TKK Dissertations 241
Espoo 2010

OPPORTUNISTIC PACKET SCHEDULING ALGORITHMS FOR BEYOND 3G WIRELESS NETWORKS

Doctoral Dissertation

Mohammed Al-Rawi

Aalto University
School of Science and Technology
Faculty of Electronics, Communications and Automation
Department of Communications and Networking

TKK Dissertations 241
Espoo 2010

OPPORTUNISTIC PACKET SCHEDULING ALGORITHMS FOR BEYOND 3G WIRELESS NETWORKS

Doctoral Dissertation

Mohammed Al-Rawi

Doctoral dissertation for the degree of Doctor of Science in Technology to be presented with due permission of the Faculty of Electronics, Communications and Automation for public examination and debate at the Aalto University School of Science and Technology (Espoo, Finland) on the 3rd of December 2010 at 12 noon.

Aalto University
School of Science and Technology
Faculty of Electronics, Communications and Automation
Department of Communications and Networking

Aalto-yliopisto
Teknillinen korkeakoulu
Elektroniikan, tietoliikenteen ja automaation tiedekunta
Tietoliikenne- ja tietoverkkotekniikan laitos

Distribution:
Aalto University
School of Science and Technology
Faculty of Electronics, Communications and Automation
Department of Communications and Networking
P.O. Box 13000 (Otakaari 5)
FI - 00076 Aalto
FINLAND
URL: http://comnet.tkk.fi/
Tel. +358-9-470 22353
Fax +358-9-470 22345
E-mail: mohammed.alrawi@tkk.fi

ISBN 978-952-60-3374-7
ISBN 978-952-60-3375-4 (PDF)
ISSN 1795-2239
ISSN 1795-4584 (PDF)
URL: http://lib.tkk.fi/Diss/2010/isbn9789526033754/

TKK-DISS-2808

Aalto-Print
Helsinki 2010

Aalto University

ABSTRACT OF DOCTORAL DISSERTATION	AALTO UNIVERSITY SCHOOL OF SCIENCE AND TECHNOLOGY P.O. BOX 11000, FI-00076 AALTO http://www.aalto.fi

Author Mohammed Al-Rawi

Name of the dissertation
OPPORTUNISTIC PACKET SCHEDULING ALGORITHMS FOR BEYOND 3G WIRELESS NETWORKS

Manuscript submitted 1 April 2010	Manuscript revised 5 August 2010
Date of the defence 3 December 2010	
☒ Monograph	☐ Article dissertation (summary + original articles)

Faculty	School of Science and Technology
Department	Communications and Networking
Field of research	S016Z Communications Engineering
Opponent(s)	Professor Mikko Valkama
Supervisor	Professor Riku Jäntti
Instructor	

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. Namely, the task of the radio resource manager is to deliver different levels of quality for these multimedia services. Radio resources are scarce and need to be shared by many users. This sharing has to be carried out in an efficient way avoiding, as much as possible, any waste in resources.

A Heuristic scheduler for SC-FDMA systems is proposed where the main objective is to organize scheduling in a way that maximizes a collective utility function. The heuristic is later extended to a multi-cell system where scheduling is coordinated between neighboring cells to limit interference. Inter-cell interference coordination is also examined with game theory to find the optimal resource allocation among cells in terms of frequency bands allocated to cell edge users who suffer the most from interference.

Activity control of users is examined in scheduling and admission control where in the admission part, the controller gradually integrates a new user into the system by probing to find the effect of the new user on existing connections. In the scheduling part, the activity of users is adjusted according to the proximity to a requested quality of service level.

Finally, a study is made about feedback information in multi-carrier systems due to its importance in maximizing the performance of opportunistic networks.

Keywords 3G, opportunistic networks, scheduling, OFDMA, LTE, SC-FDMA, ICIC, Nash , RB, admission	
ISBN (printed) 978-952-60-3374-7	ISSN (printed) 1795-2239
ISBN (pdf) 978-952-60-3375-4	ISSN (pdf) 1795-4584
Language English	Number of pages 176
Publisher Aalto University Library	
Print distribution School of Science and Technology	
☒ The dissertation can be read at http://lib.tkk.fi/Diss/2010/isbn9789526033754	

Preface

This thesis consists of research work that has been carried out at the School of Science & Technology of Aalto University (Formerly known as Helsinki University of Technology) and the University of Vaasa. Various sources have funded this work and I would like to gratefully acknowledge their contribution starting with the Tekniikan tutkimusinstituutin (TTI) of Vaasa, the Nomadic lab of Ericsson Research, Nokia-Siemens Networks and the Finnish Funding Agency for Technology and Innovation (TEKES).

I would also like to acknowledge the communications and networking department at Aalto University and the department of computer science at the University of Vaasa for facilitating the necessary resources to conduct this work. 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. It was a pleasure and a great honor to have him as a supervisor. I would like to thank him for accepting me and giving me the opportunity to attain my doctorate degree.

By finishing this thesis I can't help but remember my father who finished his phd back in the eighties with limited resources, the sound of his modest typewriter still rings in my ears. His determination and hard work was an inspiration for me to continue my studies. I would like to thank my brother Yasir who supported me during my Master's studies, completing this D.Sc. would have not been possible without his help. Finally, I would like to thank my wife for her support and patience.

Mohammed Al-Rawi
Espoo, September 2010

Contents

Preface iii

Contents v

Author's contribution viii

List of Abbreviations x

List of Symbols xiii

1 Introduction 1

1.1 General Models and Assumptions 3

1.2 The Radio Channel 5

1.2.1 Channel Model 8

1.2.2 Retransmissions 9

1.3 Multi-Carrier Model 10

1.4 Access Technologies 12

1.5 Simulator Environment 14

1.6 Objectives 14

1.7 Organization of the Thesis 15

2 Background 17

2.1 Scheduling 17

2.2 Previous work 22

2.2.1 Opportunistic Schedulers 22

2.2.2 Inter-Cell Interference Coordination 24

2.2.3 Game Theory in Scheduling 25

2.2.4 Opportunistic Admission Control 26

2.2.5 Opportunistic Rate Control 29

2.2.6 Multi-carrier Systems 30

3 Utility Based Scheduling **35**

3.1 Single-cell Multi-carrier Scheduling 35

3.1.1 Localized Gradient Algorithm LGA 35

3.1.2 Heuristic Localized Gradient Algorithm HLGA 37

3.1.3 Computational Evaluation 39

3.2 Multi-cell Multi-carrier Scheduling 48

3.2.1 Heuristic Scheduling 48

3.2.2 Optimal Scheduling 49

3.2.3 Computational Evaluation 51

3.3 Concluding Remarks 56

4 Bargaining Based scheduling **59**

4.1 System Model 59

4.2 Intra-cell Scheduling 60

4.3 Inter-cell Coordination 62

4.3.1 Nash Bargaining 62

4.3.2 Load Balancing Handovers 65

4.3.3 Joint Nash bargaining and load balancing 66

4.3.4 Bargaining objective function 67

4.4 Computational Evaluation 68

4.5 Concluding Remarks 73

5 Activity control **77**

5.1 Admission Control 77

5.2 Single-user Iterative Admission Control 79

5.2.1 Implementation in Memoryless Schedulers 79

5.2.2 Implementation in Schedulers with Memory 80

5.2.3 Iterative Admission Control 81

5.2.4 Non-stationarity 82

5.3 Multi-user Iterative Admission Control 85

5.4 Kalman Filter Estimation 87

5.5 Non-iterative Admission Control 88

5.6 Computational Evaluation 89

5.6.1 Static Traffic (Full buffer) 89

5.6.2 Dynamic Traffic 92

5.6.3 Decision Error . . . 94
5.6.4 Multi-user Admission Control . . . 94
5.6.5 Non-iterative Admission Control . . . 95
5.7 Quality Control . . . 97
5.7.1 Convergence Analysis . . . 97
5.7.2 Example . . . 99
5.8 Imperfect Estimates . . . 101
5.9 Computational Evaluation . . . 103
5.10 Concluding Remarks . . . 105

6 Feedback in Multi-Carrier Systems 109
6.1 System Model . . . 110
6.2 Feedback Based on Rank Ordering . . . 113
6.3 Numerical Analysis . . . 114
6.3.1 Decision Variable Based on Rank Ordering . . . 114
6.3.2 Comparison of Different Decision Variables . . . 115
6.3.3 Multiple Chunks . . . 116
6.3.4 Effect of Number of Feedback Bits . . . 116
6.4 Impact of Feedback Information Accuracy . . . 119
6.4.1 General Form for SNR Distribution . . . 121
6.4.2 Probability of a Correct Scheduling Decision . . . 122
6.4.3 Two-User Case . . . 124
6.5 Concluding Remarks . . . 129

7 Conclusion 131

References 134

Appendix A Validity of models A-1

Appendix B Proofs B-1

Author's contribution

The results of this thesis serve to shed more light on scheduling in opportunistic systems which is an integral part in beyond-third-generation wireless communication systems. The contributions of this work are included in Chapters 3-6 of the thesis. The chapters were based on submitted and published conference, journal papers and reports. The following is an overview of the contributions.

Chapter 3 is based on the following conference papers: i) "Opportunistic Uplink Scheduling for 3G LTE Systems" published in the proceedings of the 4th IEEE Innovations in Information Technology (Innovations07), 2007, ii) "On the Performance of Heuristic Opportunistic Scheduling in the Uplink of 3G LTE Networks" published in the proceedings of the International Symposium on Personal, Indoor and Mobile Radio Communications PIMRC 08, iii) "Channel-Aware Inter-Cell Interference Coordination for the Uplink of 3G LTE Networks" published in the proceedings of the Wireless Telecommunications Symposium WTS 09 all by M. Al-Rawi, R. Jäntti, J. Torsner and M. Sågfors.
The author proposed and introduced the HLGA algorithm as well as provide additional analysis for different traffic scenarios and verified the optimal solution. The author also revised the analysis for the coordination algorithms in the multi-cell case and provided all the necessary numerical results.

Chapter 4 is based on he conference paper "Uplink Inter-Cell Interference Coordination by Nash Bargaining for OFDMA Networks" by M. Al-Rawi, R. Jäntti submitted to the IEEE 72nd Vehicular Technology Conference (VTC'10).
In this work the author provided the background information for the system model, verified the analysis and algorithms as well as provide the numerical results.

Chapter 5 is based on the journal paper "Call Admission Control with Active Link Protection for Opportunistic Wireless Networks" by M. Al-Rawi and R. Jäntti published in the journal of Telecommunication Systems (Springer) in 2009 and the conference paper "Opportunistic Best-effort Scheduling for QoS-aware Flows" by M. Al-Rawi and R. Jäntti published in the proceedings of the 17th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '06), pp. 1-5, 2006.
The author constructed the admission control simulator providing the numerical

results. The author verified the assumptions made for the model and revised the equations as well as simulate the RLS admission controller to compare with the proposed algorithm. The author provided the theoretical analysis for the convergence of the algorithm in quality control as well as carrying out the necessary simulations.

Chapter 6 is based on the conference paper "On the block-wise feedback of channel adaptive multi-carrier systems," by R. Jäntti and M. Al-Rawi published in Proc. 65th IEEE Vehicular Technology Conference, (VTC2007-Spring), pp. 2946 - 2950, 2007 and the draft manuscript "Analysis of a Practical VOIP Scheduling" by M. Al-Rawi and J. Hämäläinen.
The author's contribution was to verify the algorithms and discover the methods that would minimize overall feedback while maintaining an acceptable performance. The author verified and corrected some of the equations as well as verify the numerical results through simulations and provide the analysis for the two-user case.

List of Abbreviations

2G	Second Generation
3G	Third Generation
3GPP	Third Generation Partnership Project
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
CS	Cumulative-based Scheduler
CSI	Channel State Information
D-FDMA	Distributed-Frequency Division Multiple Access
DS-CDMA	Direct Spread-Code Division Multiple Access
DL	Downlink
EVDO	Evolution-Voice and Data Optimized
FATB	Fast Adaptive Transmission Bandwidth
FDMA	Frequency Division Multiple Access
FDPS	Frequency Domain Packet Scheduler
FFT	Fast Fourier Transform
GIR	Gain to Interference Ratio
IIA	Independent from Irrelevant Alternatives
GSM	Global System for Mobile Communications
HARQ	Hybrid Automatic Retransmission reQuest
HDR	High Data Rate
HLGA	Heuristic Gradient Algorithm
HOL	Head of Line
HSDPA	High Speed Downlink Packet Access
IEEE	International Electrical and Electronic Engineers
IFFT	Inverse Fast Fourier Transform

IP	Internet Protocol
IS-95	Interim Standard 95
ISP	Internet Service Provider
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

UL	Uplink
VoIP	Voice over Internet Protocol
WCDMA	Wideband Code Division Multiple Access

List of Symbols

Introduction & Background

$\mathcal{A}$	Set of active users that have data in their transmission buffer
C	Throughput capacity of the radio channel
$C_{i,n}$	Throughput capacity of user i on resource block n
B	Bandwidth size of the radio channel
d_{ij}	Pyhsical distance between terminals i and j
$\mathbb{E}$	The Expectation operator
$\mathcal{F}$	Cumulative distribution function
$\frac{E_b}{N_0}$	Bit Energy to Noise power ratio
f	Frequency value
f_m	Maximum Doppler frequency shift
F	An objective function that requires maximization
$\mathbb{F}$	Cumulative distribution function
g	Switch state
G_{ij}	Link gain between transmitter j and receiver i
G_a	Directional antenna gain
h	Time response of the channel
H_k	Frequency response of subcarrier k
i^*	User selected for transmission
I_0	Zeroth order Bessel function of the first kind
γ_i	Signal-to-Interference+Noise ratio
l	Subcarrier index
L	Total number of resource blocks
m	Signal path index
M_p	Total number of signal paths
n	Resource block index
N	Number of users in the system
p	Probability density function
P_i	Transmission power of user i
$\bar{P}_i$	Average Transmission power of user i
$P_{\max}$	Maximum transmission power allocated to a user
q	Scheduling decision
Q	Set of feasible scheduling decisions

r	Number of retransmissions
R	Total number of rertansmissions
s	Switch state of the system in Stolyar's framework
t	Time
t_m	Time delay of path m
T_m	Largest delay in a fading channel
U	A utility function
$\bar{x}$	Average data throughput
$\bar{x}_i^{min}, \bar{x}_i^{max}$	Minimum and maximum constraints on the throughput of user i
$\bar{X}$	Vector of averaged throughputs
$\bar{Z}$	Average statistics of channel conditions
α	Fairness index in the α-fair scheduler
β	A positive value larger than 0 and less than 1
δ	The Dirac delta function
Δt	Time delay at the maximum Doppler shift
Δt_c	Channel coherence time
Δf_c	Channel coherence bandwidth
ϵ	System data unit error ratio at the RNC level
ι	Loss resulting from combining retransmissions of replicas of the signal
κ	Attenuation factor of the link gain which normally assumes values $2 \sim 4$
t	Time slot index
μ_i	Data rate of user i
μ_i^{req}	Requested data rate of user i
θ	Orthogonality factor between different waveforms
Θ_H	Autocorrelation function of Radio channel H
ϕ	Beam angle of a directional antenna
υ	Thermal noise power at the receiver
ξ_i	Channel condition of user i
$\chi\{A\}$	The indicator function; $\chi\{A\}$=1 if Event A happens and 0 otherwise
ζ	Slow fading component in the link gain of the channel
π	Mathematical constant equal to 3.14
σ^2	Local mean power of the received signal
$\bar{\sigma}$	Average value of the received signal
ρ	Slow fading component
Ω	Total number of subcarriers

∇	Gradient operator

Uplink Scheduling

a	Selection variable for the optimal number of edge bands
b	Base-station index
B	Number of base-stations
$\mathcal{B}$	Set of base-stations coordinating with each other
c	Slack variable of the lagrangian
d_i	Disagreement point for cell i in the bargaining process
f	Variable indicating edge bands in bargaining
F	An objective function that requires maximization
F_b	Objective function for cell b
g	State of the system at a specific time
i	User index
$\mathcal{I}_i$	Set of resource blocks that cannot be allocated to user i
$\vec{j}$	Vector or users
$\mathcal{J}_{-b}$	Vector containing users in all cells except b
k	Iteration index
K	Frequency reuse factor
$\mathcal{K}_b$	Set of possible edge resource block allocations for base-station b
l	Number of bands allocated to edge-users
$l_{min,i}$	Minimum number of possible RBs to the edge-users of BS i
$l_{max,i}$	Maximum number of possible RBs to the edge-users of BS i
L	Number of resource blocks in the system
$\mathcal{L}_i$	Set of candidate resource blocks that can be allocated to user i
m, n	Resource block index
N	Number of users in the system
N_c	Number of center users
N_e	Number of edge users
N_G	Number of groups bargaining with each other
$\mathcal{N}$	Set of system users
$\mathcal{N}_b$	Set of users in cell b
$\mathcal{N}_c$	Set of the center users
$\mathcal{N}_e$	Set of the edge users

$\mathcal{N}_E$	Set containing edge users of all coordinating cells
q	Resource block assignment decision
Q	All feasible resource block assignments
$\mathcal{R}$	set of users that have retransmitted resource blocks
s	Variable indicating edge bands in bargaining
S	Sum of edge bands in all coordinating cells
t	Time slot index
T_{slot}	Duration of the time-slot
U	Utility function
U_b	Utility function of cell b
v	Marginal utility resulting from the difference between the current selection of edge bands and the disagreement number of edge bands
V	Sum of objective functions of all coordinating cells
w	Variable indicating edge bands in bargaining
W_i	Queue size of the transmission buffer of user i
$\bar{x}$	Average data throughput
$\bar{x}_{min}$	Minimum achievable average rate
$y_{i,n}$	Resource block selection variable for user i on RB i; $y = 1$ if RB n is assigned to user i and 0 otherwise
$\mathcal{Z}_i$	Set of resource blocks assigned to user i
$\tilde{\mathcal{Z}}_i$	Set of resource blocks that are forcefully assigned to user i due to the contiguity constraint
$\mathcal{Z}_i^r$	Set of resource blocks that need to be retransmitted for user i
α	Fairness index in the α-fair scheduler
λ	Index of a resource block that cannot be assigned at the same time in neighboring cells
μ_i	Data rate of user i
$\mu_{i,n}$	Data rate of user i obtained from resource block n
$\mu_{i,j,n}$	Data rate of user i of cell 1 with RB n when user j of cell 2 is utilizing the same RB.
$\mu_{i,0,n}$	Data rate of user i of cell 1 with RB n when the same RB is free in cell 2.
ϕ_i	Fraction of resources allocated to edge user i
Φ	Set of all neighboring cells to a particular cell
φ_i	Fraction of resources allocated to center user i
ψ	Marginal utility resulting from the difference with consecutive values for

	the edge bands
ρ	Slope defining the order of the marginal utilities with different numbers of edge bands
η	Lagrangian parameter
ϱ	Lagrangian parameter
τ	Resource block representing a gap in an allocated spectrum

Activity Control

A_k	Number of users in subset k
$\mathcal{A}$	Set of active users that have data in their transmission buffer
$\mathcal{A}_l$	Set where no user is active
$\mathcal{A}_S$	Set where all users are active
$\mathcal{A}_N$	Set of active users that have data in their transmission buffer with N users in the system
B	A vector defining the tuning factors for quality control
C	Kalman filter output mapping vector
e	Measurement error resulting from the difference between the expected and actual values
$\mathbb{E}$	The Expectation operator
F	An objective function that requires maximization
h	Back-off probabilities for each of the multiple users seeking admission
i	User index
$i*$	User selected for transmission
j	Active user index
J	Vector of the sum of average throughputs for user i with different
k	Subset index of active users
K	Filter gain
$\mathcal{K}_i$	Selection of active sets that contain user i
M	Number of multiple users seeking admission
m	User index
n	Frame index
N	Number of users in the system
$\mathcal{N}$	Set of users in the system
	number of users in the system

p_b	Back-off probability in admission control
P	Error covariance matrix (A measure of the estimated accuracy of the state estimate)
q	Activity probability defining whether a user is active or not
Q	Activity probability vector
$\mathcal{Q}$	Set of feasible activity probabilities
R	Matrix form of the multiple admission control scheme
$\tilde{s}$	Expected Average QoS
$\bar{s}_i$	Average QoS of user i
$\bar{s}^{req}$	Requested average QoS by user i
$\hat{S}_i$	Expected QoS state vector of the Kalman filter
t	Time slot index
t_0	Time-slot at which a new user tries to join the system
$\mathcal{T}$	A mapping between the activity probabilities and the thinning gains
u	Input vector for the RLS algorithm
V_i	State noise of the Kalman filter
w	Filter weight
$\bar{x}$	Average data throughput
$\bar{x}_i(Z_N)$	Average data throughput of user i with N users in the system having channel statistics defined by matrix Z
$\tilde{x}$	Expected Average data throughput
$\bar{x}_{lower}$	Lower bound for the average rate
$\bar{x}_{upper}$	Upper bound for the average rate
$\bar{x}_{min}$	Threshold for the minimum acceptable throughput of active users in the system
X_i	State vector of the Kalman filter defining the throughput of user i
$\hat{X}_i$	Expected rate state vector of the Kalman filter
$\mathcal{X}_i$	Vector of average throughputs for user i with different number of users in the system
y	A continuously differentiable function representing the quality control formula at different points in time
Y	Square matrix consisting back-off probabilities
z	Non-active user index
Z_N	Statistics Matrix of channel conditions for N users
$\mathcal{Z}_N$	Set of fading vectors of length N

β	A positive value larger than 0 and less than 1
κ, ρ	Positive constants
ϵ	Estimation error
λ	Forgetting factor in the RLS algorithm
μ_i	Data rate of user i
$\tilde{\mu}$	Expected rate
π_k	Probability that set k was used at a particular time
ψ	A positive value that it larger or equal to 0 and less than 1
Ψ_{ee}	Covariance matrix for the measurement noise in the Kalman filter
Ψ_{Ve}	Cross-covariance matrix for the state and the measurement noise Kalman filter
Ψ_{VV}	Covariance matrix for the state noise Kalman filter
$\chi\{L\}$	The indicator function; $\chi\{L\}$=1 if Event L happens and 0 otherwise
$\urcorner_i\{L\}$	An indicator function; $\urcorner_i = 1$ if i provides the maximum value for L and 0 otherwise
ϕ	Channel access fraction
∇	Gradient operator

Feedback in Multi-Carrier Systems

A	Event: SNR of the first estimated channel is larger than the second
A^c	Event: SNR of the first estimated channel is smaller than the second
B	Event: SNR of the first actual channel is larger than the second
B^c	Event: SNR of the first actual channel is smaller than the second
b_f	Number of bits used for feedback
B_W	Bandwidth size of the radio channel
BEP	Bit error probability
BEP_S	Bit error probability when no scheduling is applied or scheduling is random
BEP_{min}	Minimum bit error probability
BEP_{max}	Maximum bit error probability
C	Throughput capacity of the radio channel
$C_{\max}$	Maximum throughput capacity
C_S	Throughput capacity when no scheduling is applied or scheduling is random
d	Positive constant
D_i	Packet time delay for user i

$\mathbb{E}$	The Expectation operator
e	Exponential natural value
$\mathcal{E}_1$	The exponential integral
f	Probability distribution function
f_{max}	PDF for maximum SNR
f_{min}	PDF for minimum SNR
f_S	SNR distribution when scheduling is not applied
$\mathbb{F}$	Cumulative distribution function
g	A joint distribution that is a function of the true and estimated channels
h	Time response of the channel
H	Frequency response of the channel
$\mathcal{I}$	Notational help
I_0	Zeroth order Bessel function of the first kind
K	Number of groups of resource blocks
n	Resource block index
N	Number of users in the system
L	Number of resource blocks in the system
P	Probability of a wrong scheduling decision
P^C	Probability of a correct scheduling decision
Q	Error function
r	Absolute value of the actual channel
R	Absolute value of the estimated channel
t	Time slot index
t_0	Time when a code block was first transmitted
x_i	Data throughput of user i
Z	Reported channel state information
$Z_{L,n}$	Channel state of the n^{th} smallest resource block in a group containing L blocks
δ	The Dirac delta function
Δf_c	Channel coherence bandwidth
ϵ	error from channel estimation
η	Signal-to-noise ratio
$\bar{\eta}$	Average signal-to-noise ratio
γ	The incomplete gamma function
Γ_i	The gamma function
μ_i	Data rate of user i

ξ_i	Actual channel state of user i
ρ	Total number of quantization levels
σ	Standard deviations of the underlying Gaussian channel distribution
ν	Power of the estimation error divided by the total mean received power of a resource block
ϕ	Argument between the true channel and the estimation error
Ω	Total number of subcarriers

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 many 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 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 channel SINR signal strength should be maximized while the interference part minimized. In the case of non-real time data this could be achieved by scheduling 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 and this thesis aims 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 where 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 opportunistic admission control scheme which 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.

