_{N+M-1}} \bar{x}_i(Z), \cdots, \bar{x}_i(Z_N)) \right)'$$

and a square matrix $A(n)$ consisting of elements $a_{nk} = p^{\frac{nk}{M}} \left(1 - p^{\frac{n}{M}}\right)^{M-k}$. Note that by X' we denote the transpose of the vector X. Now we can write (7.35) in matrix form as

$$R_i(n) = A(n)Q_i \tag{7.37}$$

Matrix A is a square matrix with full rank making it invertible. Thus, we could use the rate prediction CAC by checking whether the following holds

$$Q_i = A(n)^{-1} R_i(n) \geq 1_{N+M} \bar{x}_{min} \tag{7.38}$$

where 1_m denotes a vector of m ones. It is notable however, for large n and M, A would be ill-conditioned and thus its inverse A^{-1} might not behave well numerically. The main issue in performing multi admission will be the convergence rate as it will become much slower due to the fact that each user makes the back-off decision independently of the others.

7.4 Kalman Filter Estimation

So far we have assumed that mean value $\tilde{x}_i$ is available for admission control. In practice, we have to use time average values $\bar{x}_i$ that in case of ergodic channel would converge to $\tilde{x}_i$ when the frame length m approaches infinity. In practice, we have to cope with finite frame sizes and thus noisy estimates for the mean value. The main power of the formulation lies in that it allows the tracked parameter (in this case the time average rate) to change slowly. Let us define two state variables $y_{1,i}^{(n)} = \tilde{x}_i(k_n)$ and $y_{2,i}^{(n)} = \bar{x}_i(Z_{N+1})$ and state vector $Y_i^{(n)} = (y_{1,i}^{(n)}, y_{2,i}^{(n)})'$. Assume that the state noise and measurement error are Gaussian white noise processes described by $V_i^{(n)} = \left(v_{1,i}^{(n)}, v_{2,i}^{(n)}\right)'$ and $e_i^{(n)}$ respectively. Since $\tilde{x}_i(k_n)$ is the estimator of the mean value, then the error $\bar{x}(k_n) - \tilde{x}(k_n)$ becomes Gaussian due to law of large numbers. The two processes are assumed to have zero mean and follow $\Psi_{VV}(t) = \mathrm{E}\{V_i^{(n)}(V_i^{(n)})'\}$, $\Psi_{ee}(t) = \mathrm{E}\{e_i^{(n)}(e_i^{(n)})'\}$, and $\Psi_{Ve}(t) = \mathrm{E}\{V_i^{(n)}(e_i^{(n)})'\}$.

Now the state equations can be written as follows

$$Y_i^{(n+1)} = \Phi^{(n)} Y_i^{(n)} + V_i^{(n)} \tag{7.39}$$

$$\bar{x}_i(k_n) = C X_i^{(n)} + e_i^{(n)} \tag{7.40}$$

For $i = 1, 2, 3, \ldots, N$ where $\Phi^{(n)} = \begin{bmatrix} p_n & 1-p_n \\ 0 & 1 \end{bmatrix}$, $C = [\ 1 \quad 0\ ]$. It is well-known that the optimal state estimator for the process is the Kalman filter which can be written as follows

$$\hat{Y}_i^{(n+1)} = \Phi^{(n)} \hat{Y}_i^{(n)} + K^{(n)} \left(\bar{x}_i(k_n) - C\hat{Y}_i^{(n)} \right) \tag{7.41}$$

$$K^{(n)} = (\Phi^{(n)} P^{(n)} C' + \Psi_{Ve}(t))(C' P^{(n)} C + \Psi_{ee}(t))^{-1} \tag{7.42}$$

$$P^{(n+1)} = \Phi^{(n)} P^{(n)} (\Phi^{(n)})' + \Psi_{VV}(t) - K^{(n)}(C P^{(n)} C' + \Psi_{ee}(t))(K^{(n)})' \tag{7.43}$$

Where K denotes the Kalman gain and P represents the error covariance matrix (a measure of the estimated accuracy of the state estimate). Unfortunately, it is difficult to determine the covariance matrices $\Psi_{VV}(k)$, $\Psi_{ee}(k)$, and $\Psi_{Ve}(k)$ accurately. Instead they can be used as tuning parameters. The larger we set the parameters in $\Psi_{ee}(k)$ the less, we trust in the measurements. The covariance matrix $\Psi_{VV}(k)$ describes the rate at which the mean values change in the channel and is thus related to the mobility of the users. Hence, the faster the mobiles move the larger $\Psi_{VV}(k)$ and the more weight is given to the instantaneous channel estimate. Eq. (7.3) could be seen as a fixed Kalman filter due to the fact that it does not consider the impact of the new user. The Kalman filter presented in this section could be seen as an extension of Eq. (7.3) that takes into account the dynamics of the system making it better suited for the admission process.

7.5 Non-iterative admission control

In this section a Recursive Least Square (RLS) scheme is presented to later compare with the scheme. The RLS scheme is a simple one-shot admission control scheme that estimates the impact of adding a new user to the data rates of active users. This kind of admission control scheme has been suggested in (Gribanova 2004).

Recursive least squares algorithm

The basic RLS scheme was used where a filter of weight w is used to predict future values of the rates depending on past observed data.

The filter weight is updated using the RLS equations

$$K(a) = \frac{\lambda^{-1}P(a-1)u(a)}{1+\lambda^{-1}u^H(a)P(a-1)u(a)} \tag{7.44}$$

$$\epsilon(a) = \bar{x}_i(a) - w(a-1)u(a) \tag{7.45}$$

$$w(a) = w(a-1) + K(a)\epsilon(a) \tag{7.46}$$

$$P(a) = \lambda^{-1}P(a-1) - \lambda^{-1}K(a)u^H(a)P(a-1) \tag{7.47}$$

where a represents the iteration index, K represents the filter gain, P is the correlation matrix, λ is the forgetting factor, u represents the input and is given by

$$u(a) = [\ \bar{x}_i(a-4) \quad \bar{x}_i(a-3) \quad \bar{x}_i(a-2) \quad \bar{x}_i(a-1)] \tag{7.48}$$

$\epsilon(a)$ is the estimation error, $\bar{x}_i$ denotes the mean rate values for user i at admission a and w are the coefficients of the filter.