In third generation (3G) telecommunications, the introduction of universal mobile telecommunications system wideband code division multiple access (UMTS-WCDMA) was a major evolutionary step from the second generation (2G) global system for mobile telecommunications (GSM) networks. The possibility of allowing simultaneous connections and separating users with unique codes as well as the co-existence of frequency division duplex (FDD) and time division duplex (TDD) enabled WCDMA to achieve high data rates and accommodate more users [1]. The WCDMA systems' adaptability enabled more improvement by introducing the high speed packet access (HSPA) phase which is the first and one of the significant evolutionary steps in packet data access in 3G telecommunications [2]. The standard consists of two parts: high speed downlink packet access (HSDPA) and high speed uplink packet access (HSUPA). HSDPA is a key feature included in the 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 up to 10 Mbps.

HSUPA is a Release 6 feature in 3rd generation partnership project (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 (TTI) enabling faster link adaptation.
- Hybrid automatic repeat request (HARQ) with incremental redundancy making retransmissions more effective.

The evolution of HSPA systems continued and HSPA+ was introduced in release 7 to offer higher data rates and better support for internet protocol (IP) structures by having the option of an all-IP architecture through connecting base-stations directly to IP based backhauls and then to the internet service provider (ISP) edge routers [3]. This was a very important feature because it facilitated the voice over internet protocol (VoIP) technology.

3GPP Release 8 introduced the long term evolution (LTE) phase, considered an evolutionary step that is paving the way toward 4G mobile communications technology. It is expected that LTE will stretch the performance of the 3G technology and meet the growing demands for resources by users. The fundamental targets of this evolution are to further reduce user and operator costs and to improve service provisioning. This is achieved through improved coverage and system capacity as well as increased data rates and reduced latency [4].

1.1 General 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 task 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 the 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 six 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: this happens as a result of the bouncing of waves from surrounding objects such as buildings and passing vehicles.
- Refraction: is the redirection of a wave passing a boundary between two dissimilar media.
- Diffraction: that results from the bending of waves.
- Scattering: this 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.
- Blocking (Shadowing): this happens when waves are encountered by large obstacles such as walls that create shadow zones.

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 that is calculated from the following equation [5].

$$\gamma_i = \frac{G_{ii}P_i}{\sum_{j\neq i} P_j\theta_{i,j}G_{ij} + \upsilon} \tag{1.1}$$

where γ_i denotes the SINR of user i, G_{ij} represents the link gain of transmitter j at receiver i. P_i is the transmission power of i. υ denotes the (thermal) noise power at receiver i. θ_{ij} denotes the normalized squared cross correlations between waveforms. Channel quality normally defines how many information bits the channel can convey i.e. channel capacity which is proportionally related to channel condition. In 1948 Claude Shannon, an American electrical engineer and mathematician, found that there is a limitation 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 limit is defined in Shannon's formula [6] where a transmission link with a bandwidth B and SINR γ will have the following limit:

$$C = B \log_2 (1 + \gamma) \text{ (bps)} \tag{1.2}$$

1.2 The Radio Channel

Signals that are transmitted are encountered by several factors that decrease their density. Such density is inversely proportional to the distance between transmitter and receiver. 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 [7]:

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 up 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.

In a multi-carrier system, fast fading leads to two types of fading according to the number of paths as follows

- 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.

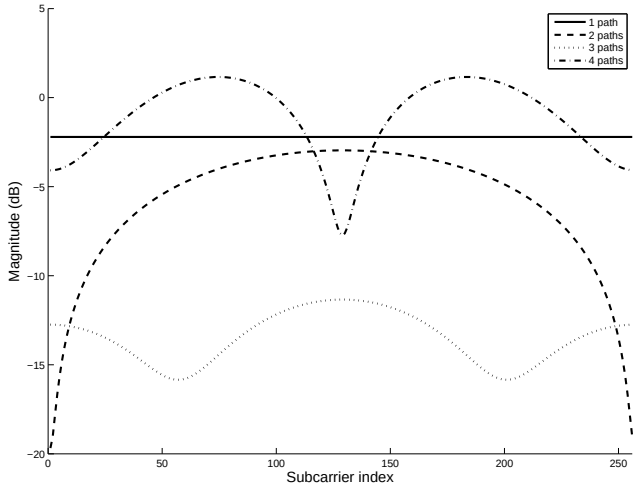

Figure 1.1: Magnitude plot for various number of multipath components

- 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.

The coherence bandwidth measures the minimum separation in frequency after which two signals will experience uncorrelated 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. Fig. 1.1 represents the effect of selective fading in a multi carrier system. One can see that as the number of components increase the frequency selectivity also increases. This shows that multipath makes the channel frequency selective. Figures 1.2 and 1.3 show the response to an example of a multiple path channel in a multi-carrier system. The example shows a 6 tap channel model for a propagation in a typical urban area described in 3GPP Release 7 [8]. 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.

Certain modulation schemes such as OFDM are 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

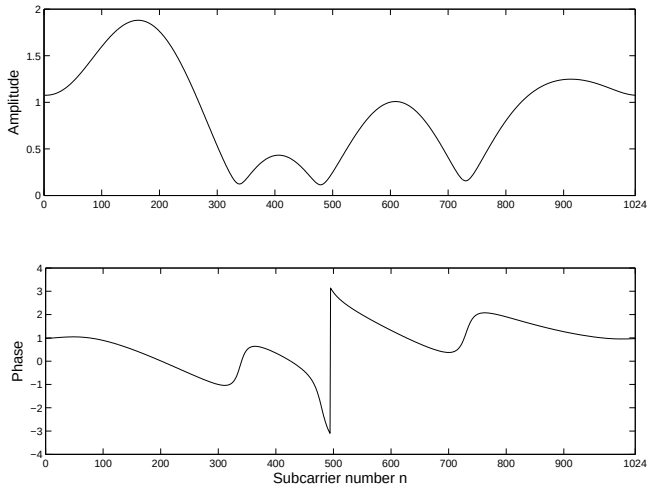

Figure 1.2: Selective frequency channel response

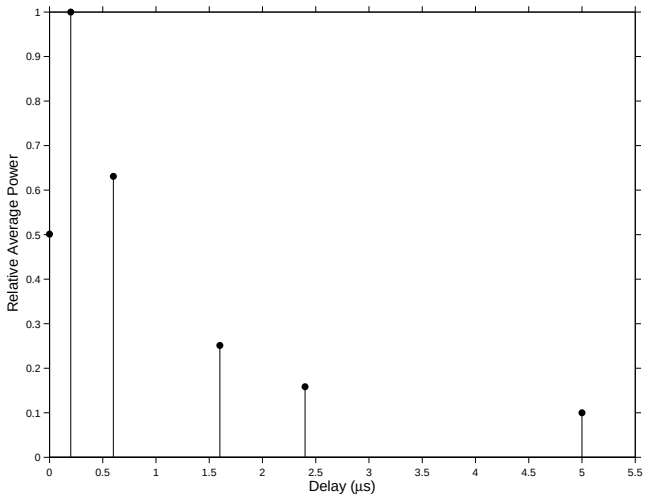

Figure 1.3: Relative powers of the delay profile

rather than frequency selective fading. This can be countered 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.

1.2.1 Channel Model

The channel model will mainly consist of the two components mentioned earlier; the large-scale fading (or slow fading) and the short-scale fading (or fast fading). The gain of a channel with frequency k can be written as follows

$$G_{i,k} = d_{ij}^{-\kappa}.10^{\rho/10}|H_l|^2 \tag{1.3}$$

where d_{ij} is the distance between terminals i and j, α is path loss component which ranges between 2 and 4, ρ is a normal random variable representing the slow fading component. The variable $|H_k|^2$ is the frequency response of the kth subcarrier channel and denotes the fast fading component and is computed from

$$h(t) = \sum_{m=1}^{M_p} h_m \delta(t - t_m), \tag{1.4}$$

where h is the wide-sense stationary channel of the mth path, M_p is the number of paths and t_m is the delay of the mth path. The frequency response of the subcarrier can be expressed by [9]

$$H_l = \int_{-\infty}^{\infty} h(t) e^{-i2\pi ft} dt|_{f=f_l} \tag{1.5}$$

where f_l is the frequency of the lth subcarrier. 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

$$p_{|h|}(\sigma) = \frac{1}{\sigma^2} \exp\left(-\frac{P}{\sigma^2}\right) \tag{1.6}$$

where P is the instantaneous power and σ^2 is the local mean power at the receiver. Hence, the envelope power $|H|^2$ will follow an exponential distribution. In this work

Jakes' model for Rayleigh fading was considered [10]. Jakes' model is based on the summing of several sinusoids. The normalized autocorrelation function of a Rayleigh faded channel with motion at a constant velocity at delay Δt, when the maximum Doppler shift is f_m, is a zeroth-order Bessel function of the first kind;

$$\Theta_H(\Delta t) = I_0(2\pi f_m \Delta t) \tag{1.7}$$

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 $\Delta t_c = \frac{1}{f_m}$ is called the *coherence time*. Each channel remains strongly correlated during this time. Analogously, the *coherence bandwidth* Δf_c 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, $\Delta f_c = \frac{1}{T_m}$, where T_m is the largest delay produced by the channel.

1.2.2 Retransmissions

Perfect channel estimation requires timely knowledge of the channel state. In practice, the selection of the rate would be 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. For the algorithms in this thesis that utilize retransmissions, synchronous non-adaptive hybrid automatic retransmission requests (HARQ) are considered. 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 may have changed. The benefit with synchronous non-adaptive HARQ is that 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, the receiver coherently combines the original code block and the retransmitted block. All the transmitted bit energy can be harnessed at 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

$$\frac{1}{\iota}\sum_{r=1}^{R}\left(\frac{E_b}{N_0}\right)_r \geq \left(\frac{E_b}{N_0}\right)_0, \tag{1.8}$$

where $r = 1, 2, \cdots, R$ is the transmission number with R being the maximum allowed number of retransmissions, $\iota > 1$ denotes the combining loss, $(\frac{E_b}{N_0})_r$ is the bit energy-to-noise ratio at the time of transmission and $(\frac{E_b}{N_0})_0$ 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 $\frac{\mu(t)}{r}$. Decoding fails if this rate still exceeds the capacity of the channel.

1.3 Multi-Carrier Model

The bandwidth B consists of Ω subcarriers that are grouped into $L = \frac{B}{\Delta f_c}$ subbands or what will henceforth be referred to as resource blocks (RB)s, shown in Fig. 1.4 with Δf_c denoting the coherence bandwidth of the channel. Each RB will contain Ω/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. The models considered in this work do not utilize power control so it is assumed that the amount of power available for every user will be constant and is represented by the maximum transmission power P_{max}. The capacity of RB n for user i at TTI t is given by Shannon's formula:

$$C_{i,n} = B_n \log_2(1 + \gamma_{i,n}(t)) \tag{1.9}$$

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

$$\gamma_{i,n} = \frac{P_{i,n}.G_{i,n}.G_a(\phi_0)}{\upsilon}, \;\; P_{i,n} = \frac{P_{max}}{L_i} \tag{1.10}$$

where $P_{i,n}$ is the amount of power allocated to RB n for user i, $G_{i,n}$ is the path gain for that subcarrier, $G_a(\phi_0)$ is a ϕ_0 degrees directional antenna gain, υ is the

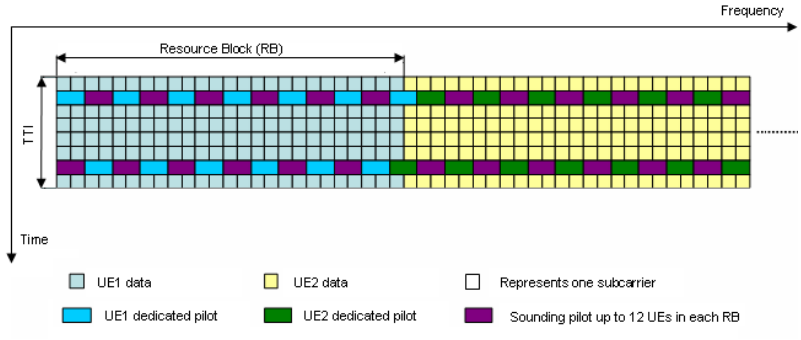

Figure 1.4: Sounding for the uplink channel in LTE [11]

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. It is assumed that all RBs consist of equal numbers of subcarriers Ω_n yielding equal sized RBs.

In a multi-cell scenario, (1.10) becomes

$$\gamma_{i,n} = \frac{P_{i,n}.G_{i,i}.G_a(\phi_0)}{\sum_{j=1,j\neq i}^{N} P_{j,n}.G_{i,j}.G_a(\phi_0) + \upsilon}, \quad P_{i,n} = \frac{P_{max}}{L_i} \tag{1.11}$$

where $G_{i,i}$ is the path gain between user i and base-station (BS) i, $P_{j,n}$ is the interfering power from user j utilizing RB n, $G_{i,j}$ represents the path gain between user j and BS i, and N is the total number of transmitters assigned to RB n.

The assumption of utilizing equal power allocation for the users in (1.10) and (1.11) is justified in single-cell systems by the fact that their users are separated in the frequency domain and thus do not cause interference. In a multi-cell model, users of neighboring cells located near the base-station (center-users) have a high link gain compared to the interference they experience from center-users or edge-users in other cells. The main concern would be the edge-users who will suffer from neighboring edge-users who transmit on the same frequency and this is where inter-cell interference coordination should take place.

Sounding

Since the channel-dependent scheduling can only be applied to low-speed UEs, usually the localized RB is assigned to transmit the traffic. In the case of localized data transmission the reference signal is also localized. This means that the reference signal occupies the same spectrum as data transmission in two short blocks. Only

one sounding pilot is required for each user equipment (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 [11]. Fig. 1.4 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 that provide the whole-band channel information for 12 channel-dependent scheduling UEs in each subframe.

1.4 Access Technologies

The access system considered in this thesis is the TDD system with access technologies ranging between 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 Single Carrier Frequency Multiple Access (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 is a channel access method for shared medium (usually radio) networks. It enables 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), thus improving transmission efficiency. In Frequency Division Multiple Access 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 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 other extension; Single Carrier Frequency Multiple Access (SC-FDMA) utilizes single carrier modulation at the transmitter and frequency domain equalization at the receiver. This technique has similar performance and essentially the same overall structure as 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 Peak to Average Power Rate (PAPR) greatly benefits the mobile terminal in terms of transmit power efficiency.

Code division Multiple Access 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 HSDPA and LTE technologies are both considered. HSDPA as mentioned in Chapter 1 is an evolutionary step of WCDMA where a time dimension is added to the access function. Adding the time property exposes the possibility of utilizing the opportunistic features of channels.

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. 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. It therefore, offers the advantage of frequency diversity. Moreover, 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.

1.5 Simulator Environment

A computer simulator is used to create a single or multi-cell packet switched network. The simulator is a quasi stationary simulator that generates N users with locations uniformly distributed over the cell area. A pedestrian profile is assumed for the the speed of the users, hence channel conditions are slowly changing and the channel is assumed to be constant during one TTI. Different users experience different channel conditions that vary depending on their distance from the base-station and speed. The speeds of the mobile users are independent random variables uniformly distributed between 3 km/h and 10 km/h. Mobility induced handovers are not considered but fast fading is simulated and 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.

The traffic models that are considered are the saturated static traffic model and the dynamic model where packets arrive according to a certain distribution. The algorithms presented in this work are mainly designated for non-real time traffic but still can be applied to real-time traffic with certain delay bound constraints.

1.6 Objectives

The objective of this work is to analyze and develop new scheduling algorithms for advanced opportunistic wireless systems. 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 service for users in cellular systems and the mechanisms to maintain that quality of these services.

The study underlying this thesis is an attempt to take a broader view of scheduling in advanced cellular communication systems that can adapt to channel condi-

tions, especially for the new long term evolution phase. The main challenge in this task is the constraints of the interface where scheduling of resources is not as flexible as other interfaces. The problem expands in multiple cells where interference becomes a factor that impacts scheduling heavily. Coordinating the transmission scheduling of neighboring cells therefore becomes a necessity. However, coordination needs to be done in a way that allows resources to be utilized to the utmost without resorting to unfair division between good and bad users.

Another issue is the concept of combining opportunistic scheduling with admission control in fast fading channels. Having a fast varying channel can consequently lead to poor admission decisions if not dealt with properly. Opportunistic scheduling algorithms also tend to provide a certain quality of service that is limited by the amount of access to the channel. Therefore, finding a mechanism that can provide a multi-level QoS is something that would be interesting to pursue.

Feedback information plays an important role in opportunistic scheduling, having the right information at the right time maximizes performance considerably. In multi-carrier systems, reporting the channel condition of every subcarrier will result in excessive overhead. Therefore, there is a need to find a suitable way to report the information back to the scheduler in a reasonable manner without severely degrading system performance.

1.7 Organization of the Thesis

Chapter 1 introduces the models and assumptions used throughout the thesis. Chapter 2 presents the necessary background information for the work. Chapter 3 proposes the LTE single-cell scheduler and multi-cell scheduler. Chapter 4 discusses the effect of bargaining on multi-cell systems. Chapter 5 studies the concept of activity control in opportunistic systems. Chapter 6 investigates the different aspects of feedback information and their impact on system performance. Finally, in chapter 7 the conclusion of this work is drawn with some discussion. In addition, the thesis contains an appendix of proof for a number of propositions as well a validation of the used models by comparing them to models from the literature.

Chapter 2

Background

2.1 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 a certain utility function. Schedulers can 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. It 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 as well as being starvation-free. In channel dependent schedulers, the scheduler forms the decision based on feedback information from the terminal (downlink) and from the Base-Station (BS) (uplink). The advantage of channel dependent scheduling is the exploitation of channel fluctuations, i.e. by assigning resources to a user who benefits the most from using them.

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 taken from Knopp and Humblet who proposed that the best strategy is to always transmit to the user with the best channel for the uplink [12] . Tse provided a similar

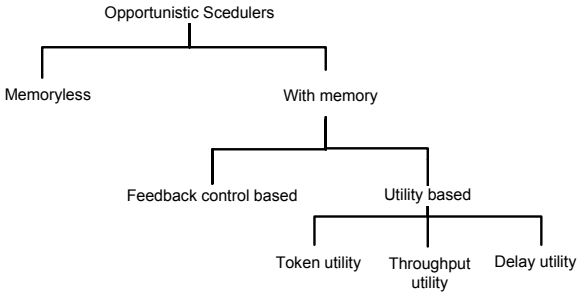

Figure 2.1: Scheduler classes

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 [13]. 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.* examined practical aspects of downlink multi-user diversity in the context of the IS-95 CDMA standard [14]. Viswanath, Tse and Laroia examined this problem for the downlink and presented a method of opportunistic beamforming via phase randomization [15].

Opportunistic schedulers are divided into a number of categories as shown in Fig. 2.1. The schedulers are mainly 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

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

   where $\xi_i(t)$ denotes the channel condition of user i at time and $\mathcal{A}(t)$ is the set of active users who have data in their buffers at time t t.

2. **Schedulers with memory**: the 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 frames, 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.

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

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 throughput.

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.

To implement the idea of opportunistic scheduling, two issues need to be addressed: fairness and the service requirements of users. 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 greatly unequal 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 exploit 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 in a short timescale as well as 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 much attention [16].

Several scheduling rules have been introduced in the literature. The maximum Carrier to Interference Ratio (max-CIR) rule always selects the user having the highest CIR [17]. This rule maximizes system throughput, but leads to very unfair allocation 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 [13, 18], which utilizes the instantaneously achievable service rate divided by the average throughput as a decision metric. 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 [19]. 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. 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 [20, 21, 22]. The gradient algorithm 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 [23]. 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 [24]. 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.

$$i^*(t) = \arg_i \max \nabla U(\bar{x}_i(t))\mu_i(t) \tag{2.3}$$

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:

$$\bar{x}_i(t) = \mathbb{E}\left\{\mu_i(t)\chi\{i = i^*\}(t)\right\} \tag{2.4}$$

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

The gradient rule can be applied to any concave utility function $U(\bar{X})$ and to systems where multiple users can be served at a time. Stolyar has proved the asymptotical optimality of this algorithm for multiuser throughput allocation [23]. Users $i = 1, \cdots, N$ are served by a switch in discrete time $t = 0, 1, 2, \cdots$ Switch state $g = (g(t), t = 0, 1, 2, \cdots)$ is a random ergodic process. In each state g, the switch can choose a scheduling decision k from a set $Q(g)$. Each decision q has the associated service rate vector $\mu^{(g)}(q) = (\mu_1^{(g)}(q), \cdots, \mu_N^{(g)}(q))$. This vector represents the service rate at a specified "time-slot" if decision q is chosen. The gradient algorithm is defined as follows; if at time t the switch is in state g, the algorithm chooses a possibly non-unique decision

$$q(t) \in \arg\max_{q \in Q(g)} \nabla U(\bar{X}(t))^T \mu^g(q) \tag{2.5}$$

where $\bar{X}(t)$ is a vector representing exponentially smoothed average service rates $\bar{x}_i$. Typically the utility function has the aggregate form $U(\bar{X}) = \sum_i U_i(\bar{x}_i)$. It has been shown that (2.5) converges to the optimal solution of $\max_{\bar{X}} U(\bar{X})$ as $t \to \infty$ [23].

Different degrees of fairness can be achieved with the gradient rule through the utilization of the α-PF fairness criterion [25], which dictates that the utility function should be defined as follows:

$$U_\alpha(\bar{x}) = \begin{cases} \log(\bar{x}) & \text{if } \alpha = 1 \\ (1-\alpha)^{-1}\bar{x}^{(1-\alpha)} & \text{otherwise} \end{cases}$$

Having $\alpha = 0$ will result in the Max-CIR rule, $\alpha = 1$ result in the PF rule and $\alpha = \infty$ gives the min-max rule.

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 [26]. 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" [15, 18, 27], first proposed for CDMA 1xEVDO, which is based on a logarithmic utility function. In [26], a larger class of utility functions is considered that allows efficiency and fairness to be traded-off. Generalized cμ-policies [28, 29, 30], such as a Max Weight policy [31, 32] 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. Andrews *et al.* considered a concave utility maximization problem with minimum and maximum rate ($\bar{x}_i^{min}$ and $\bar{x}_i^{max}$) constraints [33]. They propose a solution to this problem based on a scheduling algorithm by modifying the token counter. In the study, two specific forms of the scheduling algorithm are shown to guarantee $\bar{x}_i^{min}$ and $\bar{x}_i^{max}$.

2.2 Previous work

2.2.1 Opportunistic Schedulers

Long and Feng presented a rate-guaranteed opportunistic scheduling scheme [34], where they considered the transmission rate (throughput) as the fairness criteria, i.e. on the average the expected throughput of user i should be a fraction $\frac{\beta_i}{\sum_j \beta_j}$ of the whole system throughput, where β_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 acting 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 proposed an opportunistic scheduler that allows transmission of streaming traffic over HSDPA without losing much cell capacity [35]. The scheduler modifies the priority according to(2.6):

$$\mu_i^*(t) = \text{argmax}_{\mu_i(t)} \left\{ -\log(\epsilon) \frac{\mu_i(t)}{\bar{Z}_i} \left(1 - \frac{\frac{\mu_i(t)}{\mu_i^{req}(t)}}{\sum_{j=1}^{N} \frac{\mu_j(t)}{\mu_j^{req}(t)}} \right) \right\} \tag{2.6}$$

where $\mu_i^*(t)$ is the transmission rate for user i in the current time-slot t, μ_i is the achieved bit rate, μ_i^{req} is the required bit rate and N is the number of simultaneous streaming users in the cell. ϵ 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 taking 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.

Uplink scheduling is an important task although it is not examined as well as the downlink. In a study by Lim *et al.*, the authors examine performance in the presence of imperfect channel estimation in an uplink single carrier FDMA (SC-FDMA) system with uncoded adaptive modulation [36]. The special problem in the uplink is the lack of a broadcast channel that could be used to attain high quality channel reference in the receiver. Contrary to the downlink, in the uplink, the user equipment needs to send a specific short pilot pattern, called a 'sounding signal', related to each channel that is potentially available for scheduling. In the base-station receiver these sounding signals are used in order to estimate different channels for the scheduling decision. Since the required overhead in UL scheduling grows linearly with the number of scheduled users it is essential that sounding signals take as few radio resources as possible. Yet, such savings reflect directly on the reliability of the channel estimation results and thus, on the reliability of the scheduling decision.

Holma and Toskala layout in [37] the concept of implementing a frequency domain packet scheduler (FDPS) with fast adaptive transmission bandwidth. They present an example scheduler that starts allocating the user with the highest scheduling metric on the corresponding RB and allocates the adjacent RBs to the same user until a user with a higher metric is found. The FDPS proposed in this thesis resembles the aforementioned scheduler but is modified to allow the allocation of non-adjacent channels to the same user as long as those channels had the highest metric values

and the channels in between were free. In another study by Lim *et al.*, the authors suggest assigning resource blocks to users who obtain the highest marginal utility [38]. The FDPS in this thesis, follows a similar approach but differs in the way RBs are allocated to the users and includes the effect of imperfect channel state information and hybrid automatic repeat requests (HARQ) in the scheduling rule.

Another work considered in this review is that of Jersenius which provides a number of basic[1] allocation rules [39]. Her 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-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 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. Our work on the other hand deals with resource blocks independently

The performance of opportunistic schedulers for static user populations has been examined in a number of papers with either saturated conditions such as [40, 41, 42] or allow packet-scale dynamics but at heavy traffic limits [21, 43]. Other examples of related results can be found in [32, 44, 45, 46]. The assumption of a static population is reasonable since the scheduler works at the packet level on which the user population evolves only relatively slowly [47]. However, in the case of elastic traffic, this assumption is no longer satisfactory. Borst explored the flow-level performance in a dynamic population and later provided the necessary stability conditions [48]. Aalto and Lassila also studied dynamic traffic flows and provided the prerequisites for stability under certain conditions [49]. The instability of a system usually leads to poor performance as a consequence of growing queue sizes and packet delays. Both static and dynamic traffic conditions have therefore been considered in this thesis.

2.2.2 Inter-Cell Interference Coordination

The impact of interference in a cell differs from one user to another. Usually users situated at the edge of the cell suffer the most since the strength of a signal is inversely proportional to the distance. This in turn will result in a low SINR (Signal

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

to Interference+Noise Ratio) and consequently higher coding and lower data rates. ICIC (Inter-cell Interference Coordination) is one way to organize the shared resources among users in neighboring cells. The most common form of ICIC for the uplink is reuse partitioning [50]. The frequency band in this case is divided into several partitions and the cell is sectorized according to each partition allocation. Neighboring sectors of two cells will have different partitions. This reduces interference but at the cost of utilizing less resources. Some studies have for that reason proposed ways to reduce the effect of partitioning [51, 52]. In [51], edge users within a cell are grouped into distinct frequency groups. Users who have the same inter-cell interference quantized value are grouped into one group and assigned a chunk of the spectrum. In [52], the authors propose an adaptive reuse scheme where the bandwidth allocated to cell edge users is coordinated according to the traffic load. They later apply a reuse avoidance algorithm to verify any relocation of the assigned bandwidth. Reider suggested combining power control with ICIC where cells are divided into three sectors and low interference zones are defined [53]. The low interference zone in one sector must be different from the zone in the adjacent sector leading to the possibility of applying power control to the corresponding bands in those adjacent sectors.

The ICIC scheme proposed in this thesis introduces dynamic reuse by coordinating the transmission time of users in neighboring cells such that users who interfere with each other do not utilize the same resource block at the same time.

2.2.3 Game Theory in Scheduling

The application of game theory to radio resource management is subject to a number of considerations which include the existence of a steady state and the stability of that state. With a game theoretic analysis the network steady states can be identified from the Nash equilibriums of its associated game. Convergence of the solution is also an important factor in game theory. Neel *et al.* formulated a set of conditions necessary for modeling a wireless network as a game and these include two sets of conditions; the first is to ensure rationality and the second is a set of conditions for a non-trivial game [54]. The conditions described by Neel *et al.* are as follows:

Conditions for rationality

1- The decision-making process must be well-defined, i.e. each of the radios must follow a well-defined and deterministic set of rules for selecting an action with re-

spect to environmental factors.
2- A decision maker's choice to change an action should have a reasonable expectation to result in a positive improvement deviation.

Conditions for a nontrivial game
1- There must be more than one decision making entity in the network.
2- More than one decision maker has a nonsingleton action set.

The Nash Bargaining Solution is a well know strategy in game theory where a player enters the game with an initial point of disagreement and attempts to benefit more by negotiating with other players. The solution is mainly derived by maximizing the product of benefits and is subject to a number of axioms [55]. Han *et al.* introduced bargaining theory to OFDMA scheduling but on a local level only, i.e. intra-cell scheduling [56]. Our work extends this scheduling to the global level where the cells become the players.

2.2.4 Opportunistic 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 arrives 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 the 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 [57]. The new user is rejected if it was found that it will jeopardize the QoS of the current users, otherwise it is accepted. From a psychological point of view, it is easier for a user to be denied admittance than be admitted and later dropped during a call.

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 employed 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 [58] . Evans and Everitt used the concept of effective bandwidth to measure whether the signal to interference density ratio (SIDR) can be satisfied for each class with certain probability [59]. 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 based on their ability 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. The handoff call drop probability can be reduced 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, however, been little study on admission control in opportunistic multi-user communications.

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 distribution function $\mathbb{F}_{\xi_i}(\xi_i)$, which is independent of t. With this information at the admission controller's disposal, it can estimate the impact of the new user and form the decision on whether to accept the new user or reject it.

The importance of combining opportunistic scheduling with admission control has been recognized in several studies, see e.g. [60, 61, 62]. In [63] 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 [64], 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.

In [65] the authors propose a smooth admission control scheme. The basic idea of the controller is to gradually increase 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.

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; maximize the number of users admitted or minimize the blocking probability.

The admission control scheme presented in this thesis resembles the sliding window based call admission control scheme suggested by Zhao and Zhang [66] and can be interpreted as a modification of the active link protection scheme suggested by Bambos *et al.* to the multiuser diversity channel [67].

2.2.5 Opportunistic Rate Control

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 [63] 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 the 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 a stochastic optimal control problem. In [68], a method for controlling the resource allocation for the different users was suggested. The scheme added one control parameter to the scheduling metric that was changed based on the observed channel access time in some time window.

Patil and Veciana proposed 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 timeslot [69]. Users are divided into two categories: Real-time users and Best Effort users. Each real-time user is assigned 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. Zhang *et al.* utilized stochastic approximation algorithm to guarantee certain quality of service level in terms of minimum data rate [70]. We note that the stochastic approximation algorithm can result in either very slow convergence or very high variance of the control parameters.

The proposed scheduler that controls rate in this thesis resembles the scheduler suggested by Liu *et al.* [63], but instead of modifying the scheduling metric, the active set of users is controlled, i.e. the number of active users at a given moment

of time.

2.2.6 Multi-carrier Systems

In a multi-carrier system, channel variations of 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 [71]. In multiuser environments, the situation becomes more complicated as each user will 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. Therefore, 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 studies that discuss waterfilling in a multiuser environment such as [72] 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 proposed an iterative waterfilling algorithm based on dual composition [73]. Two decompositions are considered, one in the subcarrier domain and another in both subcarrier and user domains.

In OFDMA networks, the bandwidth is divided into many narrowband subchannels [74]. The task of the resource scheduler is to divide the transmitter power among the different channels and the channels among the different users. OFDMA schedulers can be divided into two categories; schedulers with fixed power allocation and schedulers with variable power allocation as shown in Fig. 2.2. Variable power schedulers come from the fact that 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

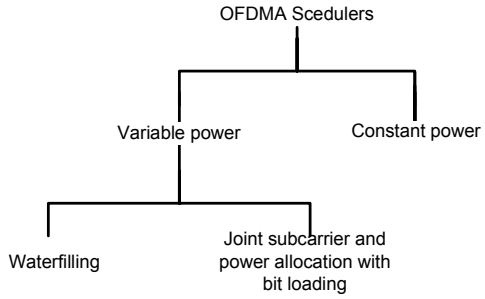

Figure 2.2: OFDMA scheduler classes

transmitter with detailed CSI using a robust feedback channel. 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). Studies regarding the application of adaptive bit loading algorithms to wireless channels can be seen in [75, 76, 77, 78].

Adaptive OFDMA has been considered for the 3G LTE [79]. Several studies have looked at scheduling in adaptive OFDMA systems and proposed optimal schemes such as [80, 81, 82]. 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 [83]. However, reporting accurate CSI requires a considerable amount of overhead, this in turn introduces a trade-off between the quality and the overall throughput of a system. For this reason, a technique called 'clustering' is introduced where subchannels are grouped into clusters of wide-band channels. This limits the amount of necessary feedback by selecting the same feedback for all narrowband channels within a cluster [84]. Cherriman *et al.* suggested to group the subcarriers into RBs and report one CSI for each RB [85].

Zhang *et al.* proposed reducing the amount of feedback bits for a cluster benefiting from the high correlation of a cluster's subchannels [86]. Gesbert and Alouini proposed in [87, 88] to utilize a one bit per user feedback approach. In their work users notify the base-station only if they exceed a certain SNR threshold. Their work was further extended to include a multi-stage version [89]. An analysis of the different number of feedback bits per user techniques has been made in [90]. The accuracy of channel estimation based on the feedback reports also plays a vital role in opportunistic scheduling. Channel estimation errors or outdated CSI reports can significantly decrease system performance as users are allocated resources that do not match their actual conditions. The effect of channel estimation errors in

OFDMA systems has been studied in [91]. Agrawal *et al.* developed an optimal and a sub-optimal scheduler for OFDMA systems by modeling the channel estimation error as a self-noise term in the decoding process [92]. In this thesis, the author extends the work to study the effect of the number of information bits as well as what the type of that information.

Part I
Uplink Scheduling

Chapter 3

Utility Based Scheduling

In this chapter we will see scheduling algorithms that maximize specific utility functions while take into account the limitations of the access system. First, a single-cell scheduler is created and is later extended to include multiple cells and act as a centralized coordinator. In addition, the optimal solutions will be presented with the help of integer programming. While the optimal solutions cannot be realized in practice due to the computational complexity, it provides good insights of how well other algorithms perform.

3.1 Single-cell Multi-carrier Scheduling

This part develops a scheduler for SC-FDMA systems. The scheduler should be consistent with the resource allocation constraints of the uplink channel for 3G LTE systems. Additionally, the scheduler must also take into account failed transmissions when forming a scheduling decision. A heuristic scheduler is one proposition and is considered a suitable choice since it is computationally feasible and is able to find a practical solution to the resource allocation problem. A study is also made on the impact of traffic reports on the overall system performance.

3.1.1 Localized Gradient Algorithm LGA

The gradient algorithm is considered as the metric for the scheduler in this chapter. Referring to Stolyar's framework presented in Chapter 2,

$$q(t) = \arg \max_{q \in Q(g)} \nabla U(\bar{X}(t))^T \mu^g(q) \tag{3.1}$$

where $Q(g)$ in this case will denote all feasible RB assignments that can be made with state g at time instant t. The set is confined by the channel capacity as well as the constraints on the allocation of the RBs. In what follows, an integer programming assignment problem is formulated for solving $q(t)$ 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 UE divides its available power evenly among the assigned RBs. Based on 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.

$$y(t) = \arg\max_{y} \sum_{i=1}^{N} \sum_{n=1}^{L} U_i(\bar{x}_i)\mu_{i,n}y_{in} \tag{3.2}$$

subject to

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

where N is the total number of users and L is the total number of RBs. It can be seen that the first inequality limits the RB to one user only. The second inequality enforces the requirement of 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 $y_{i,m} \leq -\big(1 - (1 - y_{i,(n+1)})\big)$, i.e. the inequality becomes redundant.

The gradient scheduler discussed above is optimal for perfect channel state information. In measurement delay cases and estimation errors, the selection rule occasionally picks rates that do not match the channel state. It is assumed that the synchronous non-adaptive HARQ mentioned in Section 1.2.2 is utilized to deal with the errors. Now the scheduler has to reserve those RBs to the UE that has scheduled retransmissions. To take this into account in the integer programming problem, we

need to add a constraint

$$y_{i,n} = 1, \quad \text{if user } i \text{ has an ARQ process on RB } n \tag{3.3}$$

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 that can be used as a reference. To validate this claim, it can be noted that the performance loss due to retransmissions is low as shown in Section 3.1.3.

3.1.2 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 non-feasible. Hence, there is a need for simpler algorithms that can provide adequate solutions promptly. 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 $\mathcal{Z}_i$ denote the set of RBs assigned to user i and $\mathcal{L}_i$ denote 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 $\mathcal{Z}_i$ and $\mathcal{L}_i$ for all i and t.

Algorithm 3.1

$$\begin{aligned} \mathcal{Z}_i^{(0)} &= \emptyset \\ \mathcal{L}_i^{(0)} &= \{RB_1, RB_2, \cdots, RB_L\} \end{aligned}$$

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

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

Step 2: Assign RB z^* to user i^* and update $\mathcal{L}_i$.

$$\begin{aligned} \mathcal{Z}_{i^*}^{(k+1)} &= \mathcal{Z}_{i^*}^{(k)} \cup z^* \\ \mathcal{L}_i^{(k+1)} &= \mathcal{L}_i^{(k)} \setminus \mathcal{N}_i^{(k)}, \quad i \neq i^* \end{aligned}$$

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

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

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

Update $\mathcal{L}_i$ 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.

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

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. $\mathcal{R}$ is the set of users that have ARQ processes. τ is a fixed predefined time. Iterating for $\mathcal{L}_i$ with $\mathcal{L}_i(t+\tau)^{(0)} = \{RB_1, RB_2, \cdots, L\}$, we have

$$\mathcal{L}_i^{(k+1)}(t+\tau) = \mathcal{L}_i^{(k)}(t+\tau) \setminus \mathcal{Z}_h^r(t+\tau), \quad i \neq h$$

Numerical Example

For a better understanding of the heuristics, a 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 has 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

3.1.3 Computational Evaluation

System parameters for the evaluation are described in Table 3.1. 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 fashion. Fig. 3.1 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). It can be seen that there is only a 4% gap between the LGA optimal solution and the NLGA optimal solution. This gap represents the impact of the localization requirement which implies that the localization constraint has a low impairment on the performance of the LGA. It can also be seen in the figure that the HLGA provides a close to optimal performance when compared to the NLGA and LGA optimal solutions. For the imperfect channel information case it can be seen 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 3.2 shows the exact throughput values for both cases. In Table 3.3 the methods are compared with another opportunistic scheduler obtained from the lit-

Table 3.1: System parameters

Parameter	Value
RB bandwidth	375 kHz
Total number of blocks	10 (25 subcarrier/RB)
TTI duration	1 ms
Packet arrival distribution	Log-normal
Mean inter-arrival time	60 ms
Standard deviation	5 ms
Fading model	Two path Raleigh
No. of terminals	5 (Single cell)
Site to site distance	100 m
Number of Tx antennas	1
Max. Tx power	21 dBm
Noise power	-108.5 dBm

Table 3.2: Cell throughput values in Mbps for perfect and imperfect channel estimation -Dynamic traffic

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

erature [37] which we shall call the fast adaptive transmission bandwidth (FATB) scheduler. The scheduler in this case starts allocating the user with the highest scheduling metric on the corresponding RB and allocates the adjacent RBs to the same user until a user with a higher metric is found. The assumptions for this evaluation were perfect channel estimation and full buffer traffic. Results show that their is improvement in the HLGA scheme over the FATB in small and large cells. This gain is further increased with the LGA algorithm.

Dynamic Traffic Conditions

This section demonstrates the performance of the algorithm in a dynamic traffic model with different scenarios.

Pruning: In a dynamic traffic model, the amount of resources given to a user should correspond to the amount of data in the transmission buffer. A proposition to solve that problem is to use a pruning procedure. Resource blocks are first allocated to active users regardless of the amount of data they have using the HLGA. Once the RBs have been allocated, pruning is performed to find the extra blocks allocated to a

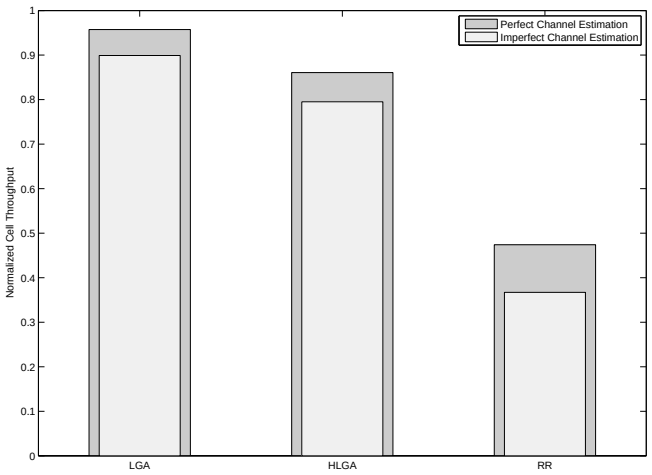

Figure 3.1: Normalized cell throughput with perfect and imperfect channel estimation

Table 3.3: Cell throughput comparison with the FATB scheduler -Static traffic (Full buffer) for different cell radii

	RR	FATB	HLGA	LGA	NLGA
100 m	24.97	26.75	26.96	27.54	28.15
1000 m	8.3861	10.6292	12.8480	13.4045	13.8310

user and re-allocate them to neighboring users in the spectrum or to users who have not been assigned any block in the scheduled TTI. Pruning is usually performed to edge blocks due to the localization constraint. The procedure can be summarized in the following algorithm

Algorithm 3.2

Step 1: Apply the HLGA algorithm to obtain the resource block allocation that maximizes (3.1).

Initiate pruning with $i = 1$, denoting the user with assignments in the beginning of the frequency spectrum.