A one-tap filter w was considered due to the limited number of users, so

$$\hat{\bar{x}}_i(a) = w\bar{x}_i(a-1) \tag{7.49}$$

The RLS admission control scheme can be summarized as follows:

RLS CAC: *Record the rate losses caused by previous admissions. Utilize RLS to* *estimate the impact of adding a new user. If the predicted rate of the worst active user is above $\bar{x}_{min}$, admit the new user; otherwise reject it.*

7.6 Numerical results

In this section different scenarios will be implemented to the controller as well as studying the admission error possibilities.

7.6.1 Static traffic (Full buffer)

Simulations were made with the help of a computer simulator that creates a mobile cell environment. The simulator would generate N users having different packet sizes. Assuming pedestrian profiles for the users, hence their channel conditions are slowly changing. However the coherence time of each user's channel will depend on its speed. In the beginning it was assumed that all users have full buffers, so they all had data to receive. The scheduling rule that was used in this simulator was the PF scheduler where users are selected in every TTI based on how good the instantaneous channel condition was relative to the time average condition, (7.1) and (7.21).

Table 7.1 shows the parameters that were used in the simulation program. The simulation was considered for WCDMA HSDPA where adaptive modulation and coding is used to guarantee high throughputs depending on the channel condition as shown in Table 7.2 (Horng, Zhang and Lao 2003).

One new user was added every time using the iterative CAC procedure described in Section 7.2. The Kalman filter described in Section 7.4 was implemented to obtain the mean values since we are averaging over small window sizes.
The initial values for $Y_i^{(k)}$were the time average rates of the active users before a new user entered the system. A Rayleigh fading vector was generated for each user. The generator is implemented in accordance with Jake's fading simulator (Jakes 2001) and (Pop and Beaulieu 2001). The SINR-rate mapping was done

Table 7.1: System parameters for Admission Control

Parameter	**Value**
Carrier frequency	2 GHz
Spreading factor	16
Number of multicodes	10
TTI duration	2 ms
Fading model	One path Rayleigh (Jake's model)
Minimum rate allowed	256 kbps
	384 kbps
Max. number of associated DPCH (N)	15 for QoS 256 kbps
	10 for QoS 384 kbps
Radio propagation	Site to site distance 500 m
	Path loss component 4
	Std. of shadow fading 6 dB
	Min mobile speed 3 km/h
	Max mobile speed 10 km/h
BS Tx power	17 W
Ec/Ior	0.7
Hybrid ARQ	Combining method Chase method
	Sched. retran. Highest priority
Back-off probability	0.999
Admission decision time-slot	4.2s
Ψ_{ee} values	$2 \times 10^5, 4 \times 10^4, 1 \times 10^3$
Ψ_{Ve}, Ψ_{VV}	$0, 10I_{2\times 2}$ respectively

according to Table 7.2 which shows the modulation indices for different SINR levels. Different users experience different channel conditions that vary depending on their distance from the base-station and velocity. The velocities of the mobile users were independent random variables uniformly distributed between the minimum 3 km/h and the maximum velocity 10 km/h. In the simulation if a user moves out of the border, it will reappear at a point on the opposite border that is symmetric to the exiting point.

In the simulation a new user was added every 6000 slots (12 sec). The user's time average rate in the beginning would experience a transient state due to averaging over a small number of time-slots. The rate gradually stabilizes as the number of time-slots increases.

Fig. 7.1 shows the Kalman estimate along with the mean (expected) rate for the

Table 7.2: Thresholds of SINR

Index i	1	2	3	4	5	6	7
Threshold (dB)	-1.9	1.25	4.5	6.5	10.2	14	16.2
Modulation and Code rate	QPSK 1/4	QPSK 1/2	QPSK 3/4	16QAM 1/2	16QAM 3/4	64QAM 5/8	64QAM 3/4
Rate (Mbps)	1.2	2.4	3.6	4.8	7.2	9.0	10.8

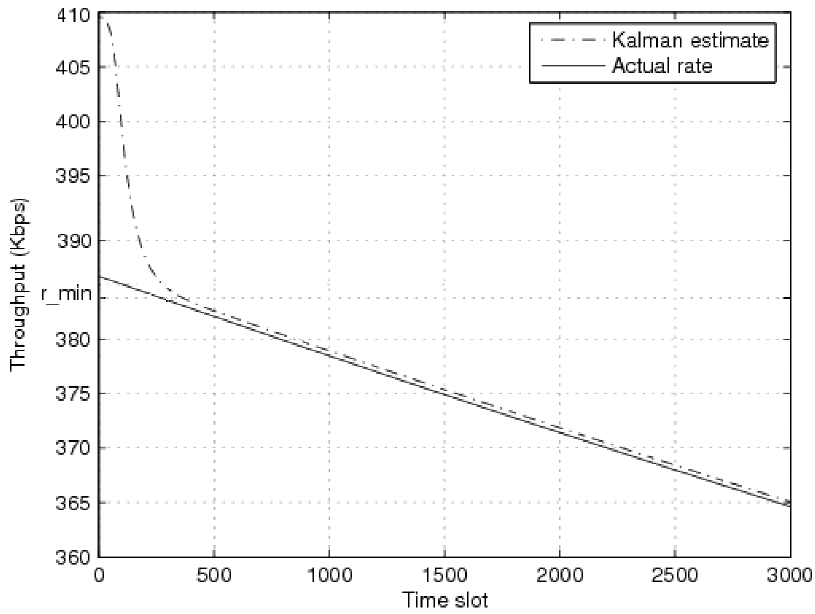

Figure 7.1: Rate of worst user (Back-off factor=0.999)

worst user when a new user is admitted. The figure shows how the Kalman estimate converges to the mean value. In this figure it could be seen that the mean rate of the worst user has declined beyond the minimum acceptable rate. With the help of the Kalman filter, we apply the scheme and the controller will make the decision to either accept or reject the new user. In this case the new user is rejected.

A new user is fully accepted if none of the active users experienced a fall in rate below $\bar{x}_{min}$ during some time window. The performance of the admission scheme is measured by the admission errors. There are two types of CAC errors:

- Type I error: Where a new user is erroneously accepted resulting in outage.
- Type II error: Where a new user is erroneously rejected resulting in blocking.