Step 2:

if for user i

$$\sum_{n=1}^{L} \mu_{i,n} y_{i,n} > W_i$$

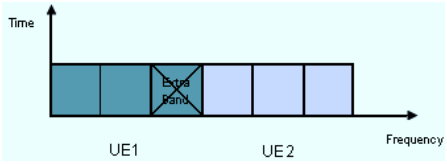

Figure 3.2: Pruning Example

where $\mu_{i,n}$ is the rate of user i obtained with RB n and $y_{i,n}$ is a selection variable: $y_{i,n} = 1$ if RB n is assigned to user i and 0 otherwise. W_i is the total amount of data in the buffer of user i.
then prune the edge block and assign it to the next user in spectrum or to a unassigned user who maximizes the allocation problem with the extra block.
Step 3: Repeat step 2 until:

$$\sum_{n=1}^{L} \mu_{i,n} y_{i,n} \leq Wi$$

Step 4: Repeat steps 2-3 to the following user in spectrum: $i = i + 1$.

It follows that pruning will spare any extra blocks assigned by the HLGA, making it possible for users in need to benefit from them. However, due to the contiguity constraint of SC-FDMA, the beneficiary users have to be either users with spectrum allocations adjacent to the extra blocks or users with no allocations in the scheduled TTI.

Example:
Figure 3.2 shows two UEs with allocated bands. Each UE has been allocated three bands. According to the transmission buffer of UE1, it was found that the amount of data can be fitted in two bands only. Therefore, UE1 can spare one band that can be pruned and allocated either to UE2 or to a non-allocated UE in the scheduled TTI. The impact of pruning on the average packet delay can be seen in Fig. 3.3. The users are sorted according to their link gains starting with the user with the best channel and ending with the worst user. It can be clearly seen that pruning has a significant impact on performance. Average packet delay is reduced dramatically with pruning. The optimal solution is included for the sake of performance evaluation.

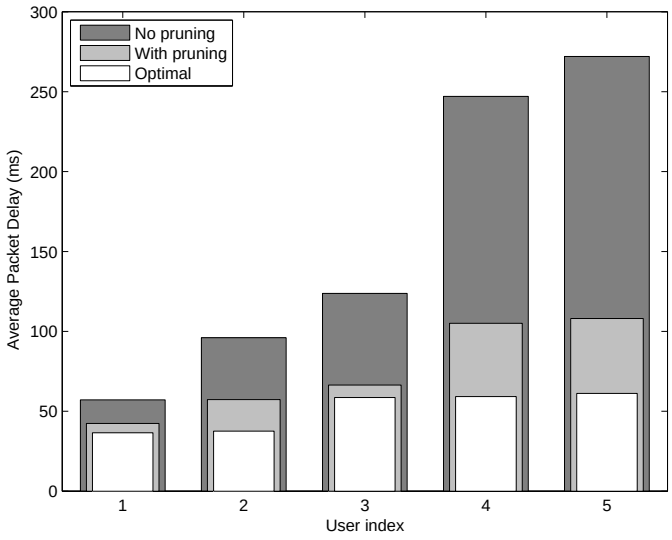

Figure 3.3: Average packet delays

Buffer Occupancy Report Delay:

Buffer occupancy (BO) status reports are generally used in data communication systems to support uplink packet scheduling decisions. Buffer reports are needed to achieve high radio resource utilization and consequently higher performance.

Detailed BO report: It is assumed so far that the buffer information of the users is always available at the scheduler. In reality when a UE has data it wants to transmit, it makes a request for resources and reports its BO information to the BS. The UE will be scheduled the necessary time and frequency resources based on that report. However, due to the delay of the next buffer report, the BS will keep on scheduling the subject UE based on old buffer information, when in fact the buffer status has changed due to new packet arrivals. Large intervals between reports lead to packet delays increasing dramatically due to the accumulation of packets arriving before the next report and consequently staying in the buffer because of the absence of the necessary resources. On the other hand, when the report interval is small, the number of arrivals is much smaller. This will lead to the fact that at the time of the second buffer report an empty user will be considered non-active for the next report interval giving all the resources to other users for the complete interval. Figures 3.4, 3.6 and 3.8 represent the cumulative distribution functions (CDF) for the average packet delay of the best, worst and median users respectively. It can be seen that

in the case of the best user, performance is at its best when there are no delays in the reports. As the report delay increases, system performance decreases. In the worst user case it is interesting to see that the best performance is with small buffer delays since good users who are able to empty their buffers will become idle. This in turn will make the set of active users who need to be scheduled smaller. Performance however decreases with higher report delays since the probability that good users will have a packet arrival becomes higher and in turn makes them active. It is also interesting to see that the user with the weakest link experienced the worst performance when there were no report delays. In the median user case, there is almost similar performance with small and no report delays.

Limited BO report: Detailed BO information will help the BS to assign resources more efficiently at the cost of increased uplink signaling overhead. Therefore, there will be a trade-off between the BO accuracy and the scheduling gain. This part will look into the case where signalling is at its minimum with the report consisting of only 1 bit of information declaring either a full or an empty buffer, i.e. active or non-active without reporting how much data there is in the buffer. In this case the base-station will not limit the resources granted to an active user that maximizes the allocation optimization problem. In this case bad channel users will suffer more delay since they will only be selected when their rate quality factor is higher than others. The rate quality here is the instantaneous achievable rate divided by the average throughput on a band, i.e. the allocation of bands will be a function of channel condition only rather than channel condition + buffer size. Again, the performances of the best, worst and median users are presented in Figures 3.5, 3.7 and 3.9 respectively. Naturally, packet delays grow larger due to the limited amount of information available at the scheduler. This is clearly noted when comparing the cases of the worst and median users of Figures 3.6 and 3.8 with 3.7 and 3.9 respectively. This result also shows the impact of pruning due to the fact that in limited BO reports, pruning cannot be utilized. The overall outcome is similar to the detailed BO report results except with larger delays.

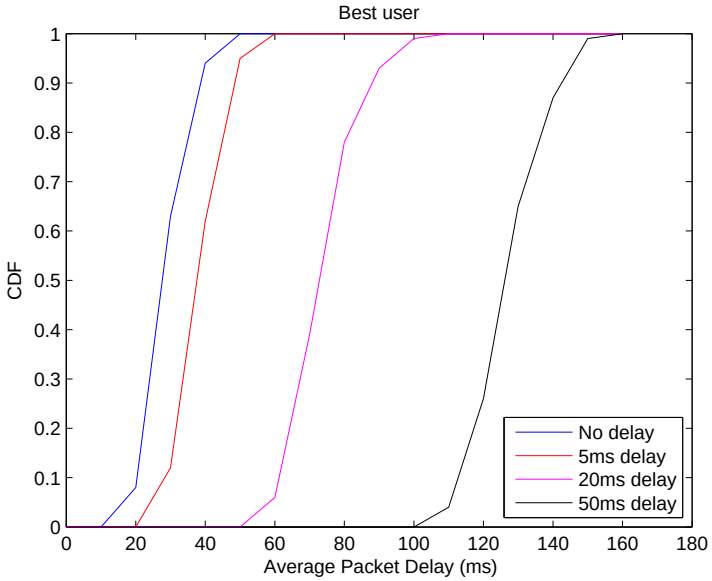

Figure 3.4: Average packet delay CDF for best user - Detailed BO report delay

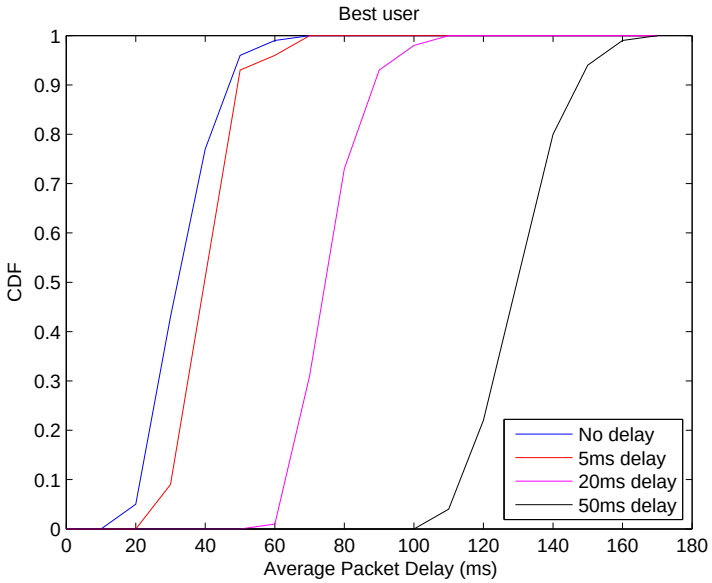

Figure 3.5: Average packet delay CDF for best user - Limited BO report delay

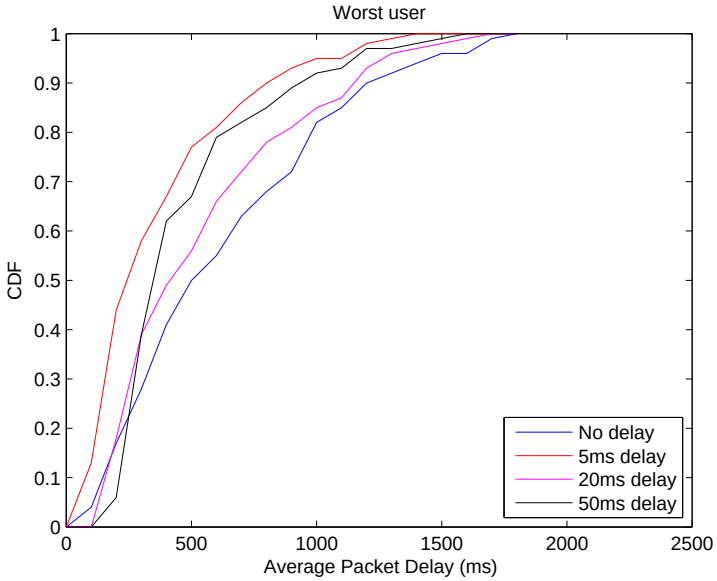

Figure 3.6: Average packet delay CDF for worst - Detailed BO report delay

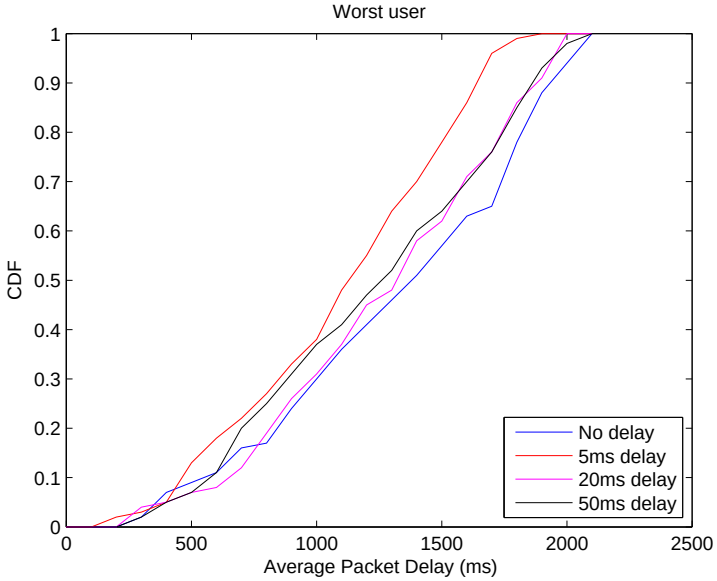

Figure 3.7: Average packet delay CDF for worst user - Limited BO report delay

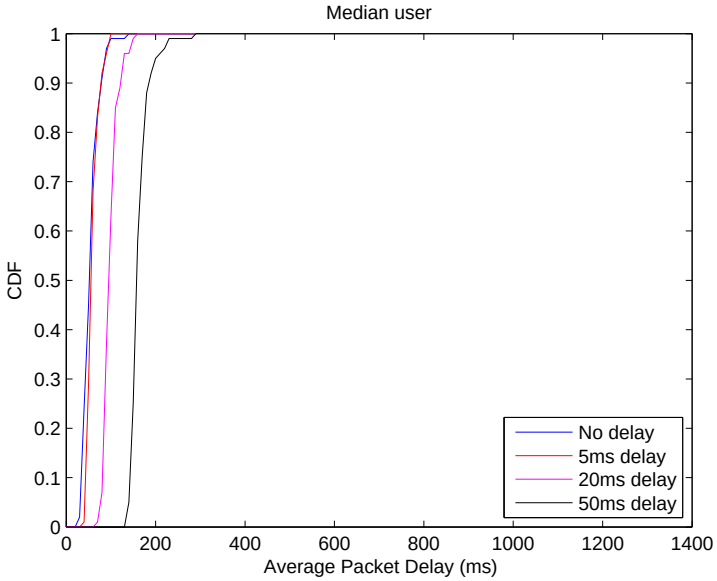

Figure 3.8: Average packet delay CDF for median user - Detailed BO report delay

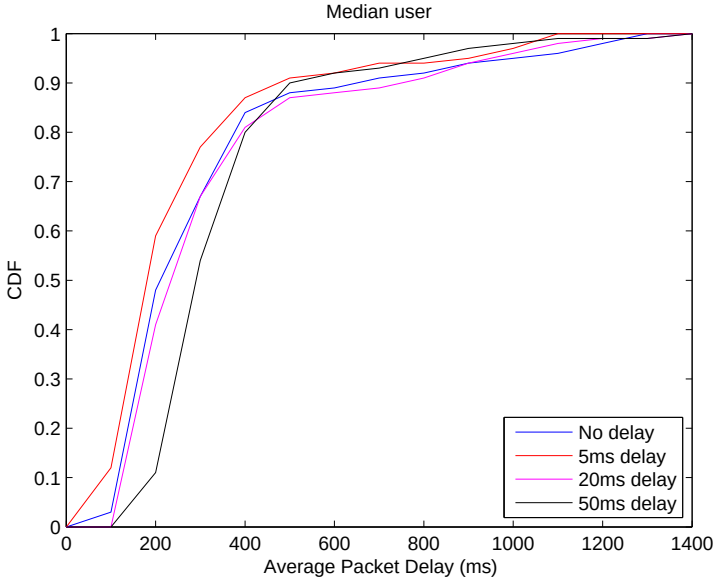

Figure 3.9: Average packet delay CDF for median user - Limited BO report delay

3.2 Multi-cell Multi-carrier Scheduling

The single-cell algorithm will now be generalized to include interference from other cells. The objective of the scheme in this case is to coordinate the uplink transmission in neighboring cells such that inter-cell interference is mitigated and the aggregate utility of the users is maximized.

3.2.1 Heuristic Scheduling

In this section, a heuristic approach is proposed to coordinate the scheduling of users in neighboring cells. The basic idea of the scheme is to switch users on and off to maximize a certain utility function taking into account the constraints of the uplink interface.

We first start by assigning the RBs in each of the neighboring cells separately using the single-cell heuristic scheduler proposed in Section 3.1.2. It is also possible to use other methods for the assignment at this stage. We then proceed with a 2-cell approach where coordination is carried between 2 cells at a time. The result coordination of each cell is later coordinated with another neighboring cell and so on, i.e. iterative improvement is implemented for the coordination. This pairing approach has also been implemented in for instance [56]. For the users of two cells sharing RB n, the marginal utility when both users transmit is compared with the case where only one user transmits.

Algorithm 3.3

Step 1: Start with cell pair $\mathcal{B} = \{1, 2\}$.

If

$$\nabla U_i(\bar{x}_i)\mu_{i,j,n}(t) + \nabla U_j(\bar{x}_j)\mu_{j,i,n}(t) \geq \nabla U_i(\bar{x}_i)\mu_{i,0,n}(t),$$
$$i \in \mathcal{N}_1(n), \quad j \in \mathcal{N}_2(n), \quad n = 1, 2, \cdots, L$$

and

$$\nabla U_i(\bar{x}_i)\mu_{i,j,n}(t) + \nabla U_j(\bar{x}_j)\mu_{j,i,n}(t) \geq \nabla U_j(\bar{x}_j)\mu_{j,0,n}(t),$$

then;

Both users can utilize RB n

otherwise;

$$q^* = \arg \max_{q \in \mathcal{N}_b(n)} \nabla U_k(\bar{x}_q)\mu_{q,n}(t), \quad b \in \mathcal{B}$$

where $\mathcal{N}_b(n)$ is the set of users in cell b who have been assigned RB n. The notation of $\mu_{i,j,n}$ here denotes the data rate of user i of cell 1 with RB n when user j of cell 2 is utilizing the same RB. If j doesn't transmit on RB n then the notation is written as $\mu_{i,0,n}$

Step 2: We check for gaps in the spectrum that happened due to disabling users from disputed RBs. Since SC-FDMA demands an intact spectrum, gaps cannot be present in the spectrum allocated to a user except at the edges to preserve the locality constraint. Therefore, if a hole was found, for example, in the middle, a suggestion would be to disable all the RBs that come before or after the gap which belong to the allocated user. The decision whether to disable the RBs that come before or after the gap is made based on the following criteria:

For a gap in a resource block with the index τ;

If

$$\sum_{n<\tau, n\in\mathcal{L}_i} \nabla U_i(\bar{x}_i)\mu_{i,n}(t) > \sum_{n>\tau, n\in\mathcal{L}_i} \nabla U_i(\bar{x}_i)\mu_{i,n}(t)$$

then;

$$\mathcal{L}_y := \mathcal{L}_i \setminus n, \quad \forall n > \lambda$$

otherwise;

$$\mathcal{L}_i := \mathcal{L}_i \setminus n, \quad \forall n < \lambda$$

Step 3: With the new assignment, Repeat steps 1-2 for other cell pairs until all the pair combinations are exhausted.

3.2.2 Optimal Scheduling

This section formulates the required optimization problem for coordinating the uplink transmission of multiple users in multiple cells. For that purpose, the algorithm finds the allocations that would maximize the aggregate of the marginal utilities.

$$
\begin{aligned}
& y(t) = \\
& \arg\max_{y} \sum_{n=1}^{L} \sum_{b\in\mathcal{B}} \sum_{i\in\mathcal{N}_b} \sum_{\overrightarrow{j}\in\mathcal{J}_{-b}} \nabla U_i(\bar{x}_i)\mu_{i,n}(\overrightarrow{j},b,t)y_{i,n}(\overrightarrow{j},b), \\
& \mathcal{B} = \{\text{Set of base-station indices}\} \\
& \mathcal{J}_{-b} = \{[j_{\Phi(1)}, j_{\Phi(2)}, \cdots j_{\Phi(\mathcal{B}-1)}] \mid j_{\Phi(k)} \in \mathcal{N}_{\Phi(k)}\}], \quad \Phi = \mathcal{B} \setminus b
\end{aligned}
\tag{3.4}
$$

subject to

$$
y_{i,n}(\overrightarrow{j}) \in \{0,1\}, \quad \forall(\overrightarrow{j}) \in \mathcal{J}_{-b}
$$

$$
\sum_{i\in\mathcal{N}_b} y_{i,n}(\overrightarrow{j}) \leq 1, \quad \forall(\overrightarrow{j}) \in \mathcal{J}_{-b} \tag{3.5}
$$

$$
\sum_{i\in\mathcal{N}_b} y_{i,n+1}(\overrightarrow{j}) - \sum_{i\in\mathcal{N}_b} y_{i,n}(\overrightarrow{j}) + \sum_{i\in\mathcal{N}_b} y_{i,m}(\overrightarrow{j}) \leq 1, \tag{3.6}
$$

$$
n = 1, 2, \cdots, L \quad m = n+2, n+3, \cdots, L, \quad \forall(\overrightarrow{j}) \in \mathcal{J}_{-b}
$$

$$
\sum_{n=1}^{L} \sum_{i\in\mathcal{N}_b} \min\Big(W_i(t)\ ,\ \mu_{i,n}(\overrightarrow{j},t)y_{i,n}(\overrightarrow{j})T_{slot}\Big) \leq W_i(t), \tag{3.7}
$$

$$
\forall(\overrightarrow{j}) \in \mathcal{J}_{-b}
$$

where $y(t)$ is the RB allocation at time t. The total number of RBs is L and $\mathcal{N}_b$ is the set of users in cell b. The variable $y_{i,n}(\overrightarrow{j},t)$ is the selection probability that RB n is allocated to user i in cell b and users in the vector $\overrightarrow{j}$ for the other cells at time t. W_i denotes the transmission buffer occupancy for user i. The duration of the time slot is represented by T_{slot}.

Equation (3.5) will limit a RB to one user only in each cell. Equation (3.6) enforces the requirement of consecutive blocks. If $y_{i,n}(\overrightarrow{j},t) = 1$ and $y_{i,n}(\overrightarrow{j},t) = 0$, then $y_{i,m}(\overrightarrow{j},t) \leq 0$ for $m > n+1$. If on the other hand both $y_{i,n}(\overrightarrow{j},t) = 1$ and $y_{i,n+1}(\overrightarrow{j},t) = 1$, the inequality requires that $y_{i,m}(\overrightarrow{j},t) \leq 1$. If $y_{i,n}(\overrightarrow{j},t) = 0$ then the inequality states that $y_{i,m}(\overrightarrow{j},t) \leq -\Big(1-(1-y_{i,n+1}(\overrightarrow{j},t))\Big)$. This means that the inequality becomes redundant. Equation (3.7) is to insure the matching between the amount of granted resources to the actual need. The constrained optimization problem above is solved by integer programming.

Table 3.4: System parameters

Parameter	Value
RB bandwidth	375 kHz
Total number of blocks	6 (25 subcarrier/RB)
TTI duration	1 ms
Fading model	TU-6 Raleigh fading
Radio propagation	Site to site distance 100 m
Max. Tx power	21 dBm
BS antenna gain	18 dBi
UE antenna gain	0 dBi
Noise power	-108.5 dBm

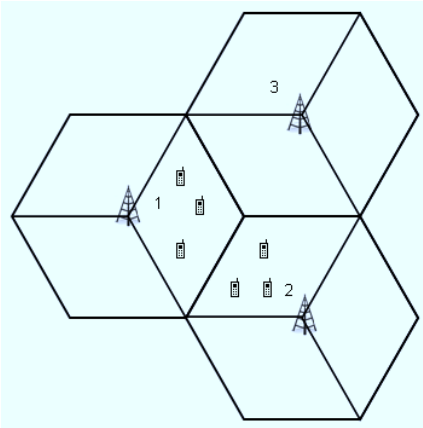

Figure 3.10: Cellular model

3.2.3 Computational Evaluation

The performance of the scheme has been evaluated in terms of uplink average throughput, utility and delay, via Monte Carlo simulations. The simulations consider only a two-cell case where the focus is on two adjacent sectors 1 and 2 of two neighboring cells as shown in Fig. 3.10. Two scenarios for the traffic are considered; one with packets arriving according to a log-normal distribution with fixed packet sizes arriving with different inter-arrival times and the other being a full buffer case where there are packet arrivals in every TTI causing a congestion in the buffers. System parameters are shown in Table 3.4. $\log(\bar{x})$ is used as the utility function $U(\bar{x})$ for the considered model with $\bar{x}$ being the average throughput [93]. This in turn will produce the proportional fair rule when used in the gradient algorithm. Perfect channel estimation is considered so retransmissions are not needed.

In the simulations the proposed scheme is compared with cases when there is no

coordination using both the PF and round robin schedulers with a reuse 1 factor. Reuse 1 means that the whole spectrum is available to all the users in all cells. Other references that are used in the comparison are proportional fair and round-robin schedulers in a reuse static reuse partitioning model. Here, a reuse factor of 3 is used which allows only one third of the spectrum to users who are situated near the edge of the cell. A user is defined as a cell-edge user if the difference in the link gain to the home and neighboring base-stations is less than 5 dB. Simulations reveal that using the proposed coordination scheme yields near-optimal results. This can be seen in Fig. 3.11 which plots the total average throughput of the system. The simulations had a standard deviation of less than 1 Mbps. It can clearly be seen that with coordination the output of the system is matched to the input up to very high traffic loads. On the other hand non-optimal methods tend to produce varying amounts of outage as loads grow higher. Throughputs are also plotted for increasing numbers of users shown in Fig. 3.12 where it can be seen that using the coordination scheme will always provide higher data rates as long as the number of users in each cell is larger than 1.

Figures 3.13 and 3.14 represent the aggregate utility value for both cases. It shows that using the coordination scheme maximizes the utilities of the users providing a reasonable degree of satisfaction for all users in the system. Finally, Figures 3.15 and 3.16 show the mean amount of delay the users suffer with the different methods. The optimal solution was not considered in Figures 3.12, 3.14 and 3.16 due to the computational complexity in solving the optimization problem as the number of users increases. Therefore, based on the results obtained from the 2 user/sector case in Figures 3.11, 3.13 and 3.15 the assumption was made that the optimal solution can be generalized. It is also worth noting that the reason PF and RR methods provide similar performances is due to the fact that we are operating in a high SINR region where the logarithmic rate-SINR mapping is almost constant.

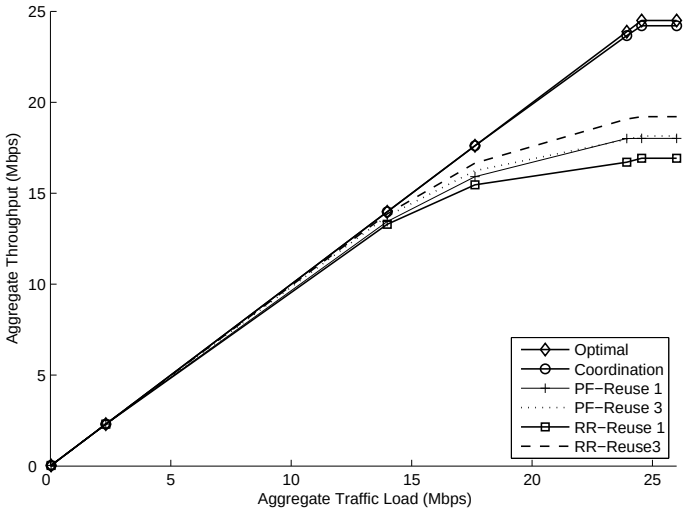

Figure 3.11: Aggregate throughput as a function of the traffic load for a fixed number of users; 2 users/cell

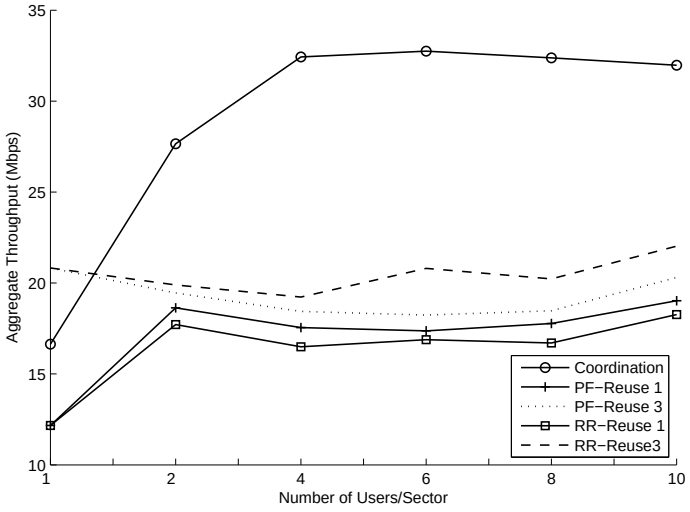

Figure 3.12: Aggregate average throughput as a function of the number of users for a fixed traffic load; Full buffer

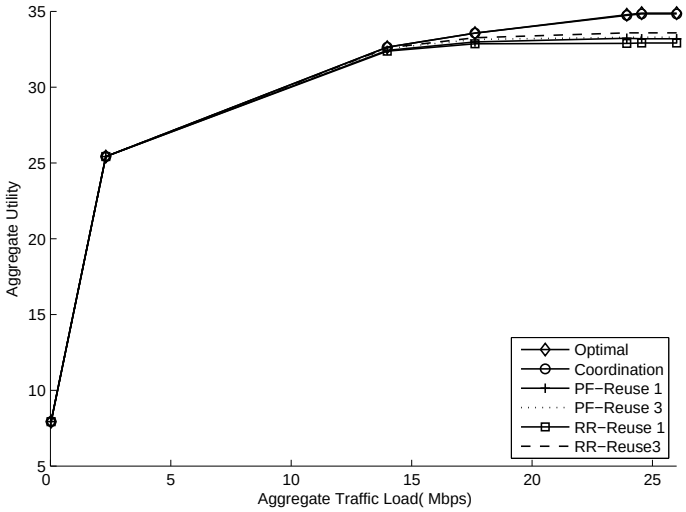

Figure 3.13: Aggregate utility as a function of the traffic load for a fixed number of users; 2 users/sector

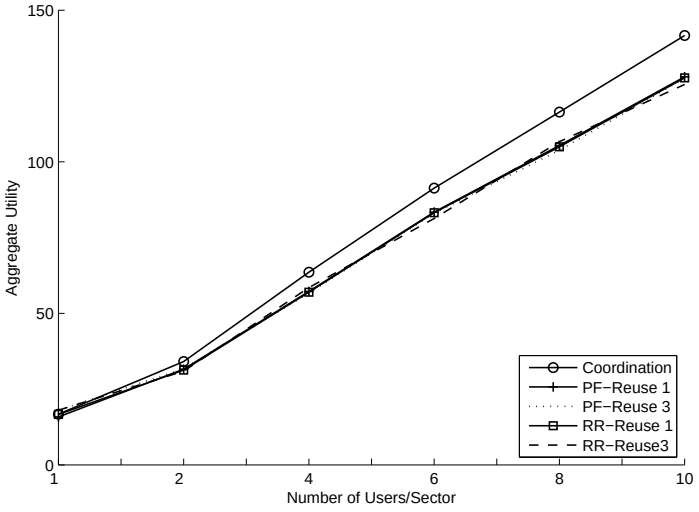

Figure 3.14: Aggregate utility as a function of the number of users for a fixed traffic load number of users; Full buffer

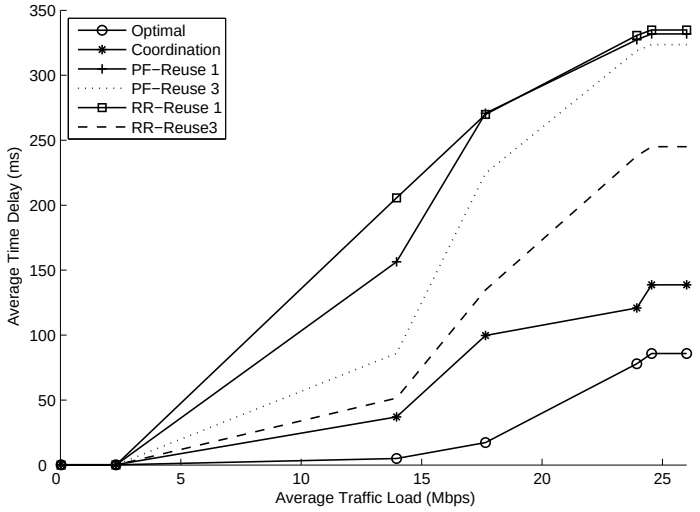

Figure 3.15: Average packet delay as a function of the traffic load for a fixed number of users; 2 users/sector

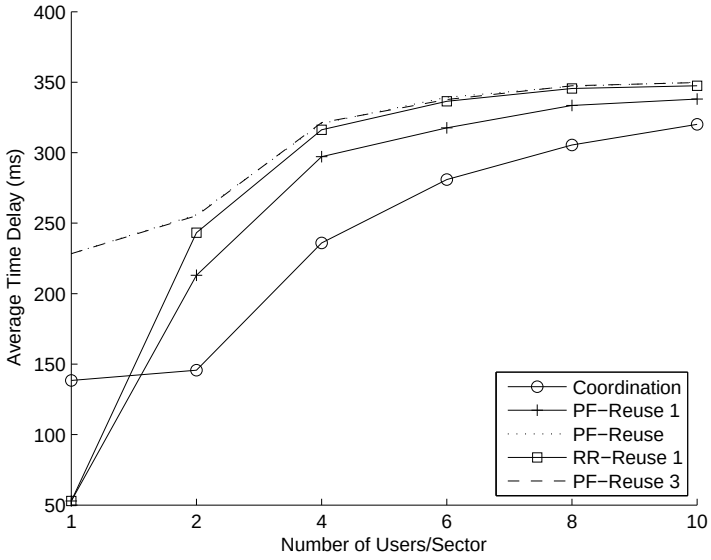

Figure 3.16: Average packet delay as a function of the number of users for a fixed traffic load number of users; Full buffer

3.3 Concluding Remarks

The resource allocation constraint in the uplink channel of 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 and therefore, a heuristic approach to the scheduling problem would be a suitable choice to find a practical solution to the allocation optimization problem. For this purpose, the HLGA was suggested and provided a good performance when compared to the optimal solution of the LGA. The LGA, on the other hand, was considered realizable only in theory due to its complexity making it a good benchmark to measure the performance of other channel dependent scheduling rules for localized SC-FDMA.

The work in this chapter also proposed a procedure that can match resources to the buffer occupancy of users in the case of dynamic traffic models. With pruning, resources are allocated more efficiently, taking into consideration the amount of data in users' buffers as well as the maximization of the resource allocation problem. The results show a significant impact on performance. It is worth noting that pruning can also be used to discard weak bands from a set of bands scheduled for a user and re-distribute their power on the remaining bands. The effect of buffer occupancy report delays were also examined and was found that small report delays were in fact beneficial from the fairness perspective as users having good link quality occasionally backed off from the scheduling after emptying their buffers. This in turn left resources free for the others to utilize whereas, in the case of good users, they are non-active more frequently. The result of this is that when a user is declared non-active at the time of the report, it will not be allocated any resources until the next report, giving more chances to the bad users. Limited buffer information reports were also considered. It was found that limited information about the buffer status would not lead to utilizing resources efficiently causing a waste of these resources.

The single-cell scheme was later extended to a multi-cell system. Results show that cell coordination can provide a significant amount of improvement to system performance. The necessary multi-user multi-cell optimization problem was also formulated to find the optimal scheduling of resources to maximize the utility of the users and reduce inter-cell interference. The solution obtained from this optimization provides a good reference to other opportunistic and non-opportunistic coordination

methods. The main restriction of the schemes is the assumption of perfect channel estimation, fixed transmission power, fixed noise power and non-real time data.

Chapter 4

Bargaining Based scheduling

This chapter discusses bargaining as a means of scheduling by coordinating the sharing of resources. Cells negotiate with each other to determine the number of channels that should be allocated to their edge users. Thus, the distribution of the channels will be a result of a bargaining process that works on maximizing a global optimization problem. The bargaining can be viewed as a dynamic reuse scheme where the number of edge bands are changed according to the optimization process.

4.1 System Model

The Resource blocks in this chapter are divided into fractional reused RBs and are denoted as low interference RBs (LI-RBs) and the reuse 1 RBs as high interference RBs (HI-RBs). The channel is assumed to be subject to slow fading only.

The focus in this work is on average throughputs that can be obtained in a given scheduling interval. The interval is assumed to be short enough so that the underlying channel and traffic processes can be assumed wide sense stationary, but at the same time it is assumed that the scheduling interval is large enough so that the frequency selective fast fading is essentially averaged out from the resulting mean data rates and arbitrary fractions of the RBs can be allocated to the users via time domain scheduling. The mean data rate of user j per RB is $\bar{x}_j$ and α-fair scheduling is considered where the utility per user is defined as:

$$U_\alpha(\bar{x}) = \begin{cases} \log(\bar{x}) & \text{if } \alpha = 1 \\ (1-\alpha)^{-1}\bar{x}^{(1-\alpha)} & \text{otherwise} \end{cases}$$

The α-fair scheduling defines a continuum of scheduling laws ranging from max

throughput ($\alpha = 0$) to Proportional Fair ($\alpha = 1$) and min-max fair ($\alpha \to \infty$). The parameter α could be also interpreted as a resource split between users having favorable channel conditions and users having poor channel conditions. The larger α is, the more time-frequency resources are given to users with low data rates. The case $\alpha = 0$ leads to a greedy solution where few users close to the base-station grab all the resources leaving the users close to the edge of the cell without any service.

4.2 Intra-cell Scheduling

Consider the following scheduling problem

$$F_b(l) = \max_{\phi,\varphi} \sum_i U_\alpha[\bar{x}_i(l)] \tag{4.1}$$

s.t.

$$\bar{x}_i = \begin{cases} \mu_i l \phi_i & \text{if } i \in \mathcal{N}_{e,b} \\ \mu_i \left(l\phi_i + (L-l)\varphi_i\right) & \text{if } i \in \mathcal{N}_{c,b} \end{cases}$$

$$\begin{aligned} \sum_{i \in \mathcal{N}_{c,b} \cup \mathcal{N}_{e,b}} \phi_i &\leq 1 \\ \sum_{i \in \mathcal{N}_{c,b}} \varphi_i &\leq 1 \end{aligned}$$

$\bar{x}_i \geq 0, \phi_i \geq 0, \varphi \geq 0$

where $\mathcal{N}_{c,b}$ is the set of center users in cell b and $\mathcal{N}_{e,b}$ is the set of edge users. Both sets have cardinalities of N_c and N_e respectively. The variables ϕ_i and φ_i denote the asymptotic fraction of the resources that are utilized by user i and $\bar{x}_i$ is the resulting throughput. Eq. 4.2 will yield the optimal scheduling of resources (ϕ^*, φ^*) with respect to the implemented scheduling metric. In fractional frequency reuse cases, the number of LI-RBs is $d_b = \lfloor \frac{L}{K} \rfloor$, where $K = 3$ is a typical value and $d_b = \lfloor . \rfloor$ denotes the floor operation. Hence, the corresponding cell utility is given by $F_b(d_b)$.

Proposition 4.1 The α-Proportional Fair optimal throughput of the users satisfies one of the following:

1. $\bar{x}_i = L\frac{\mu_i^{\frac{1}{\alpha}}}{\sum_{i\in\mathcal{N}_{c,b}\cup\mathcal{N}_{e,b}}\mu_i^{\frac{1}{\alpha}-1}}$, $i \in \mathcal{N}_{c,b} \cup \mathcal{N}_{e,b}$

   for $l \geq \frac{\sum_{i\in\mathcal{N}_{e,b}}\mu_i^{\frac{1}{\alpha}-1}}{\sum_{i\in\mathcal{N}_{c,b}\cup\mathcal{N}_{e,b}}\mu_i^{\frac{1}{\alpha}-1}}L$

2. $\bar{x}_i = \begin{cases} l\frac{\mu_i^{\frac{1}{\alpha}}}{\sum_{i\in\mathcal{N}_{e,b}}\mu_i^{\frac{1}{\alpha}-1}}, & i \in \mathcal{N}_{e,b} \\ (L-l)\frac{\mu_i^{\frac{1}{\alpha}}}{\sum_{i\in\mathcal{N}_{c,b}}\mu_i^{\frac{1}{\alpha}-1}}, & i \in \mathcal{N}_{c,b} \end{cases}$

   for $l < \frac{\sum_{i\in\mathcal{N}_{e,b}}\mu_i^{\frac{1}{\alpha}}}{\sum_{i\in\mathcal{N}_{c,b}\cup\mathcal{N}_{e,b}}\mu_i^{\frac{1}{\alpha}-1}}L$

Corollary 4.1 The aggregate utility has the property that

$$U_b(l) = U_b(l_{max,b}), \quad l \geq l_{max,b} \tag{4.2}$$

where

$$l_{max,b} = \left\lceil \frac{\sum_{i\in\mathcal{N}_{e,b}}\mu_i^{\frac{1}{\alpha}-1}}{\sum_{i\in\mathcal{N}_{c,b}\cup\mathcal{N}_{e,b}}\mu_i^{\frac{1}{\alpha}-1}}L \right\rceil \tag{4.3}$$

and $\lceil . \rceil$ denotes the ceiling (round up) operation. The proof follows directly from Proposition 5.1.

The scheduling problem can be easily extended to the case in which the data rates of the users are confined to some fixed interval: $\bar{x}_{min,i} \leq \bar{x}_i \leq \bar{x}_{max,i}$. The associated optimization problem is called constrained rate cell wise scheduling problem. It is noted that the objective function (aggregate utility) is concave and the set of constraints are linear. Hence the constrained rate cell wise scheduling problem can be solved easily with standard numerical tools.

Proposition 4.2 The solution to the constrained rate cell wise scheduling exists if

$$l \geq \frac{1}{\sum_{i\in\mathcal{N}_{e,b}}\frac{\bar{x}_{min,i}}{\mu_i}} \geq l_{min,b} \text{ and} \tag{4.4}$$

$$\frac{1}{\sum_{i\in\mathcal{N}_{c,b}}\frac{\bar{x}_{min,i}}{\mu_i}} + \frac{1}{\sum_{i\in\mathcal{N}_{e,b}}\frac{\bar{x}_{min,i}}{\mu_i}} \leq L \tag{4.5}$$

4.3 Inter-cell Coordination

4.3.1 Nash Bargaining

Due to minimum rate constraints, each cell requires at least LI-RBs. If $\sum_{b\in\mathcal{B}} l_{min,b} \leq L$, where $\mathcal{B}$ is the base-station index, there then exists a surplus of resource blocks that the cells can divide between themselves. In economics, a common approach to resource allocation problems is to formulate them as collaborative bargaining games. Several different bargaining game formulations have been proposed that differ based on the utilized axioms. The most famous is the Nash bargaining game [55].

Definition 4.1
A bargaining solution $F(U, d)$ is called Nash bargaining solution if it satisfies the following axioms

- *Symmetry.* A bargaining solution F satisfies symmetry if for all symmetric bargaining problems (U,d)
  $(z_1, z_2) \in F(U, d) \Leftrightarrow (z_1, z_2) \in F(U, d)$

- *Pareto optimality.* A solution $X(U, d)$ satisfies Pareto optimality if for all bargaining problems (U, d), $F(U, d)$ is a subset of the weakly efficient points in U. It satisfies Pareto optimality if for all bargaining problems (U, d), $F(U, d)$ is a subset of the efficient points in U

- *Invariance.* A bargaining solution satisfies invariance if whenever (U', d') is obtained from the bargaining problem (U, d) by means of the transformations $z_b \rightarrow \varrho_b z_b + \beta_b$, for $b = 1, 2$,where $\beta_b > 0$ and $\kappa_b \in \mathbb{R}$, we have that $F_b(U', d') = \varrho_b F_b(U, d) + \kappa_b$, for $b = 1, 2$.

- *IIA.* A bargaining solution F satisfies independence of irrelevant alternatives if $F(U', d) = F(U, d) \cap U'$ whenever $U' \subseteq U$ and $F(U, d) \cap U' \neq \emptyset$.

The Independent from Irrelevant Alternatives Axiom (IIA) is slightly controversial, since it has been argued that the more alternatives an agent has, the better bargaining power he or she has. Some authors suggest alternative solutions that do not use this axiom like [94]. Dargan *et al.* however, proposed replacing this axiom by three independent properties after adding the following property to the axioms [95].

- *Single-valuedness in symmetric problems.* A bargaining solution F satisfies single-valuedness in symmetric problems if for every symmetric problem $V \in \mathcal{V}$, $F(V)$ is a singleton.

The three alternative properties for the IIA axiom are

- *Independence of non-individually rational alternatives.* A solution satisfies independence with respect to non-individually rational alternatives if for every two problems (U, d) and (U', d) such that $F\ \mathcal{V}(U, d) = F\mathcal{V}(U', d)$ we have $F(U, d) = F(U', d)$.

- *Twisting.* A bargaining solution F satisfies twisting if the following holds: Let (U, d) be a bargaining problem and let $(\hat{z}_1, \hat{z}_2) \in J(U, d)$. Let (U', d) be another bargaining problem such that for some agent $b = 1, 2$
  $U \setminus U' \subseteq \{(z_1, z_2) : z_b > \hat{z}_b\}$
  $U' \setminus U \subseteq \{(z_1, z_2) : z_b < \hat{z}_b\}$
  Then, there is $(U'_1, U'_2) \in M(U', d)$ such that $s'_b \leq \hat{z}_b$.

- *Disagreement point convexity.* A bargaining solution F satisfies disagreement point convexity if for every bargaining problem $V = (U, d)$ for all $z \in F(U, d)$ and for every $\psi \in (0, 1)$ we have $U \in F(U, (1 - \psi)d + \psi z)$.