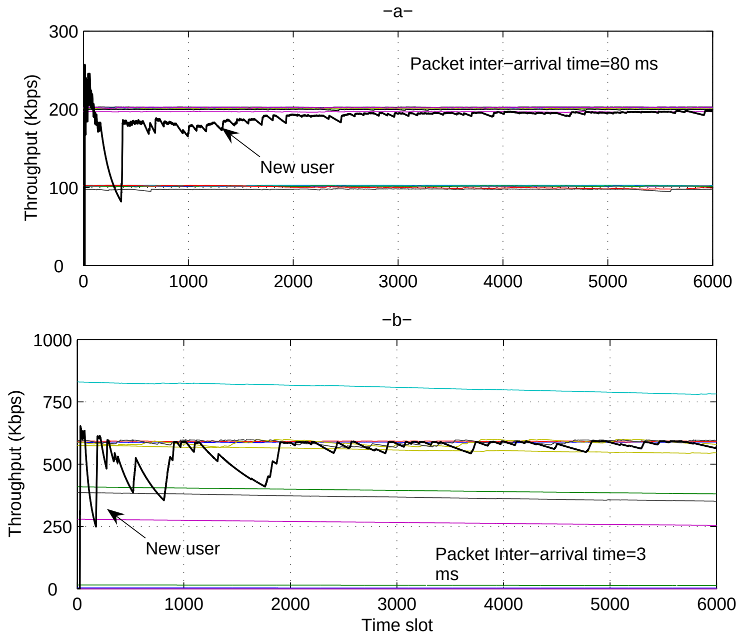

Figure 7.2: Dynamic traffic users with different packet inter-arrival times

Pedestrian channels experience a longer coherence time than that of vehicle users and consequently this will affect the mean rate of these users. For example an admission error may occur because an active user had a good channel and came under a bad fading pattern causing a temporary drop in its mean rate below $\bar{x}_{min}$ but later recovered after moving away from the cause of that fading dip. This in turn causes an admission error type II if it happened during the probing time window and error type I if after.

7.6.2 Dynamic Traffic

In this case the buffer occupancy of the users is allowed to vary so that not all users necessarily have data to receive in each time-slot. This differs from the previous static case discussed earlier, in which all the users were assumed to have full transmission buffers all the time. The mean inter-arrival time of the packets will play an important role as rates will be a function of packet arrival as well as channel conditions. With few packet arrivals the rates of users will be divided into groups as shown in the two lines in Fig. 7.2 mainly due to the fact that the rates are more of a function of packet arrivals than channel conditions, the impact of the channel condition in this case could be regarded similar to a quantizing impact. Our interest is studying the new user's channel impact with respect to the active users' channels. Therefore, the mean inter-arrival time was decreased to create more packets and consequently make the rates be more of a function of channel conditions than packet arrivals. The difference between dynamic users with two different mean inter-arrival times is illustrated in Fig. 7.2. In Fig. 7.2-a we could see that the impact of the new user is small due to the fact that the rates here are more functions of packet arrivals than channel conditions. However in Fig. 7.2-b the impact of the new user is more noticable where we see the decrease in the mean rate of the other users as the rate of the new user increases. In the simulation, the inter-arrival times follow the log-normal distribution. The decision to use a log-normal process for packet arrivals is due to its longer tail probability property which makes it more appropriate to model the bursty nature of the data traffic than the Poisson process.

Table 7.3 illustrates the possibilities of a user being erroneously accepted (Type I CAC error) or erroneously rejected (Type II CAC error). Both static and dynamic cases are considered and two minimum QoS levels. The results shown in the table were obtained by repeating the simulation 100 times and computing the percentage of error for the total number of simulations. Tuning the noise covariance matrix Ψ_{ee} in some cases can affect the admission decision as it determines how fast the Kalman filter estimate converges to the actual value leading to an increase or decrease in CAC errors as shown in Table 7.4. In this table different values are assigned to Ψ_{ee} and observe the impact on the percentage of errors. It could be seen that the higher the estimation error is the more errors we get.

7.6.3 Decision validity

Different patterns of multipath fading have a significant impact on the mean rate which in turn will cause confusions in the admission decision as noticed earlier. In this section, a discussion is made for this problem with a suggestion for a solution

Table 7.3: Type I and II CAC errors

Traffic type	QoS level	Error type I	Error type II
Static	256 kbps	4%	6%
	384 kbps	8%	4%
Dynamic	256 kbps	9.09%	3.03%
	384 kbps	6.25%	3.12%

Table 7.4: Impact of Ψ_{ee} on static traffic users

Ψ_{ee}	2×10^5	4×10^4	1×10^3
256 kbps QoS type I error	6%	4%	4%
384 kbps QoS type I error	12%	8%	4%

using an illustrative example.

Example

For error type I shown in Fig. 7.3. The new user was admitted and later there was a temporary fall in the rate of the worst user but then recovered. This temporary fall resulted in declaring a type I error.

According to Fig. 7.4 the new user will be blocked due to the fall in rate of the worst user below the acceptable threshold. The worst user will then have a temporary improvement in the fading pattern causing an increase in its rate, but later reverts to its original state. In this case, we will have a type II admission error, but clearly this will not be a genuine error because the decision to block the new user was in fact accurate, the temporary rise in rate was the main reason to trigger the error flag.

Solution

We propose to solve this problem the use of a heuristic scheme, in which we observe the number of times the rate crosses a reference (in our case the rate threshold) and the time between each crossing and associate these observations with the decision making. This will enable us to form a good picture about the variations in throughputs and consequently back up the admission decision.

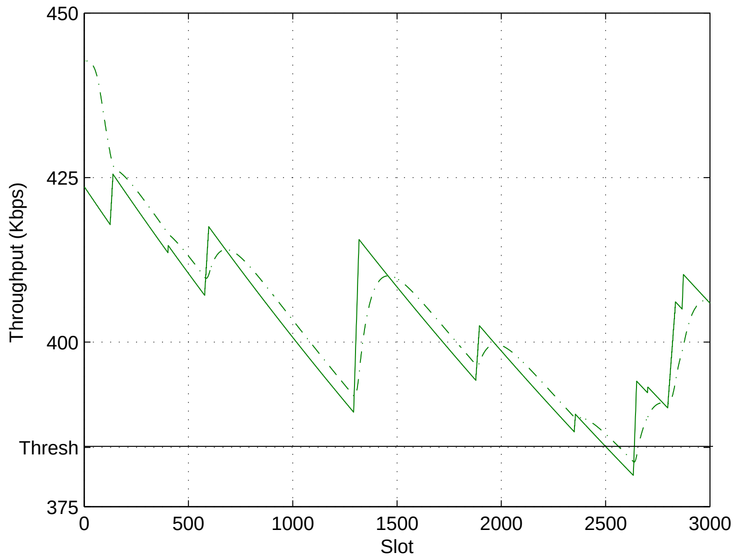

Figure 7.3: Questionable error I

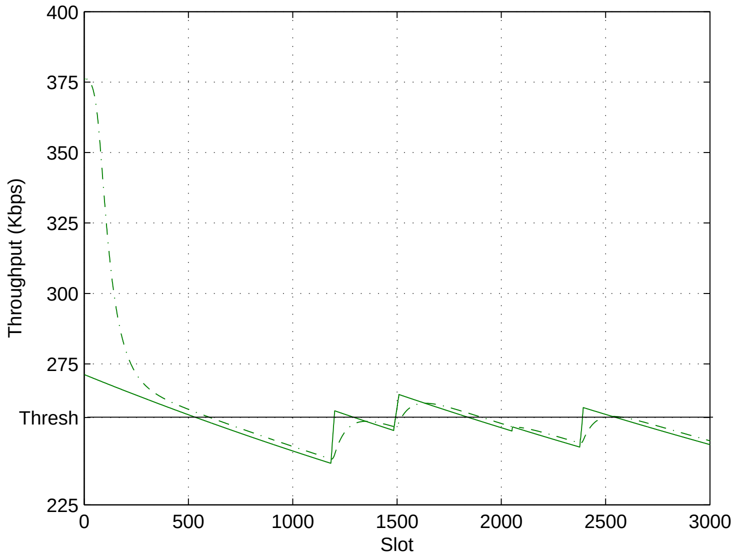

Figure 7.4: Questionable error II

Procedure

Error type I

In the case of error type I, we observe how many times the mean rate crossed the threshold after the probing time window, naturally if the crossing occurred more than once then it should be investigated. Let's apply the technique to Fig. 7.3. In the figure, we note that the threshold has been crossed twice, the time between the crossings is 134 slots (which is very short). Hence we could easily conclude that this fall in rate was not inflicted by the new user but rather a change in the user's fading pattern and the decision was successful and this wasn't a genuine error.

Error type II

The same procedure is carried out. Looking at Fig. 7.4 we can see that there were 6 crossings. Since it is a type II error, we are mostly interested in the number of times the mean rate crosses from below to above the threshold. Again we observe the time between the consecutive crossings to determine the ratio of the rate being above the threshold to it being under.

After the first crossing, the ratio of the time the rate was above threshold is relatively smaller to the time it was under and thus we can conclude that this wasn't a genuine error.

7.6.4 Multi-user admission control

Fig. 7.5 shows the rate of the worst user when 3 users are being added.

An alternative would be that all users use the same back-off probability and jointly back off. In this case, the mean rate obtained in frame n becomes

$$\tilde{x}_i^{(n+1)} = p\tilde{x}_i^{(n)} + (1-p)\bar{x}_i(Z_{N+M}) \tag{7.50}$$

Naturally, the convergence will be faster, but on the other hand the drawback will be that all M users will be denied admission if $\tilde{x}_i^{(n+1)}$ dropped below the minimum rate indicating the unfairness of this alternative. Table 7.5 represents the results obtained

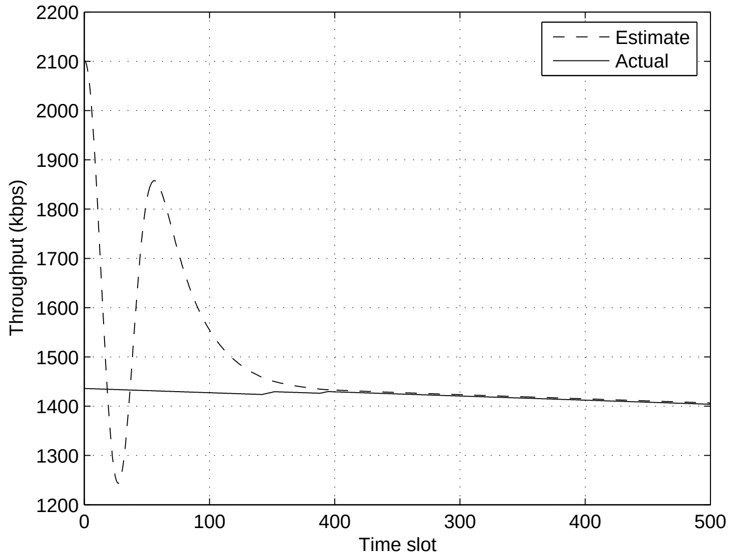

Figure 7.5: Multiuser case - independent probability factors

Table 7.5: Type I and II CAC errors (Multiuser case - same probability factor for all)

Traffic type	QoS level	Error type I	Error type II
Static	256 kbps	5%	0%
	384 kbps	12.5%	2.5%
Dynamic	256 kbps	7.14%	3.57%
	384 kbps	3.57%	0%

for joint back-off admission. The admission was made for 3 users at a time. The joint back-off option is likely to be the best alternative despite its drawback since time is the most critical element for users requesting admission.

7.6.5 Non-iterative admission control

Fig. 7.6 shows the RLS algorithm application to the worst user. It could be seen that when an 11th user is added, the rate of the worst active user drops below the minimum acceptable rate. Table 7.6 shows a comparison between the RLS and ALP-CAC schemes in terms of admission error where the results for the ALP-CAC are from Table 7.3 and the results for the RLS were obtained by running repetitive

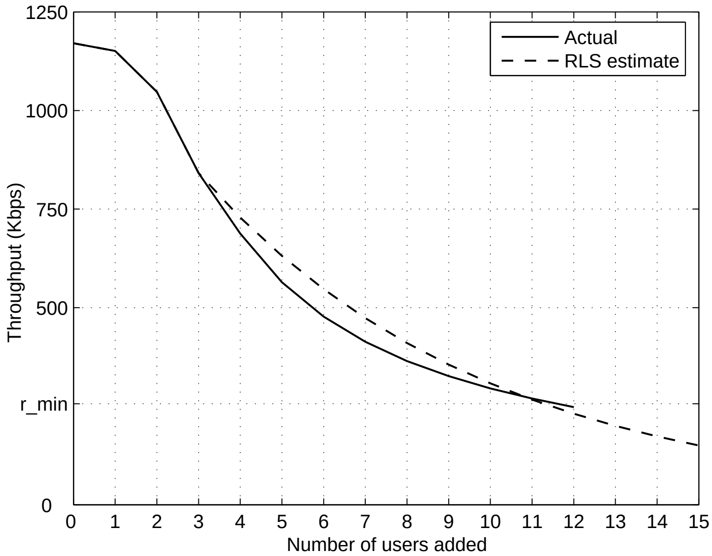

Figure 7.6: RLS admission scheme (Worst user)

Table 7.6: Comparison of ALP-CAC and RLS schemes

Traffic type	Scheme	Error type I	Error type II
Static	ALP-CAC	4%	6%
	RLS	13.7%	18.5%
Dynamic	ALP-CAC	9.09%	3.03%
	RLS	16%	22%

simulations and computing the percentage of admission errors I and II for those simulations. In the RLS case an admission error occurs due to the estimation error resulting in denying a new user admission at the time the actual rate of the worst user was still above the minimum rate or vice versa. The results indicate that the ALP-CAC scheme is superior in all cases.