Theorem 4.1
The only function $F(U, U(d)$ satisfying the axioms given by Definition 4.1 is the Nash barging solution given by [96]

$$F(U, U(d)) = \arg\max_{l \in L} \prod_b (U_b(l_b) - U_b(d_b)) \tag{4.6}$$

Parento optimality implies that any bargaining solution should be better than the initial agreement point d. In the cell coordination case, the initial agreement point can refer to the allocation that merely guarantees minimum rates for the edge user or $d_b = l_{min,b}$ it can refer to fractional reuse case, in which the reuse $k = 3$ is utilized to serve the edge users. In such case $d_b = \lfloor \frac{L}{k} \rfloor$.

The proposition below provides the feasibility condition for the cell coordination problem.

Proposition 4.3 The cell-coordination problem is feasible if $\sum_{b \in \mathcal{B}} l_{min,b} \leq L$,

$d_b \geq l_{min,b}$ and $\frac{1}{\sum_{i \in \mathcal{N}_{c,b}} \frac{\bar{x}_{min,i}}{\mu_i}} + \frac{1}{\sum_{i \in \mathcal{N}_{e,b}} \frac{\bar{x}_{min,i}}{\mu_i}} \leq L$ holds for all cells b.

The cell coordination problem can be stated in terms of Multiple-Choice Knapsack packing Problem (MCKP) as follows

$$\max_{a} \sum_{b} \sum_{f \in \mathcal{K}_b} v_{b,f} a_{b,f} \tag{4.7}$$

s.t.

$$\sum_{f \in \mathcal{K}_b} a_{b,f} \leq 1$$
$$\sum_{f \in \mathcal{K}_b} s_{b,f} a_{b,f} \leq L$$

where

$$\begin{aligned} v_{b,f} &= \log[U_b(f) - U_b(d_b)], \quad f \in \mathcal{L}_b \\ s_{b,f} &= f \\ \mathcal{L}_b &= \{l_{min,b}, l_{min,b}+1, \ldots, L\} \end{aligned}$$

Definition 4.2
If two allocations $\theta, \vartheta \in \mathcal{L}_b$ satisfy that $s_{b,q} \leq s_{b,\theta}$ and $a_{b,\vartheta} \geq a_{b,\theta}$, then we say that ϑ dominates θ.

It is known that for MCKP, if ϑ dominates θ, then the optimal solution satisfies $a_{b,\theta} = 0$ [97]. Therefore the search space can be pruned by throwing out all items for which $a_{b,f} \leq a_{b,f'}$, $f > f'$. By Corollary 4.1 there exists $l_{max,b}$ such that $U(l_{max,b}) = U(f)$, $\forall f > l_{max,b}$, hence the search space can be limited to $\mathcal{L}_b = \{l_{min,b}, l_{min,b}+1, \ldots, l_{max,b}\}$.

Proposition 4.4 The Cell-coordination problem can be solved using a greedy algorithm in polynomial time.

Algorithm 4.1 Greedy approach

1. Set $a_{b,f} = \begin{cases} 1 & f = l_{min,b} \\ 0 & \text{otherwise} \end{cases}$ for all cells b

Set $S = \sum_{b\in\mathcal{B}} l_{min,b}$ and $V = \sum_{b\in\mathcal{B}} v_{b,l_{min,b}}$.

2. Define $\psi_{b,l} = \log[U_b(l) - U(l-1)]$,
   $f = \{l_{min,b}, l_{min,b}+1, \ldots, l_{max,b}$ and order the slopes $\{\psi_{b,f}\}$ in a non-decreasing order.

3. Let h, f be the indices corresponding to the next slope in $\{\psi_{h,f}\}$. If $S+s_{h,f} > L$ Stop; otherwise set $a_{h,f} = 1$ and $a_{h,(f-1)} = 0$ and update the sums $S := S+s_{h,f}$ and $V := V + v_{h,f}$. Repeat step 3.

The order of complexity for bargaining for one cell will therefore be $O(N_b.C - l_{min,b})$ and is carried out in B cells. In practice, the method will work reasonably well for moderate number of users.

4.3.2 Load Balancing Handovers

Instead of moving frequency resources between the cells by means of cell co-ordination, the load among the cells can be balanced by performing load balancing handovers [98].

It is assumed that an edge user i is able to make several handovers during the scheduling interval such that it can be connected to base station b arbitrary fraction of the scheduling interval. Mobility induced handovers are omitted due to the assumption that the channels remain stationary during the scheduling process.

Consider the following load balanced scheduling problem

$$F(d) = \max_{\phi,\varphi} \sum_{b\in\mathcal{B}} \sum_{i\in\mathcal{N}_{c,b}\cup\mathcal{N}_{e,b}} (U_\alpha[\bar{x}_i]) \tag{4.8}$$

s.t.

$$\bar{x}_i = \begin{cases} \sum_{b\in+mathcalB} \mu_{ji} d_b \phi_{ji} & \text{if } i \in \mathcal{N}_{e,b} \\ \mu_i \left(d_b\phi_i + (L-d_b)\varphi_i\right) & \text{if } i \in \mathcal{N}_{c,b} \end{cases}$$

$\bar{x}_{min,i} \leq \bar{x}_i \leq \bar{x}_{max,i}\ \forall i$

$$\begin{aligned}
\sum_{i\in\mathcal{N}_{e,b}} \phi_{i,b} + \sum_{i\in\mathcal{N}_{c,b}} \phi_i &\leq 1, \quad \forall b \in \mathcal{B} \\
\sum_{i\in\mathcal{N}_{c,b}} \varphi_i &\leq 1, \quad \forall b \in \mathcal{B} \\
\sum_{b\in\mathcal{B}} \phi_{i,b} &\leq 1, \quad i \in \mathcal{N}_E
\end{aligned}$$

$\bar{x}_i \geq 0$, $\phi_i \geq 0$, $\phi_{ji} \geq 0$, $\varphi \geq 0$ $\forall b, i$

where $\mathcal{N}_E = \cup_{b\in\mathcal{B}}\mathcal{N}_{e,b}$. The above problem is concave optimization and can be easily solved with standard numerical techniques.

4.3.3 Joint Nash bargaining and load balancing

This section combines the Nash bargaining solution with the load balancing solution. The problem for a two cell case can be written in a MCKP form as follows.

$$\max_a \sum_{f\in\mathcal{L}_2} \sum_{w\in\mathcal{L}_1} v_{w,f} a_{w,f}$$

s.t.

$$\begin{aligned}
a_{w,f} &\leq 1 \\
\sum_{f\in\mathcal{L}_2} \sum_{s\in\mathcal{L}_1} w_{s,f} a_{s,f} &\leq L
\end{aligned}$$

where

$$\begin{aligned}
v_{w,f} &= \log[U(s,f) - U(d)], \quad w \in \mathcal{L}_1, f \in \mathcal{L}_2 \\
s_{w,f} &= w + f \\
\mathcal{L}_b &= \{l_{min,b}, l_{min,b}+1, \ldots, L\}
\end{aligned}$$

Similar to Algorithm 4.1, we can write;

Algorithm 4.2 Joint greedy approach

1. Set $a_{w,f} = \begin{cases} 1 & w = l_{min,1}, f = l_{min,2} \\ 0 & \text{otherwise} \end{cases}$

   Set $\mathcal{S} = \sum_{b \in \mathcal{B}} l_{min,b}$ and $\mathcal{V} = v_{b,l_{min,1},l_{min,2}}$.

2. Define $\psi_{w,f} = \log[U(w,f) - U(w-1,f-1)]$,
   $w = l_{min,1}, l_{min,1}+1, \ldots, l_{max,1}$,
   $f = l_{min,2}, l_{min,2}+1, \ldots, l_{max,2}$ with U defined in (4.8) and order the slopes $\{\rho_{w,f}\}$ in a non-decreasing order.

3. Let w, f be the indices corresponding to the next slope in $\{\rho_{w,f}\}$. If $S+s_{w,f} > L$ Stop; otherwise set $a_{w,f} = 1$ and $a_{w-1,f-1} = 0$ and update the sums $S := S + s_{w,f}$ and $V := V + v_{w,f}$. Repeat step 3.

4.3.4 Bargaining objective function

The resource division among the cells can be viewed as a group bargaining problem in which users belonging to a particular cell, form a group or coalition and negotiating with the other groups (or coalitions). The group bargaining problem has been discussed in economical theories and mathematical social science [99]. At least three different targets for the bargaining process in (4.6) can be identified:

1. Every player represents himself and we have a $\sum_{b=1}^{N_G}(\mathcal{N}_{e,b} + \mathcal{N}_{c,b}) = n$ player bargaining game with N_G representing the number of groups bargaining. Projecting this criteria on the system, the objective function can be, for example, the product of the users' utilities.

$$U_b(l) = \prod_{i \in \mathcal{N}_{e,b} \cup \mathcal{N}_{c,b}} (U_\alpha[\bar{x}_i(l, \phi_i^*, \varphi_i^*)])$$

2. Each group has a single representative who represents the whole group in the N_G player bargaining game. The user should be selected to maximize the bargaining power of the group. One example would be the user with the lowest utility value.

$$U_b(l) = \min_{i \in \mathcal{N}_{e,b} \cup \mathcal{N}_{c,b}} \{U_\alpha[\bar{x}_i(l, \phi_i^*, \varphi_i^*)]\}$$

3. A new player is constructed for each group who then represents the whole group. Zhang has suggested that the new player who can represent the group

can be obtained by simply taking the average of the utilities [100].

$$U_b(l) = \frac{1}{|N_{e,b}| + |N_{c,b}|} \sum_{i \in \mathcal{N}_{e,b} \cup \mathcal{N}_{c,b}} (U_\alpha[\bar{x}_i(l, \phi_i^*, \varphi_i^*)])$$

From a signaling point of view, target 2 requires the least amount of signaling since only one utility is reported per cell. In targets 1 and 3, the base-stations need to share the information of every user in their cells leading to a considerable amount of signalling.

4.4 Computational Evaluation

The performances of the schemes have been evaluated in terms of the uplink system throughput, the aggregate throughput of all the edge users and the average delay. The result is averaged over multiple simulations. A three-cell case is considered where the base-stations are situated in the middle of the cell and have 120 degrees antennas dividing their cells into 3 sectors as shown in Fig. 4.1. Each cell has a distinct set of edge bands that are orthogonal to the edge bands of the neighboring cells. This is depicted in the different edge patterns in Fig. 4.1. The neighboring sectors have different loads of users and it is assumed that there is one lightly loaded cell, one medium loaded and one heavily loaded and users are divided into center and edge users. System parameters are shown in Table 4.1. Perfect channel estimation is assumed so retransmissions are not needed. A scheduling frame of 10 ms is used where, at the beginning of every frame, the bargaining procedure is carried out based on the current information. Simulations are carried out for 4 schemes which are: Static frequency reuse (SR) with a reuse factor K=3, load balancing handovers (LBH), Nash bargaining (NB) and, finally, Nash bargaining with load balancing handovers (NB+LBH). Figures 4.2 and 4.3 show the system throughputs and edge users throughput respectively with different objective functions described in Section 4.3.4. The average and 5th percentile user throughput are displayed in Figures 4.4 and 4.5 respectively as well as the the average packet delay and 80th percentile delay in Figures 4.6 and 4.7 respectively. It can be seen that bargaining clearly enhances system performance and can be further improved by adding the load balancing handovers technique. The impact of the different objective functions can also be seen.

Table 4.1: System parameters

Parameter	Value
RB bandwidth	375 kHz
Total number of blocks	9 (25 subcarrier/RB)
TTI duration	1 ms
Radio propagation	Site to site distance 100 m
Max. Tx power	21 dBm
BS antenna gain	18 dBi
UE antenna gain	0 dBi
Noise power	-108.5 dBm
Number of users	$3/BS_1$, $5/BS_2$, $7/BS_3$
Traffic model	Full buffer

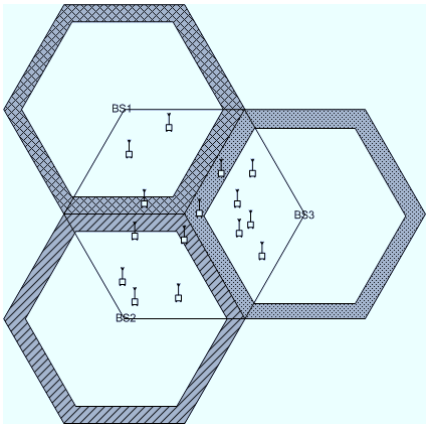

Figure 4.1: Cellular model

The results demonstrate the impact of individual and group bargaining. Similar performance is noticed in the case where a certain utility function is elected to represent the group (Objective: average utility) and the case where every user bargains for himself (Objective: product of utilities). In the case of the minimum utility, it can be seen that bargaining alone, although provides better performance when compared to the static reuse case, is less than load balancing handovers. Nevertheless, when both techniques are merged it was found that the minimum utility gains a good performance.

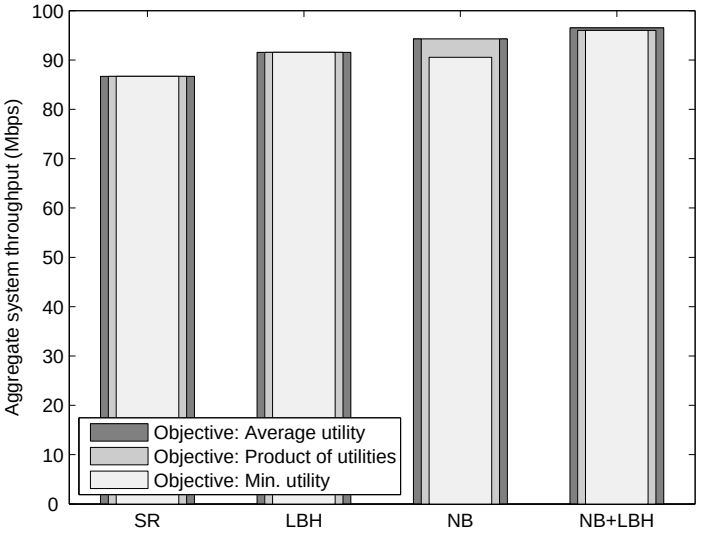

Figure 4.2: System aggregate throughput

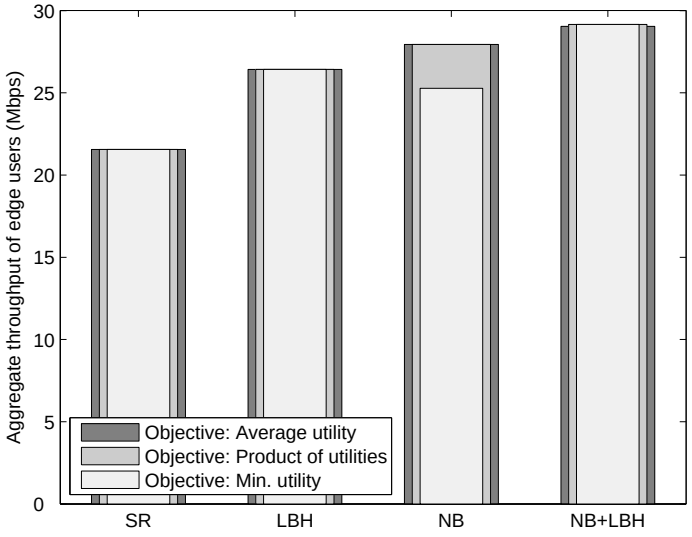

Figure 4.3: Aggregate throughput of the edge users for all cells

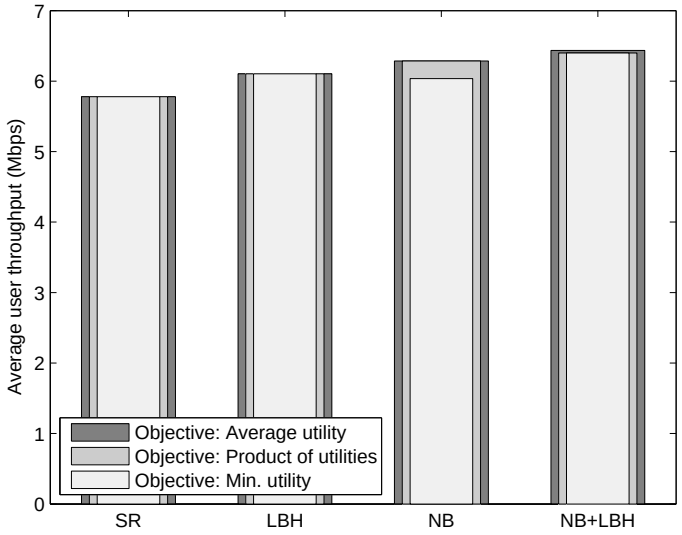

Figure 4.4: The average throughput a user obtains in the system

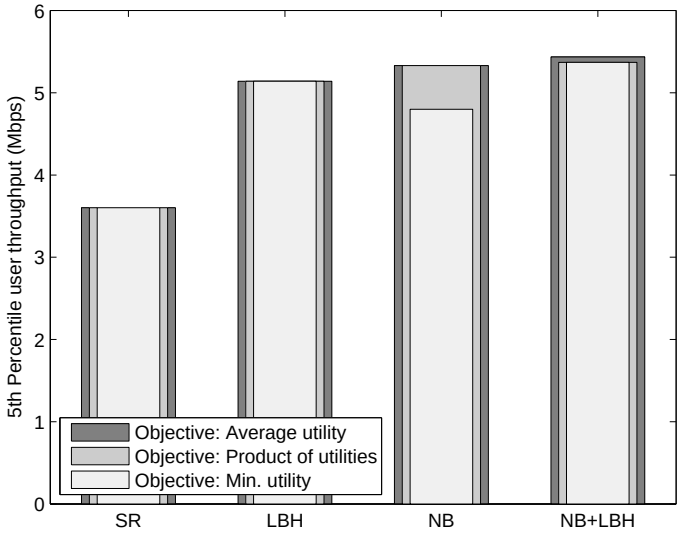

Figure 4.5: 5th Percentile for the throughput of the users in the system

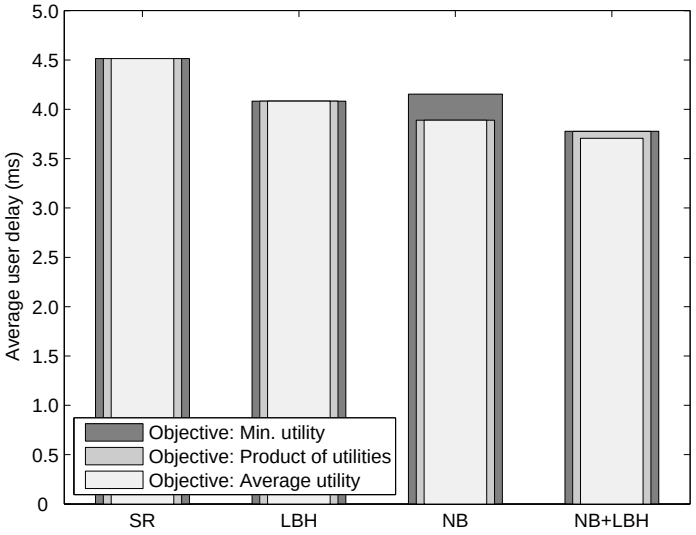

Figure 4.6: The average delay a user experiences in the system

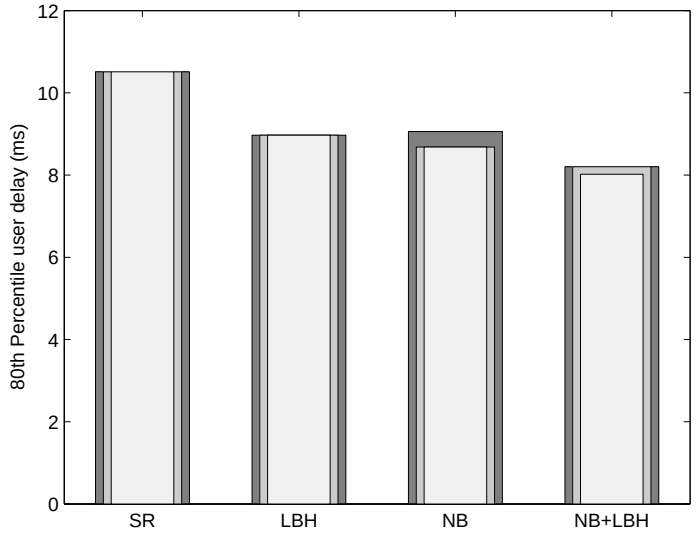

Figure 4.7: 80th Percentile for the delay of the users in the system

4.5 Concluding Remarks

In this chapter a proposal was made that can mitigate the effect of interference suffered by users situated on the cell edge through bargaining. Neighboring cells would negotiate with each other to obtain the necessary resources that would maximize certain objective functions. The negotiation was presented in the form of Nash bargaining with the bargaining taking place in every scheduling frame. This resembled a case of having dynamic reuse where the reuse factor K is changed in accordance with the bargain. The closed form solution was also provided and proved for the bargaining problem of non rate-constrained situations. For rate-constrained cases, providing a closed form solution is very difficult because of the growing complexity. Performance of the bargaining scheme was further improved by merging it with load balancing handovers which introduced a gain of approx. 10% in system performance. The impact of different objectives for the bargaining was also demonstrated. It was shown through simulations that it is feasible to have one representative for a group of users in a cell to carry out the bargaining with other cells, thus reducing the amount of signalling between base-stations. In conclusion, bargaining proves to be a powerful way to maximize the utilization of shared resources.

Part II
Activity Control

Chapter 5

Activity control

In this chapter the concept of activity control is proposed where the activity of a user is subject to the proximity of a certain QoS level. In cellular systems, users are considered active if they have data in their transmission buffers, otherwise the scheduler considers them to be non-active and are excluded from the scheduling process. In this chapter an attempt is made to utilize the activity control concept in some of the main RRM mechanisms. First, an admission control scheme is proposed, that mainly controls the activity of a new user entering the system by assigning a back-off factor to limit the maximum impact the new user can cause to ongoing connections in terms of throughput loss. Also in this chapter, an activity controlled scheduler is proposed for flows that have target QoS levels, such as mean data rates or mean HOL delays. This corresponds, for instance, to the case where a leaky bucket traffic shaping filter is applied at the edge router of the radio access network. For the sake of simplicity, we assume a single-channel system such as HSDPA. However without loss of generality, the results obtained in this chapter are also valid for multi-channel systems.