7.7 Concluding remarks

The suggested iterative CAC method proves to be very promising and satisfies minimum QoS rate levels for all users in an ongoing system as it tends to protect the active users and guarantee that the new user being admitted will not violate the QoS level provided to them. The scheme is applicable with stationary and non-stationary scheduling rules. The scheme is further extended allowing multiple users to request admission simultaneously. However, it was noticed that the scheme suffered the problem of slow convergence. In order to speed up the convergence, a method is proposed where the new users jointly back-off. This scheme can be too cautious because of the issue of fairness by denying admission to users that do not cause the active users' rates to drop below threshold. Finally, a comparison is made with a one-shot admission control scheme. The results suggest that our scheme can achieve smaller admission error probabilities than the RLS based one-shot scheme.

Chapter 8

Conclusion

The challenge of resource scheduling is not to solve high dimensional optimization problems, but to develop algorithms that could be implemented in practice. For this reason, control engineering and computing methods were utilized in this research. The work mainly focuses on the exploitation of the varying nature of communication channels. The result of this kind of exploitation is an efficient use of resources since resources are granted to users who utilize them the best at a certain time.

In this work opportunistic schedulers were proposed for the downlink and uplink directions. The downlink scheduler would deliver requested quality of service levels to the users. In the event the scheduler failed to provide a requested level, the user is still guaranteed to obtain the quality of service it would obtain with a proportional fair scheduler. In this thesis we analyzed rate control as one example of the QoS requirements. Other QoS classes are possible such as packet delay control, however the analysis for such QoS is quite complicated.

For the uplink, a heuristic scheduler was proposed. The scheduler takes into account ARQ processes in the scheduling decision as well as the resource allocation constraint. The main goal of the scheduler is to find the set of user-RB pairs that would maximize a certain constrained optimization problem. Due to the localization constraint, we suggested a heuristic that could be implemented in practice and follows a greedy approach to solve the problem. We were also able to modify the optimization problem so that it could be solved theoretically and provide the optimal solution. When comparing the heuristic solution with the optimal solution, it was found that the heuristic solution provided relatively good results.

A study was carried out to examine feedback in multi-carrier systems. The study highlights system performance under different feedback criteria. The model we considered for the study resembles a SC-FDMA model, in a sense that subcarriers are grouped together. However, we did not consider restrictions in group assignment. We used order statistics to find the optimal feedback information. The study compared this optimal information with the information that was based on the smallest SINR subcarrier, the maximum SINR subcarrier, the median SINR subcarrier and the average SINR of the block. It was found that at low and practical SINR regions, the median SINR provided the nearest result to that of the optimal. At a high SINR region, the minimum SINR decision variable was the closest to the optimal.

An opportunistic admission controller was introduced in this work. Contrast to common controllers that base their decision of directly rejecting or accepting a new user on an estimation of the load this user will cause to the system, the opportunistic controller instead gradually integrates the new user and examines the system online. The user is later rejected or accepted. The main issue in admission control is the possibility of decision error. This leads to either accepting a user in an already loaded system and consequently resulting in the drop of the quality of service for active users or rejecting a user at a time its admittance would not have degraded the quality of the active connections. The errors would usually happen due to temporary changes in the channel conditions of users that would be interpreted wrongly. The author therefore made a simplistic approach in an attempt to reduce the admission errors. The approach studies the behaviour of the channel before forming a decision. The author also considered multi-user admission, where more than one user are admitted at a time.

Finally, an appendix is attached to validate the results obtained in this work by verifying that the models considered in the thesis deliver the same results of other models when using the same scenarios and assumptions.

8.1 Future Work

In chapter 5 we considered only localized SC-FDMA. One suggestion for future work is to consider distributed FDMA. Another suggestion would be trying to find a different approach in resource block assignment other than the greedy approach. A study could be made in examining and comparing different approaches. In this work we only considered static traffic models where it was assumed all users had full buffers all the time. For future work we could consider dynamic traffic cases with different classes of traffic. The models utilized in this thesis only consider single-case scenarios. Thus, interference from neighboring cells was neglected. Naturally, interference will have a significant impact on cell edge users. Since, we only consider WCDMA and LTE systems, the frequency reuse factor is 1. Therefore, one solution would be cell coordination to limit the amount of interference. Designing such a coordination could be considered in future work. One question that comes to mind is: could the admission control and scheduling algorithms suggested for the HSDPA systems be also suitable for LTE systems. The answer would be yes since LTE also utilizes TDMA making it possible to implement those algorithms. This could also be examined in future work.

In chapter 7, the main issue was to detect an original fall in QoS for ongoing calls to make an admission decision. Users moving with different speeds and in different directions are constantly subjected to different fading patterns. This affects the suggested algorithm dramatically. Hence, our algorithm is more sensitive to channel condition variations than others. Our suggestion of a level crossing counter to verify a drop or rise in channel state may not be a very good choice. A study could be made to find a more efficient way to verify these channel states for the controller.

Appendix A

Validity of models

Our models are simplifications of real systems. The main goal of them is to capture the basic characteristics of these real systems. In this appendix we will attempt to verify the models in order to prove that the results they produce are valid results. The verification is made by comparing the performance obtained with our models and the performance of other literature models using their parameters and scenarios. The comparison is merely to give an insight of the credibility of our results. We address the two main system models considered in this thesis. The first representing the HSDPA system and the second is the LTE system.

A.1 HSDPA

We will first look into the HSDPA model. We make the reference in this section to the HSDPA model used in a 3GPP TSG RAN document (3GPP RP-01.0586). The document mainly discusses HSDPA throughputs with fixed and variable sized TTIs. In our model we assume only fixed TTIs. Therefore, the comparison will be done only with the fixed TTI result. We consider the same simulation parameters used in the paper and apply them to our simulator and try to compute the cumulative distribution function (CDF) for packet call throughputs for 37 UEs.

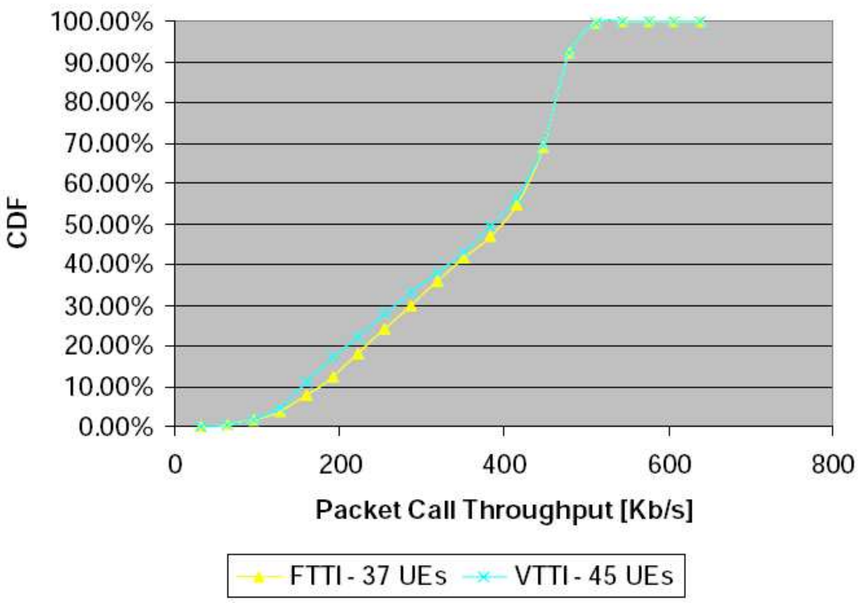

Figure A.1: CDF of packet call throughput from reference HSDPA model (3GPP RP-01.0586)

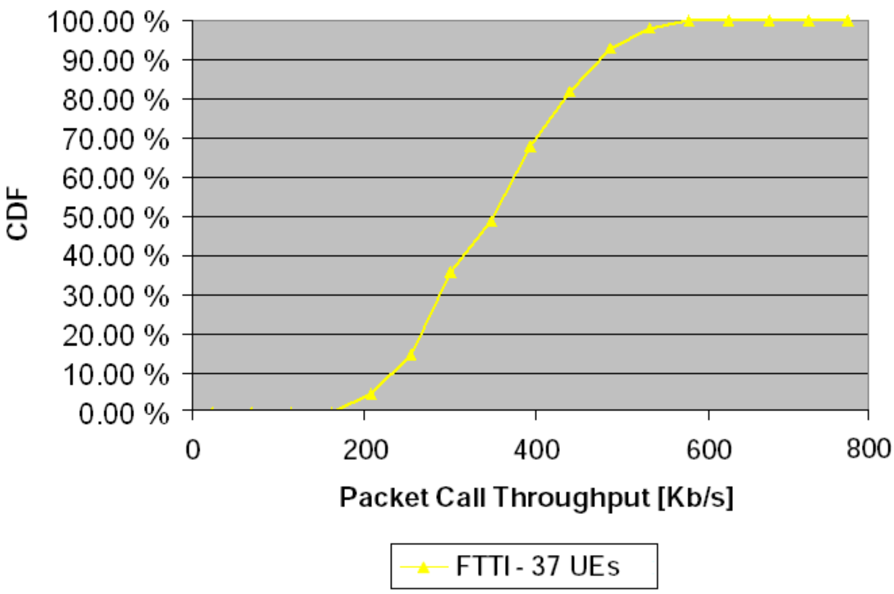

Figure A.2: CDF of packet call throughput from our model

A.2 LTE

We are only interested in the uplink part of LTE, hence the comparison will be with another work employing a SC-FDMA model. For this part, we will refer to a thesis work (Jersenius, 2007). The thesis mainly suggests simplistic scheduling approaches for localized SC-FDMA. We will apply some of the scheduling methods proposed in the thesis to our model using the same scenarios, simulation assumptions and parameters. We will consider the following scheduling algorithms:

1. Channel Dependent Time Domain Scheduling (CDT): In every TTI, the active user with the largest average gain to interference ratio (GIR) (average over all RBs) is assigned all resource units.

2. Channel Dependent Frequency and Time Domain Scheduling (CDFT): Resource units are grouped into resource groups, the number of these groups depends on the number of active users. Selection is made by choosing the users that maximize the 'user-resource group' pair.

A.3 Validation

Comparing the reference Figures A.1 and A.3 with the figures produced with our models using the same scenarios and parameters A.2 and A.4, we could see that our simulator provides a roughly close result or at least provides the same order of magnitude to the ones obtained with the simulators considered in the references. What we can conclude from this result is that we could comfortly say that the results we obtain from our HSDPA and LTE models can be considered reliable since the models provide similar performance to others when similar simulation environments are assumed.

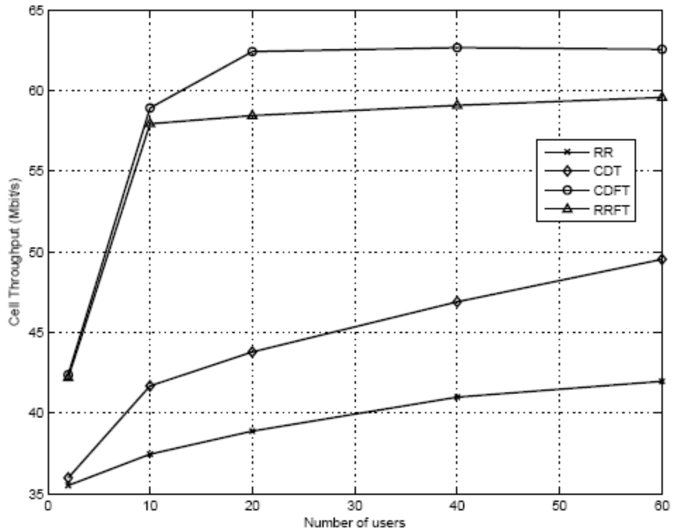

Figure A.3: Cell throughput for 2, 10, 20, 40 and 60 users from reference model (Jersenius 2007)