5.1 Admission Control

Consider the downlink of a time-slotted system where time is the resource to be shared among all users. The channel is assumed to be stationary, more specifically, the fast fading process is assumed to be stationary and ergodic so that the time average approaches the mean value as the number of samples increases. The fading distributions of the users can be different but their channel should be stationary. This assumption is later relaxed in Section 5.8 where parameter estimation is dis-

cussed and the case of where the channel statistics are slowly changing is considered. The UE is assumed to estimate the channel based on a pilot signal transmitted by the base-station. Based on the channel measurement, the UE then determines the maximum data rate $\mu_i(t)$ that it can achieve under the current channel condition at time-slot t, and reports this back to the base-station which then utilizes the information in the scheduling decision.

Let $\mathcal{A}(t)$ 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 $\mu_i(t), i \in \mathcal{A}(t)$ and the estimate of the long term average throughput $\bar{x}_i(t)$ using a simple selection rule of the form:

$$i^*(t) \in \text{argmax}_{i \in \mathcal{A}(t)} F(\mu_i(t), \bar{x}_i(t)) \tag{5.1}$$

where $F(\mu_i(t), \bar{x}_i(t))$ is some non-decreasing function of $\mu_i(t)$ and $\bar{x}_i(t)$ is updated as follows:

$$\bar{x}_i(t+1) = (1-\beta)\bar{x}_i(t) + \beta\mu_i(t)\chi(i = i^*(t)) \tag{5.2}$$

where $\beta > 0$ is a fixed (small) parameter. The operator $\chi\{L\}$ is an indicator function of an event L: $\chi\{L\} = 1$ if the event L occurs and zero otherwise. It can be seen from this definition that, $\bar{x}_i(t)$ 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 (5.2) yields:

$$\mathbb{E}\{\bar{x}_i(t+1)\} = (1-\beta)\mathbb{E}\{\bar{x}_i(t)\} + \beta\mathbb{E}\{\mu_i(t)\chi(i = i^*(t))\} \tag{5.3}$$

The scheduler is very general and contains all the memoryless scheduling rules suggested in [101], as well as gradient scheduling rules suggested in [20] as special cases. Thus, the two types of schedulers mentioned in Chapter 2 are further defined as follows:

1- Memoryless Schedulers

A memoryless policy is a stationary policy whose decision does not depend on time-slot t 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 $\mathbb{E}\{\mu_i(t)\chi(i = i^*(t))\}$ is also stationary as long as

the selection process (5.1) which is a function of $\mu_i(t)$ and $\bar{x}_i(t)$ is stationary. Since $\mu_i(t)$ is assumed to be stationary, consequently the selection process is stationary as long as the estimate $\bar{x}_i(t)$ is stationary. Let us define $\bar{x}_i(t) = \mathbb{E}\{\mu_i(t)\chi(i = i^*(t))\}$ and $\tilde{x}_i(t) = \mathbb{E}\{\bar{x}_i(t)\}$. If we assume that $\bar{x}_i(t) = \bar{x}_i, \forall t$ then $\tilde{x}_i(t) = \bar{x}_i, \forall t$. This will allow the implementation of the admission control scheme in a slot-by-slot fashion.

2- Schedulers with Memory

Schedulers with memory are dynamic schedulers whose policy depend on time-slot t. In this case the memory of the scheduler contains information of past throughput $\bar{x}_i(t)$. For a memory scheduler, the rate in (5.3) can be written as:

$$\begin{aligned}
\tilde{x}_i(1) &= (1-\beta)\bar{x}_i(0) + \beta\bar{x}_i(0) = \bar{x}_i(0) \\
\tilde{x}_i(2) &= (1-\beta)\tilde{x}_i(1) + \alpha\bar{x}_i(1) = (1-\beta)\bar{x}_i(0) + \alpha\bar{x}_i(1) \\
\tilde{x}_i(3) &= (1-\beta)\tilde{x}_i(2) + \beta\bar{x}_i(2) = (1-\beta)^2\bar{x}_i(0) + (1-\beta)\bar{x}_i(1) + \beta\bar{x}_i(2) \\
&\vdots \\
\tilde{x}_i(t) &= \sum_{m=0}^{t-1}(1-\beta)^m\beta\bar{x}_i(t-1-m)
\end{aligned} \tag{5.4}$$

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

5.2 Single-user Iterative Admission Control

This section describes the admission control scheme with its application to the two types of schedulers.

5.2.1 Implementation in Memoryless Schedulers

Assume that a new user tries to join the system at time-slot t_0. At time-slot $t > t_0$, the new user is excluded from the active set $\mathcal{A}$ with a back-off probability p_b and included with probability $1 - p_b$. 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. Thus, $\bar{x}_i(Z_{N+1})$ represents the mean rate of user i with $N+1$ users in the system, i.e. with the new user fully admitted. At timeslot t_0 the new user $N+1$ is 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 $\tilde{x}_i(t) = \bar{x}_i = \bar{x}_i(Z_N), \forall t \leq t_0$. Based on the analysis of the mean rates, it is possible to observe the impact of the new user on the active users in the following manner:

$$\tilde{x}_i(t) = p_b\bar{x}_i(Z_N) + (1-p_b)\bar{x}_i(Z_{N+1}) \quad i = 1, 2, \cdots N, \quad t \geq t_0 \tag{5.5}$$

$$\tilde{x}_{N+1}(t) = (1-p_b)\bar{x}_{N+1}(Z_{N+1}) \quad t \geq t_0 \tag{5.6}$$

and

$$\tilde{x}_i(t) = \bar{x}_i(Z_N) \quad i = 1, 2, \cdots N, \quad t < t_0 \tag{5.7}$$

$$\tilde{x}_{N+1}(t) = 0 \quad t < t_0 \tag{5.8}$$

giving,

$$\tilde{x}_i(t_0) - \tilde{x}_i(t_0 - 1) = -(1-p_b)(\bar{x}_i(Z_N) - \bar{x}_i(Z_{N+1})) \quad i = 1, 2, \cdots N \tag{5.9}$$

$$\tilde{x}_{N+1}(t_0) - \tilde{x}_{N+1}(t_0 - 1) = (1-p_b)\bar{x}_{N+1}(Z_{N+1}) \tag{5.10}$$

Thus, in a slot the impact of the new user is limited to the fraction $1 - p_b$. This behavior resembles the active link protection (ALP) scheme suggested by Bambos for power controlled systems [67]. Therefore, the scheme will be referred to as the ALP-CAC.

5.2.2 Implementation in 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 l slots. Assuming l is large, so that the

estimator will converge.

$$\tilde{x}_i(t_0 + l) \rightarrow p_b \bar{x}_i(Z_N) + (1 - p_b)\bar{x}_i(Z_{N+1}) \tag{5.11}$$

similarly,

$$\tilde{x}_i(t_0 + l) - \tilde{x}_i(t_0) = -(1 - p_b)(\bar{x}_i(Z_N) - \bar{x}_i(Z_{N+1})) \geq -(1 - p_b)\bar{x}_i(Z_N) \tag{5.12}$$

The rate has dropped by no less than a factor $1 - p_b$.

5.2.3 Iterative Admission Control

Decreasing the back-off probability p_b exponentially result in having a back-off probability of p_b^n at timeslot t_n (or frame n for schedulers with memory) as shown below:

$$\begin{aligned}
\tilde{x}_i(t_1) &= p_b \bar{x}_i(Z_N) + (1 - p_b)\bar{x}_i(Z_{N+1}) && (5.13)\\
\tilde{x}_i(t_2) &= p_b^2 \bar{x}_i(Z_N) + (1 - p_b^2)\bar{x}_i(Z_{N+1}) && (5.14)\\
&\vdots \\
\tilde{x}_i(t_n) &= p_b^n \bar{x}_i(Z_N) + (1 - p_b^n)\bar{x}_i(Z_{N+1}) && (5.15)
\end{aligned}$$

From the above equations, it is possible to write the rate obtained in a timeslot or frame in terms of the rate obtained in the previous one. This way, the equations can be solved with the help of the Kalman filter estimate since the value of $\bar{x}_i(Z_{N+1})$ is not known.

$$\begin{aligned}
\tilde{x}_i(t_2) &= p_b \tilde{x}_i(t_1) + (1 - p_b)\bar{x}_i(Z_{N+1}) > p_b \tilde{\mu}_i(t_1) \\
&\vdots \\
\tilde{x}_i(t_n) &= p_b \tilde{x}_i(t_{n-1}) + (1 - p_b)\bar{x}_i(Z_{N+1}) > p_b \tilde{x}_i(t_{n-1}) && (5.16)
\end{aligned}$$

Hence, in a timeslot (or frame) the rate does not drop more than by a factor $(1-p_b)$. The size of p_b limits the impact of the new user and therefore the new user can be rejected if $\tilde{x}_i(t_n) < \frac{\bar{x}_{min}}{p_b}$. This impact can be clearly seen in Fig. 5.1 where the theoretical rates were computed with two different values of p_b. It can be seen that with $p_b = 0.99$, the new user does cause the worst active user to drop below the

threshold $\frac{\bar{x}_{min}}{p_b}$ unlike the case of $p_b = 0.9$.

$$\tilde{x}_i(t_n) < \tilde{x}_i(t_{n-1}) \tag{5.17}$$

Similarly for the new user we can write

$$\tilde{x}_{N+1}(t_1) = (1 - p_b)\bar{x}_{N+1}(Z_{N+1}) \tag{5.18}$$

and

$$\tilde{x}_{N+1}(t_n) = \tilde{x}_{N+1}(t_{n-1}) + p_b^{n-1}(1 - p_b)\bar{x}_{N+1}(Z_{N+1}) > \tilde{x}_{N+1}(t_{n-1}) \tag{5.19}$$

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

$$\tilde{x}_{N+1}(t_n) > \tilde{x}_{N+1}(t_{n-1}) \tag{5.20}$$

According to the analysis above an iterative control scheme that is applicable to both memoryless schedulers and schedulers with memory can be implemented. An outline for 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* $\tilde{x}_i(t_n) < \frac{\bar{x}_{min}}{p_b}$ *for any* $i = 1, 2, ..., N$ *at some iteration* m*. Otherwise* $p_b^m \to 0$ *in (5.15) and the user is admitted to the network.*

In practice, $\bar{x}_i(Z_{N+1})$ is unknown and the fact that a new user starts from a rate of 0 leads to a transient state that in turn will affect $\tilde{x}_i(t_m)$. Therefore, it can be difficult to base the decision of admission solely on the criteria $\tilde{x}_i(t_m) < \frac{\bar{x}_{min}}{p_b}$. Thus, there is a need for a fixed decision time for admission. If $\tilde{x}_i(t_m)$ at the decision time is found to be greater than $\bar{x}_{min}$, then the new user should be admitted, otherwise rejected. The order of complexity of the scheme can thereby derived: $O(N.n)$, with n depending on the decision time making the scheme possible to realize in practice.

5.2.4 Non-stationarity

This subsection discusses a slot-by-slot approach 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

$$i^*(t) = \text{argmax}_{i\in\mathcal{A}(t)}\{\nabla F(\bar{x}_i(t))\mu_i(t)\} \tag{5.21}$$

where $\mu_i(t)$ is the actual data rate process (equivalently the instantaneous channel state), $\bar{x}_i(t)$ denotes the estimated throughput and $\mathcal{A}(t)$ denotes the set of active users at time-slot t. Assume that the actual data rate process is wide sense stationary and ergodic. The rate estimator is assumed to be unbiased. It is assumed that F is a non-increasing function of $\bar{x}_i(t)$ and a non-decreasing function of $\mu_i(t)$. Let us define

$$\begin{aligned} \urcorner_i(\mu_i(t), \bar{x}_i(t), \mathcal{A}(t)) &= 1 \quad i = \text{argmax}_{i\in\mathcal{A}(t)}\{\nabla F(\bar{x}_i(t))\mu_i(t)\} \\ &= 0 \quad \text{otherwise} \end{aligned} \tag{5.22}$$

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

$$\bar{x}_i(t) = \mathbb{E}\{\mu_i(t)\urcorner_i(\mu_i(t), \bar{x}_i(t), \mathcal{A}(t))\} \tag{5.23}$$

Let us consider the saturated condition in which $\mathcal{A}(t) = \mathcal{A} = \{1, 2, ..., N\}\ \forall t$. It has been shown by Stolyar that the scheduling rule (5.21) converges $\bar{x}_i(t) \to \bar{x}_i$ which maximizes the utility function $F(\bar{x}_i(t))$ [23]. In the steady state, the scheduling rule becomes

$$i^*(t) = \text{argmax}_{i\in\mathcal{A}(t)}\{\nabla F(\bar{x}_i(t))\mu_i(t)\} \tag{5.24}$$

Lemma 5.1
For $\mathcal{A}_N \subseteq \mathcal{A}_{N+1}$, we have

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

and

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

Hence, it was possible to write the expected rate in a slot-by-slot basis with the help of the steady lower bound mean rate $\bar{x}_{lower,i}$ that asymptotically converges with the expected rate $\tilde{x}_i$. Fig. 5.2 is presented for illustration purposes of the above analysis. It shows how that the rate is always bounded and eventually the bounds converge to the mean rate.

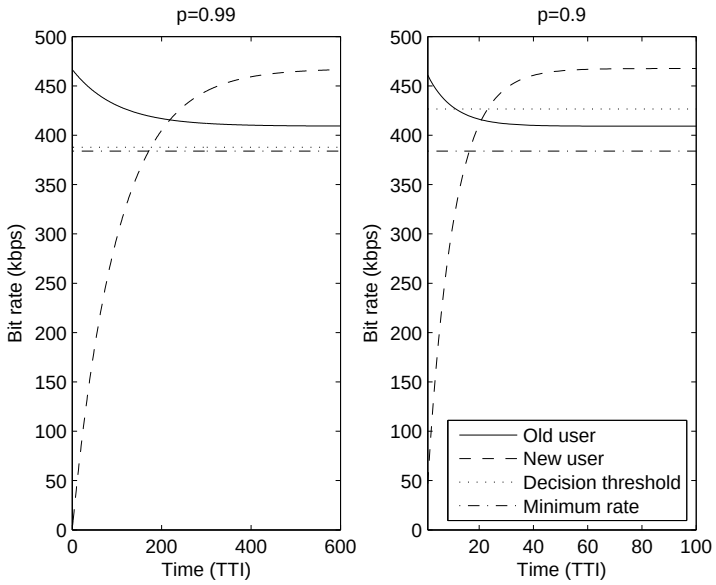

Figure 5.1: Theoretical rates obtained for old and new users with back-off probabilities 0.99 and 0.9

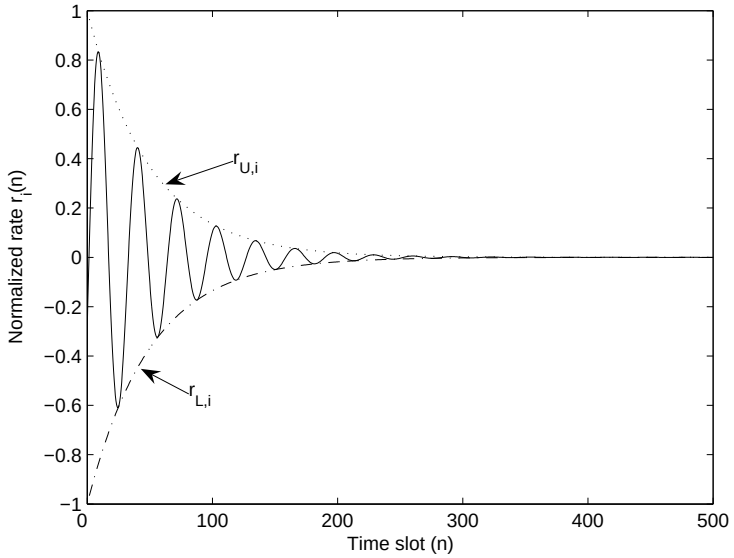

Figure 5.2: Illustration of the non-stationarity behaviour

5.3 Multi-user Iterative Admission Control

Let us limit the number of new users to M. The back-off probability is now set as $p_b^{\frac{1}{M}}$. 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 fading vectors 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

$$\tilde{x}_i^{(n)} = \sum_{t=0}^{M} p_b^{\frac{nk}{M}} \left(1 - p_b^{\frac{n}{M}}\right)^{M-t} \sum_{Z \in \mathcal{Z}_{N+M-t}} \bar{x}_i(Z) \tag{5.25}$$

Notice that when $M = 1$ (5.25) is reduced to (5.5). An example for implementing (5.25) can be seen in Fig. 5.3. For $0 < p_b < 1$, we have $(1 - p_b^{\frac{n+1}{M}})^m \geq (1 - p^{\frac{n}{M}})^m$ for all $m = 0, 1, 2, \cdots$. Hence, we have

$$\begin{aligned} \tilde{x}_i^{(n+1)} = p_b \tilde{x}_i^{(n)} + \sum_{t=0}^{M} p_b^{\frac{(n+1)t}{M}} (p^{\frac{t}{M}} (1 - p_b^{\frac{(n+1)}{M}})^{M-t} \\ - p_b (1 - p_b^{\frac{n}{M}})^{M-t}) \sum_{Z \in \mathcal{Z}_{N+M-t}} \bar{x}_i(Z) \geq 0 \end{aligned} \tag{5.26}$$

This means that the rate of an active user cannot drop more than a factor of $(1 - p_b)$ from frame to frame even if M new users are trying to get access to the system. Thus, the iterative CAC described in Section 5.2 can be directly used . Let us define two column vectors $\mathcal{X}_i(n) = \left(\bar{x}_i^{(m)}, \bar{x}_i^{(m+1)}, \cdots, \bar{x}_i^{(m+M)}\right)'$ and $J_i = \left(\bar{x}_i(Z_{N+M}), \sum_{Z \in \mathcal{Z}_{N+M-1}} \bar{x}_i(Z), \cdots, \bar{x}_i(Z_N)\right)'$ and a square matrix $G(n)$ consisting of elements $y_{nm} = p_b^{\frac{nm}{M}} \left(1 - p_b^{\frac{n}{M}}\right)^{M-m}$. Note that by $\mathcal{X}'$ we denote the transpose of the vector $\mathcal{X}$. Now we can write (5.25) in matrix form as

$$R_i(n) = Y(n) J_i \tag{5.27}$$

Matrix Y is a square matrix with full rank making it invertible. Thus, the rate prediction CAC can be used by checking whether the following holds

$$J_i = Y(n)^{-1} \mathcal{X}_i(n) \geq 1_{N+M} \bar{x}_{min} \tag{5.28}$$

where 1_m denotes a vector of m ones. It is notable however that for large m and M, Y would contain very small elements and thus its inverse Y^{-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 independent of others.

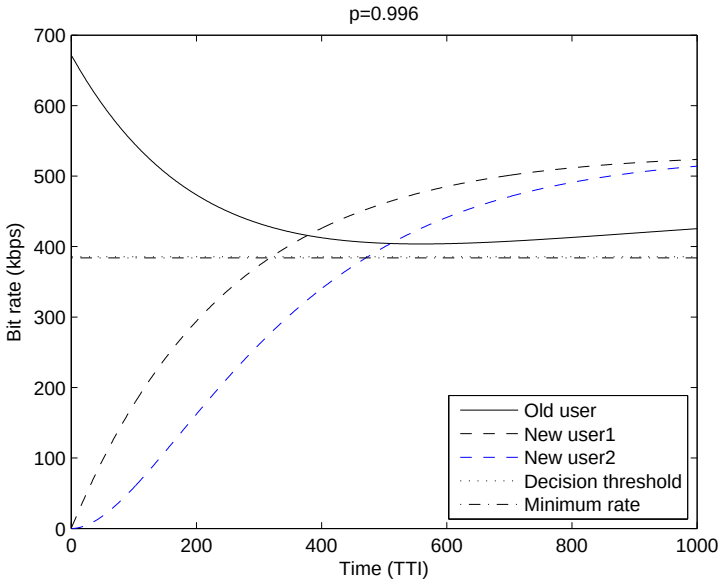

Figure 5.3: Theoretical Rates using multiple admission control with back-off probability 0.996

5.4 Kalman Filter Estimation

A Kalman filter is essentially a set of mathematical equations that implement a predictor-corrector type estimator. This estimator is optimal in the sense that it minimizes the estimated error covariance - when certain presumed conditions are met. The strength of the formulation is that it allows the tracked parameter (in this case the average rate or delay) to change slowly in time regardless of the channel's statistics.

So far it has been assumed that the mean value $\tilde{\mu}_i$ is available for admission control. In practice, we have to use time average values $\bar{x}_i$ that in ergodic channel cases would converge to $\tilde{x}_i$ when the frame length of n 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 $x_{1,i}^{(n)} = \tilde{x}_i(t_n)$ and $x_{2,i}^{(n)} = \bar{x}_i(Z_{N+1})$ and state vector $X_i^{(n)} = (x_{1,i}^{(n)}, x_{2,i}^{(n)})'$. 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 $\tilde{x}_i(t_n)$ is the estimator of the mean value, then the error $\bar{x}(t_n) - \tilde{x}(t_n)$ becomes Gaussian due to the 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

$$X_i^{(n+1)} = \Phi^{(n)} X_i^{(n)} + V_i^{(n)} \tag{5.29}$$

$$\bar{x}_i(t_n) = C\hat{X}_i^{(n)} + e_i^{(n)} \tag{5.30}$$

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

$$\hat{X}_i^{(n+1)} = \Phi^{(n)} \hat{X}_i^{(n)} + K^{(n)} \left(\bar{x}_i(t_n) - C\hat{X}_i^{(n)} \right) \tag{5.31}$$

$$K^{(n)} = (\Phi^{(n)} P^{(n)} C' + \Psi_{Ve}(t))(C' P^{(n)} C + \Psi_{ee}(t))^{-1} \tag{5.32}$$

$$P^{(n+1)} = \Phi^{(n)} P^{(n)} (\Phi^{(n)})' + \Psi_{VV}(t) - K^{(n)} (C P^{(n)} C' + \Psi_{ee}(t))(K^{(n)})' \tag{5.33}$$

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 $\Psi_{VV}(t)$, $\Psi_{ee}(t)$, and $\Psi_{Ve}(t)$ accurately. Instead they can be used as tuning parameters. The larger the parameters in $\Psi_{ee}(t)$ the less we trust in the measurements. The covariance matrix $\Psi_{VV}(t)$ 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 $\Psi_{VV}(t)$ and the more weight is given to the instantaneous channel estimate. Equation 5.3 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 (5.3) that takes into account the dynamics of the system making it better suited for the admission process.

5.5 Non-iterative Admission Control

This section presents a Recursive Least Square (RLS) scheme to later compare with the proposed 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 by Gribanova [102].