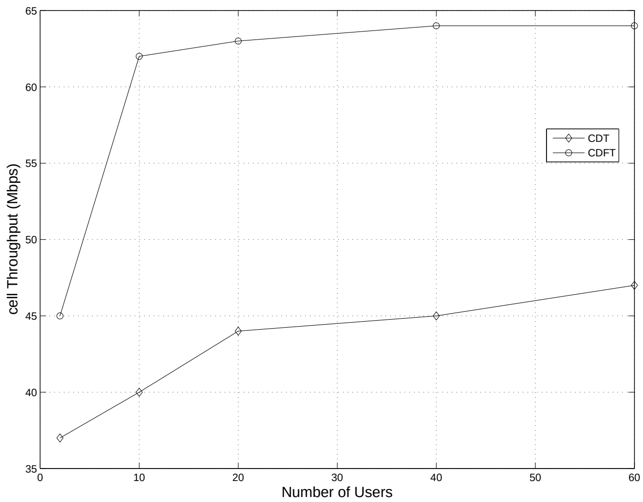

Figure A.4: Cell throughput for 2, 10, 20, 40 and 60 users from our model

Bibliography

3GPP TSG-RAN-1, "TSGR1#20(01)0586: HSDPA System Performance with Variable TTI", Lucent Technologies, May 22-25.

3GPP RP-040461, "Proposed study item on evolved utra and utran. Technical Report".

3GPP RP-060907, Nortel "Sounding channel for UL channel-dependent scheduling," May 2006.

3GPP RP-25.892, "TR 25.892: Feasibility Study for OFDM for UTRAN Enhancement", Version 1.2.0, May 2004.

Andrews M., Qian L., Stolyar A.L., "Optimal Utility Based Multi-User Throughput Allocation subject to Throughput Constraints" , INFOCOM'2005, Miami, March 13-17.

Agrawal R., Bedekar A. and Subramanian V. "A Class and Channel-Condition based Weighted Proportionally Fair Scheduler," Proc. of ITC 2001, Salvador, Brazil, Sept. 2001.

Agrawal R. and Subramanian V. "Optimality of Certain Channel Aware Scheduling Policies," Proc. of 2002 Allerton Conference on Communication, Control and Computing, Oct. 2002.

Assaad M. and Zeghlache D. "Opportunistic Scheduling for streaming services in HSDPA," in *in Proc. 17th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '06)*, 2006.

Bambos B., Chen S.C. and Pottie G. "Channel access algorithms with active link protection for wireless communication networks with power control", IEEE/ACM Transactions on Networking 8, 583 - 597, 2000.

Barreto A.N. and Furrer S. "Adaptive bit loading for wireless OFDM systems," *in Proc. 12th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '01)*, Vol. 2, pp. 88-92, 2001.

Bender P. Black P., Grob P., Padovani R., Sindhushayana N. and Viterbi A. "CDMA/HDR: A bandwidth-efficient high-speed wireless data service for nomadic users", IEEE Communications Magazine, 38, 70-77, 2000.

Berggen F. and Jäntti R. "Asymptotically fair transmission scheduling over fading channels", IEEE Transactions on Wireless Communications, 3, 326-336, 2004.

Bonald T. and Proutire A. "Insensitive bandwidth sharing in data networks", Queueing Systems, Vol 44, pp 69-100, 2003.

Chang K. and Han Y. "QoS-based Adaptive Scheduling for a Mixed Service in HDR System", Proc. IEEE PIMRC 2002, 4, 1914-1918.

Cherriman P.J., Keller T. and Hanzo L. "Subband-adaptive turbo-coded OFDM-based interactive video telephony ", *IEEE Transactions on Circuits* *and Systems for Video Technology*, Vol. 12, pp. 829-839, 2002.

Czylwik A. "Adaptive OFDM for wideband radio channels," in *in Proc. IEEE Global Telecommunications Conference (GLOBECOM '96)*, Vol. 1, pp. 713-718, 1996.

Dardari, D. "Ordered subcarrier selection algorithm for OFDM-based high-speed WLANs ," *in Proc. 15th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '04)*, Vol. 2, pp. 1220-1224, Sept. 2004.

Ekström A., Furuskär A., Karlsson A., Meyer S., Parkvall M., Torsner J. and Wahlqvist M. "Technical solutions for the 3G long-term evolution," *IEEE Communications Magazine*, Vol. 44, 2006.

Evans J.S. and Everitt D. "Effective bandwidth-based admission control for multiservice CDMA cellular networks," IEEE Trans. on Vehicular Technology, vol. 48, no. 1, pp. 36-46, Jan. 1999.

Fazel K. and Fettewis G. "Multi-Carrier Spread-Spectrum" Norwell, MA: Kluwer, p. 260, 1997.

Gribanova K. "Packet Scheduling for Video Streaming Data in the HDR system", Master's thesis, Telecommunication software and multimedia laboratory, Helsinki University of Technology, 2004.

Haykin S. and Moher M. "Modern Wireless Communications," Pearson Prentice Hall, 2005.

Horng J., Zhang J. and Lao D. "Throughput analysis for W-CDMA system with MIMO and AMC", TR-2003-48.

Hu M., Zhang J., "Two novel Schemes for Opportunistic Multiuser Communications," special session on wireless networks in IEEE Multimedia Signal Processing Workshop, Dec 2002.

Hummels D.R. and Ratcliffe F.W. "Calculation of error probability for MSK and OQPSK systems operating in a fading multipath environment," *IEEE Trans.* Veh. Technol., vol. 30, pp. 112-120, Aug. 1981.

Jakes W.C. "Microwave Mobile Communications. Wiley, NY", Section 1.7.2, 2001.

Jalali A., Padovani R. and Pankaj R. "Data throughput of CDMA-HDR: A high efficiency-high data rate personal communication wireless system," *IEEE 51st Vehicular Technology Conference, Tokyo, Japan*, Vol. 3, pp. 1854-1858, May 2000.

Jersenius k. "Uplink channel dependent scheduling for future cellular systems," *Master thesis, Linköping university*, 2007.

Kalman R.E. "A new approach to linear filtering and prediction problems," *Transactions of the ASME-Journal of Basic Engineering,* pp. (82):35-45, 1992.

Keller T. and Hanzo L. "Adaptive multicarrier modulation: a convenient framework for time-frequency processing in wireless communications," *in Proc. of the IEEE*, Vol. 88, No. 5, pp. 611-640, 2000.

Kim T. "Quantized Feedback for Slow Fading Channels", Licentiate thesis, KTH, paper A, 2006.

Knopp R. and Humblet P. ”capacity and power control in single cell multiuser communications,” in Proc. International Conference on Communications, Seattle, WA, June 1995.

Kobayashi M. and Caire G., “A practical approach for the weighted sum rate maximization in MIMO-OFDM BC”, Asilomar 2007

Kushner H. and Whiting P. ”Asymptotic Properties of Proportional-Fair Sharing Algorithms,” Allerton Conference on Communication, Control and Computing, Oct. 2002.

Lee H. and Chong S. “Combined QoS Scheduling and Call Admission Control Algorithm in Cellular Networks,” *Proc. IEEE WiOpt 2006*, pp. 114-123, 2006.

Lim J., Myung H. Oh K. and Goodman D. “Proportional fair scheduling of uplink single-carrier FDMA systems,” in *Proc. IEEE PIMRC 06*, pp. 1-6, 2006.

Liu X., Chong E.K. and Shroff N.B. X. Liu, E.K. Chong, N.B. Shroff, “Opportunistic Transmission Scheduling with Resource-sharing Constraints in Wireless Networks,” in *Proc. IEEE Journal on Selected Areas in Communications,* 2001a.

Liu X., Chong E.K. and Shroff N.B. “Transmission scheduling for efficient wireless resource utilization with minimum-performance guarantees,” in *Proc. IEEE VTC2001 Fall* Vol. 2, pp. 824 - 838. 2001b.

Liu Z. and Zarki M El ”SIR-based call admission control for DS-CDMA cellular systems," IEEE Journal on Selected Areas in Communications, vol. 12, no. 4, pp. 638-644, May 1994.

Liu P., Berry, R. and Honig M. "A Fluid Analysis of a Utility-Based Wireless Scheduling Policy," IEEE Trans. on Information Theory, 2006.

Long F. and Feng G. "Efficient Rate Guaranteed Opportunistic Scheduling for Wireless Networks," *IEEE Transactions on Vehicular Technology*, 2007.

Mandelbaum A. and Stoylar A.L. "Gcμ Scheduling of Flexible Servers: Asymptotic Optimality in Heavy Traffic," Proc. of 2002 Allerton Conference on Communication, Control and Computing, Oct. 2002.

Münz G., Pfletschinger S. and Speidel J. "An efficient waterfilling algorithm for multiple access OFDM", Proc. IEEE International Conference on Global Communications (Globecom'02), November 2002.

Park D., Seo H., Kwon H. and Lee B. "A new wireless packet scheduling algorithm based on the CDF of user transmission rates", Proc. IEEE GLOBECOM, 528-532, 2003.

Parkvall S., Dahlman E., Frenger P., Beming P. and Persson M. "The high speed packet data evolution of WCDMA", Proc. IEEE VTC Spring, 3, 2287-2291, 2001.

Patil S. and Veciana G. de "Managing resources and quality of service in wireless systems exploiting opportunism," *to appear in IEEE Transactions on Networking*, 2007.

Perre L. van der, Theon S., Vandenameele P., Gyselinckx B. and Engels

M. "Adaptive loading strategy for a high speed OFDM-based WLAN," *in Proc. IEEE Global Telecommunications Conference (GLOBECOM '98),*, Vol. 4, pp. 1936-1940, 1998.

Pop and Beaulieu "Limitations of sum of sinsoids fading simulator simulators", IEEE Trans.Commun.,49, 669-708, 2001.

Sanayei S., Nosratinia A. and Aldhahir N. "Code Combining: A maximum-likelihood decoding approach for combining an arbitrary number of noisy packets", *IEEE Trans. on Commun. 1985*, Vol. 33, pp. 593-607, 1985.

Shakkottai S. and Stolyar A.L. "Scheduling for Multiple Flows Sharing a Time Varying Channel: the Exponential Rule", Bell Labaratories, Tech. Rep. 2000.

Shakkottai S. and Stolyar A.L. "Scheduling Algorithms for a Mixture of Real-Time and Non-Real-Time Data in HDR", Proc. Int. Teletrafic Congress, 793-804, 2001.

Shannon C. E. 1948, "A Mathematical Theory of Communication", Bell Syst. Techn. J., Vol. 27, pp. 379-423, 623-656.

Stolyar A.L. "MaxWeight scheduling in a generalized switch: state space collapse and equivalent workload minimization in Heavy Traffic," submitted, 2001.

Stolyar A.L. "On the Asymptotic Optimality of the Gradient Scheduling Algorithm for Multiuser Throughput Allocation", Operations Research, Informs, Vol. 53, No. 1, pp 12-25, 2005.

Tassiulas L. and Ephremides A. ”Dynamic server allocation to parallel queue with randomly varying connectivity”, in IEEE Transactions on Information Theory, Vol. 39, pp. 466-478, March 1993.

Tse D. ”Optimal power allocation over parallel Gaussian channels,” *in Proc. of IEEE* International Symposium on Information Theory, Ulm, Germany, June 1997.

Tse D. and Hanly S. “Multiaccess Fading Channels-Part I: Polymatroid Structure, Optimal Resource Allocation and Throughput Capacities,” *IEEE Transactions on Information Theory*, No. 7, 1998.

Van Mieghem J.A. ”Dynamic Scheduling with Convex Delay Costs: the Generalized cμ Rule,” Annals of Applied Probability, 5(3), 1995.

Viswanath P., Tse D. and Laroia R. ”Opportunistic beamforming using dumb antennas,” IEEE Trans. on Information Theory, vol. 48, no. 6, pp. 1277.1294, June 2002.

Vukadinovic V. and Drogou E. 2006, ”Opportunistic Scheduling in Wireless Networks”, Project report, Royal Institute of Technology.

Willink T.J. and Wittke P.H. ”Optimization and performance evaluation of multicarrier transmission,” IEEE Trans. Inform. Theory, vol. 43, pp. 426-440, Mar. 1997.

Wu ”QoS Provisioning in Wireless Networks,” Wireless Communications and Mobile Computing, vol. 5, pp. 957–969, 2005. Wiley.

Zhang Z., He Y. and Chong E.K.P. “Opportunistic downlink scheduling for multiuser OFDM systems," In *Proc. IEEE Networking Conference 2004* Vol. 2 pp. 1206 - 1212, 2005.

Zhao P. and Zhang H.M. "Sliding window based CAC for adaptive service in mobile network", Proc IEEE PIMRC 2002 5, 2165 - 2169.