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

$$
\begin{aligned}
K(n) &= \frac{\lambda^{-1}P(n-1)u(n)}{1+\lambda^{-1}u^H(n)P(n-1)u(n)} && (5.34)\\
\epsilon(n) &= \bar{x}_i(n) - w(n-1)u(n) && (5.35)\\
w(n) &= w(n-1) + K(n)\epsilon(n) && (5.36)\\
P(n) &= \lambda^{-1}P(n-1) - \lambda^{-1}K(n)u^H(n)P(n-1) && (5.37)
\end{aligned}
$$

where K represents the filter gain, Y is the correlation matrix, λ is the forgetting factor, u represents the input and is given by

$$u(n) = [\ \bar{x}_i(n-4) \quad \bar{x}_i(n-3) \quad \bar{x}_i(n-2) \quad \bar{x}_i(n-1)] \tag{5.38}$$

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

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

$$\bar{x}_i(n) = w\bar{x}_i(n-1) \tag{5.39}$$

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.*

5.6 Computational Evaluation

5.6.1 Static Traffic (Full buffer)

In the beginning it was assumed that all users have full buffer traffic. The scheduling rule that was used here was the proportionally fair scheduler (2.2). Table 5.1 shows the parameters that were used in the simulation program. The simulation was considered for WCDMA HSDPA where adaptive modulation and coding (AMC)

Table 5.1: System parameters

Parameter	Value
TTI duration	2 ms
Fading model	One path Rayleigh
Minimum rate allowed	256 kbps
	384 kbps
Max. number of associated	15 for QoS 256 kbps
DPCH (N)	10 for QoS 384 kbps
Radio propagation	Site to site distance 500 m
Hybrid ARQ	Chase Combining
Back-off probability	0.999
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

is used to guarantee high throughputs depending on the channel condition. The scheme itself can be used with any opportunistic single or multi-channel scheduler. An HSDPA type of system was chosen due to simplicity of the simulation model. In addition, using a single channel model makes it easier to illustrate the impact of channel impairments.

One new user was added every time using the iterative CAC procedure described in Section 5.2. The Kalman filter described in Section 5.4 was implemented to obtain the mean values since we are averaging over small window sizes. The initial values for $X_i^{(t)}$were the time average rates of the active users before a new user entered the system. A Raleigh fading vector was generated for each user in accordance with Jakes' fading simulator [10, 103]. Discrete SINR-rate mapping was carried out for the AMC. Different users experience different channel conditions that vary depending on their distance from the base-station and velocity.

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 stabilized as the number of time-slots increased.

Fig. 5.4 shows the Kalman estimate along with the mean (expected) rate for the worst user when a new user is admitted. The figure shows how the Kalman estimate converges to the mean value. In this figure it can be seen that the mean rate of the worst user has declined beyond the minimum acceptable rate. With the help of the Kalman filter, the scheme is implemented and the controller will make

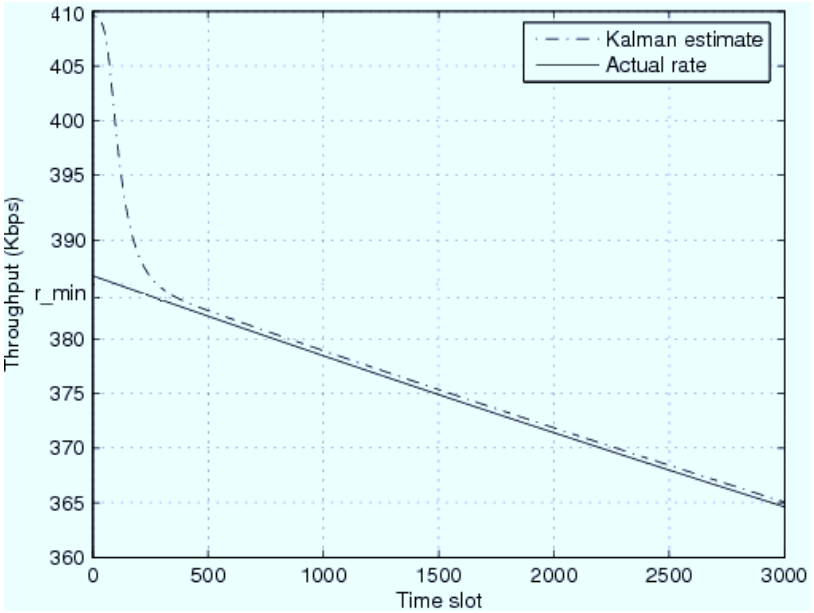

Figure 5.4: Rate of worst user (Back-off factor=0.999)

the decision to either accept or reject the new user where in this case the new user was rejected. It was found after running rigorous simulations that time-slot 2100 ($t = 4.2s$) appeared to be the most suitable decision time (with respect to having the minimum amount of admission errors) to either accepting the new user or rejecting it. A new user is fully accepted if none of the active users experienced a fall in rate below μ_{min} before time-slot 2100. 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.

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 μ_{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 before time-slot 2100 and error type I if after.

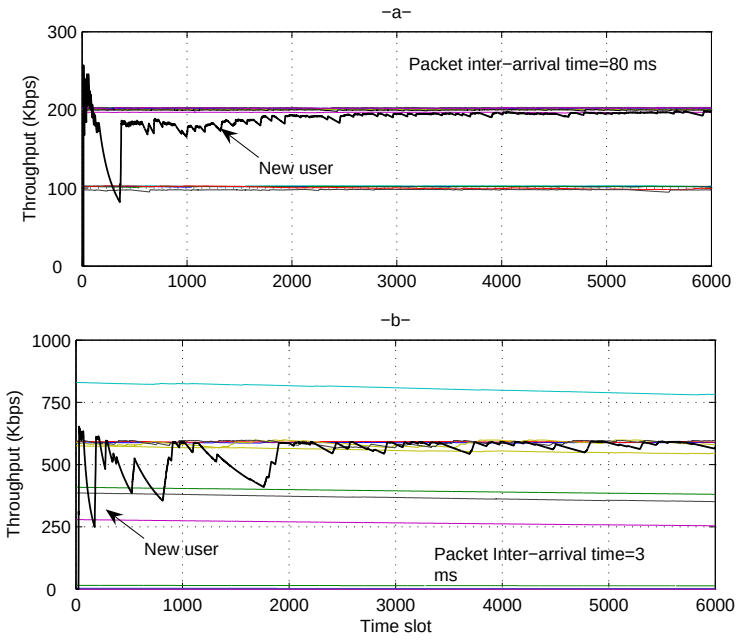

Figure 5.5: Dynamic traffic users with different packet inter-arrival times

5.6.2 Dynamic Traffic

The case in which the buffer occupancy of the users is allowed to vary so that not all users necessarily have data to receive in each time-slot is considered. 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. 5.5 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 can be regarded as similar to a quantizing impact.

We are interested in observing the new user's impact on the active users' channels. Therefore, the mean inter-arrival time was decreased to create more packets and consequently make the rates 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. 5.5. In Fig. 5.5-a it can be seen 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. 5.5-b the impact of the new user is more noticeable where the decrease in the mean rate of the other users is seen 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 modeling the bursty nature of the data traffic more appropriate than the Poission process.

Tables 5.2 and 5.3 illustrate the possibilities of a user being erroneously accepted (Type I CAC error) or erroneously rejected (Type II CAC error) at decision times 4.2s and 2.1s respectively. Both static and dynamic cases are considered and two minimum QoS levels. The results shown in the table were obtained by running multiple simulations 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 5.4. In this table different values to Ψ_{ee} are assigned and the impact on the percentage of errors is observed. It is clear that the higher the estimation error the more errors obtained.

Table 5.2: Type I and II CAC errors (Decision time=4.2s)

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 5.3: Type I and II CAC errors (Decision time=2.1s)

Traffic type	QoS level	Error type I	Error type II
Static	256 kbps	7.02%	6.1%
	384 kbps	10.3%	4.51%
Dynamic	256 kbps	12.8%	3.6%
	384 kbps	9.13%	3.4%

Table 5.4: 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%

5.6.3 Decision Error

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. This section discusses this problem and suggests a solution with the help of an illustrative example.

Example
A new user is admitted and later has a temporary fall in rate but then recovers. This temporary fall results in declaring a type I error. In a second case, the user is denied admission and 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 is in fact accurate, the temporary rise in rate is the main reason to trigger the error flag.

A proposition to solve this problem is to use a simple heuristic scheme in which we observe the number of times the rate crosses a reference (in this case the rate threshold) and the time between each crossing and by computing the ratio, a conclusion can be made about the decision validity. This will enable us to form a good picture about the variations in throughputs and consequently back up the admission decision.

5.6.4 Multi-user Admission Control

In this part we can see the impact of admitting multiple users with independent back-off probabilities on the rate of the worst shown in Fig. 5.6. An alternative would be that all users use the same back-off probability and jointly back off. In

Table 5.5: 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%

this case, the mean rate obtained in frame n becomes

$$\tilde{x}_i^{(n+1)} = p_b \tilde{x}_i^{(n)} + (1 - p_b)\bar{x}_i(Z_{N+M}) \tag{5.40}$$

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 5.5 represents the results obtained for joint back-off admission. The admission is 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.

5.6.5 Non-iterative Admission Control

Fig. 5.7 shows the RLS algorithm application to the worst user. It can be seen that when an 11th user is added, the rate of the worst active user drops below the minimum acceptable rate. Table 5.6 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 5.2 and the results for the RLS were obtained by running repetitive 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.

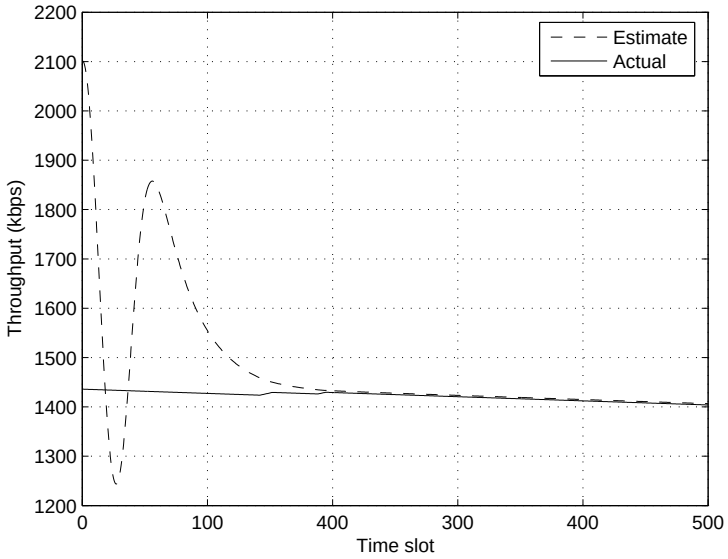

Figure 5.6: Multiuser case - independent probability factors

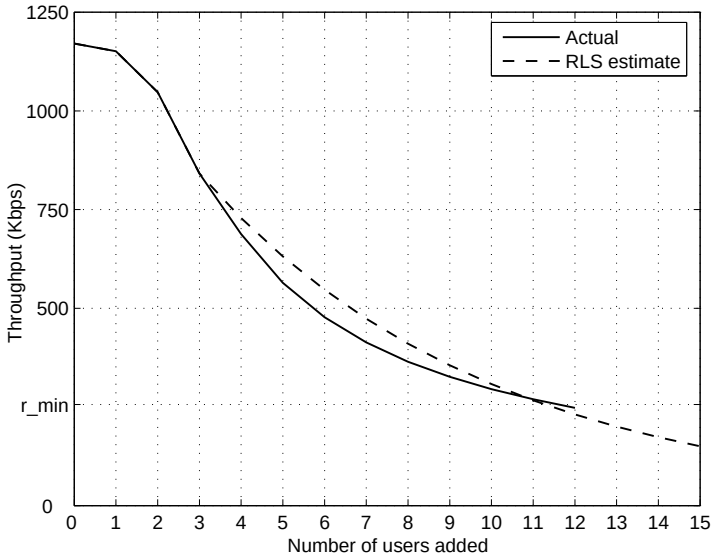

Figure 5.7: RLS admission scheme (Worst user)

Table 5.6: 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%

5.7 Quality Control

Assume that user i requests a 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 providing this way a best-effort framework. For this reason, the algorithm can be combined with, for example, the PF scheduler (2.2). That is; in the congested case, the quality control scheme will 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:

$$q_i^{(n+1)} = \min\left\{1, \max\left\{0, q_i^{(n)} + \beta_i\left(\bar{s}_i^{req} - \bar{s}_i^{(n)}\right)\right\}\right\} \tag{5.41}$$

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

It can be noted 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.

5.7.1 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:

$$\mathcal{T}_i(Q, B) = q_i + \beta_i(\bar{s}_i^{req} - \bar{s}_i(q_i)) \tag{5.42}$$

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

Proposition 5.1 *Suppose that $\mathcal{Q}$ is convex. If $\bar{s} : \Re^n \mapsto \Re^n$ is continuously differentiable and there exists a scalar $\psi \in [0, 1)$ such that*

$$\|I - \beta_i(\nabla_i \bar{s}_i(Q))'\|_\infty + \Sigma_{j \neq i}\|\beta_i(\nabla_j \bar{s}_i(Q))'\|_\infty \leq \psi, \quad \forall i \tag{5.43}$$

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

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 $\rho > 0$ such that

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

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

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

$$\begin{aligned} &|1 - \beta_i \nabla_i \bar{s}_{q_i}| + \beta_i \Sigma_{j \neq i}|\nabla_j \bar{s}_i(q_i)| \\ &= 1 - \beta_i(\nabla_i \bar{s}_{q_i} - \Sigma_{j \neq i}|\nabla_j \bar{s}_i(q_i)|) \\ &\leq 1 - \beta_i \rho < 1 \end{aligned} \tag{5.44}$$

which shows that inequality (5.43) holds. The result follows from Prop. (1). ■

5.7.2 Example

Rate Control

In this case, the requested QoS is depicted in mean service rate. It is useful to request specific rates when 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 $\mathcal{A}(t) \subset \mathcal{N}$. 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

$$\pi_k(Q) = \prod_{j \in \mathcal{A}_k} q_j \prod_{z \in \mathcal{N} \setminus \mathcal{A}_k} (1 - q_z) \tag{5.45}$$

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

$$\bar{x}_i = \sum_{k \in \mathcal{K}_i} \bar{x}_i(\mathcal{A}_k)\pi_k(Q) \tag{5.46}$$

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 $A_k \geq A_l$, i.e. 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

$$\pi_{k,i} = \prod_{\substack{j \in \mathcal{A}_k \\ j \neq i}} q_j \prod_{z \in \mathcal{N} \setminus \mathcal{A}_k} (1 - q_z) \quad k \in \mathcal{K}_i \tag{5.47}$$

which is independent of q_i.
The mapping in eq. (5.42) will now become:

$$\mathcal{T}_i(Q; B) = q_i + \beta_i \left(\bar{x}_i^{req} - q_i \sum_{k \in \mathcal{K}_i} \bar{x}_i(\mathcal{A}_k) \pi_{k,i} \right) \tag{5.48}$$

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

$$q_i^{(m+1)} = \min \left\{ 1, \max \left\{ 0, \mathcal{T}_i(Q^{(m)}, B^{(m)}) \right\} \right\} \tag{5.49}$$

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 5.2 *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* $Q \in \mathcal{Q}$.

Proposition 5.2 *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.*

Proposition 5.3 *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.*

It is worth noting that there is a difference between controlling the activity set $\mathcal{A}$ and controlling channel access ϕ. An example of a scheduler that controls channel access is the CDF based scheduler (CS) which selects the user for transmission based on the cdf of user rates, in such a way that the user whose rate is high enough but least probable to become higher is selected first [104]. This makes the capacity region for this type of schedulers larger than the RCA which limits the rate space for users between three rate vectors. This can be seen in Fig. 5.8 which depicts the normalized achievable mean rates for the RCA and CS algorithms in a simple two-user example [105].

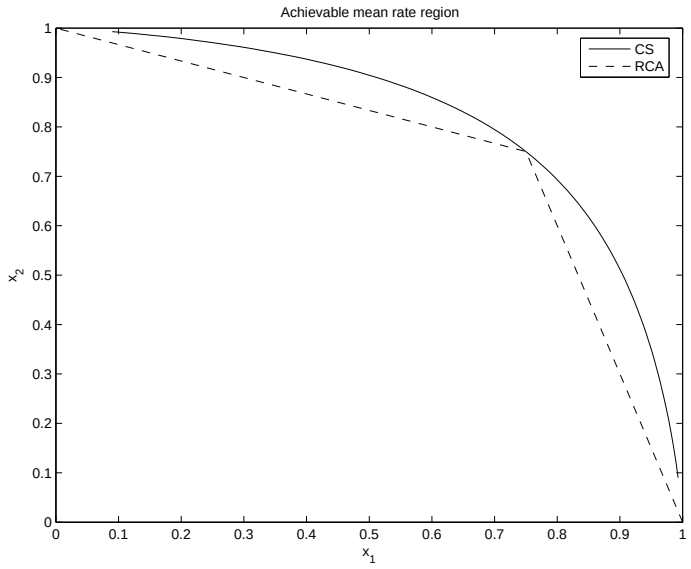

Figure 5.8: Achievable throughput for RCA and CS scheduling schemes

5.8 Imperfect Estimates

Since the mean value $\tilde{s}_i(t)$ is not available we have to use instead the average $\bar{s}_i(t)$ over some time window. $\bar{s}_i(t)$ can be treated as a noisy estimate for $\tilde{s}_i(t)$. As discussed earlier in Section 5.4, in order to cope with imperfect estimates the Kalman state estimation is utilized. 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 a relatively small value of N, the number of states grows to be far too large for real-time operation.

Let $\bar{s}_{n,i} = \bar{s}_i(\mathcal{A}_k)$ for all $A_k = m$, $k \in \mathcal{K}_i$. Furthermore, define

$$\bar{\pi}_{n,i} = \sum_{k \in \{k \in \mathcal{K}_i : A_k = m\}} \pi_{k,i} \tag{5.50}$$

The mean quality parameter equation can be defined as

$$\bar{s}_i = \sum_{m=1}^{N} \bar{\pi}_{m,i} \bar{s}_{m,i} \tag{5.51}$$

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

Table 5.7: System parameters for downlink scheduling in HSDPA

Parameter	Value
Spreading factor	16
Number of multicodes	10
TTI duration	2 ms
Fading model	One path Rayleigh (Jakes' model)
No. of associated DPCH	3
Radio propagation	Site to site distance 500 m
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

Let $S_i^{(n)} = (\bar{s}_{1,i}^{(n)}, \bar{s}_{2,i}^{(n)} \cdots \bar{s}_{N,i}^{(n)})'$ 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 neither sparse nor diagonal. In addition, we have the measurement noise $e_i^{(n)}$ that has a covariance of $\Psi_{ee} = \mathrm{E}\{(e_i^{(n)})^2\} = \upsilon > 0$. If the channel is varying rapidly, this may have a negative impact on the estimation accuracy as well. To model this, it is assumed that the cross covariance $\Psi_{Ve} = \mathrm{E}\{V_i^{(n)} e_i^{(n)}\} \geq 0$. The process dynamics can be expressed as

$$S_i^{(n+1)} = S_i^{(n)} + V_i^{(n)} \tag{5.52}$$

$$\bar{s}_i^{(n)} = C\hat{S}_i^{(n)} + e_i^{(n)} \tag{5.53}$$

The Kalman state estimator can be written as

$$\hat{S}_i^{(n+1)} = \hat{S}_i^{(n)} + (K)_i^{(n)} \left(\bar{s}_i^{(n)} - C_i^{(n)} \bar{X}_i^{(n)}\right) \tag{5.54}$$

$$(K)_i^{(n)} = \left(P_i^{(n)}(C_i^{(n)})' + \Psi_{Ve}\right) ((C_i^{(n)})' P_i^{(n)} C_i^{(n)} + \Psi_{ee})^{-1} \tag{5.55}$$

$$P_i^{(n+1)} = P_i^{(n)} + \Psi_{VV} - (K)_i^{(n)}(C_i^{(n)} P^{(n)}(C_i^{(n)})' + \Psi_{ee})((K)_i^{(n)})' \tag{5.56}$$

5.9 Computational Evaluation

The parameters considered in this part are listed in Table 5.7. Full buffer traffic was assumed. The initial values of 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. 5.9 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 5.8. 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.

Fig. 5.10 illustrates the activity probabilities that achieve the target values. It can be seen from the figures how the probabilities converge to fixed points confirming the convergence analysis outlined in Section 5.7.1 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 requiring more activity. In data rates, all users will have high activity probabilities at the beginning to enable them to reach the target QoS and once they are reached the activities decrease.

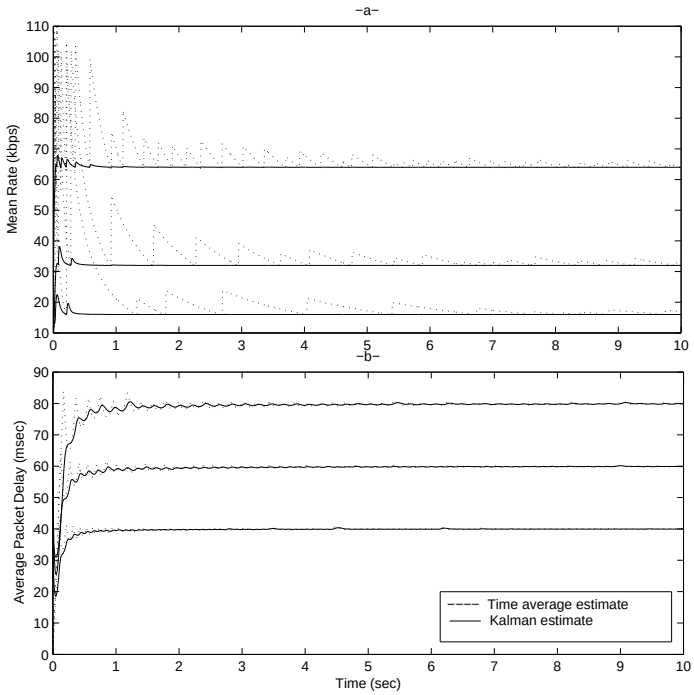

Figure 5.9: a) Time average based estimate & Kalman filter based estimate rates b) Time average based estimate & Kalman filter based estimate delays

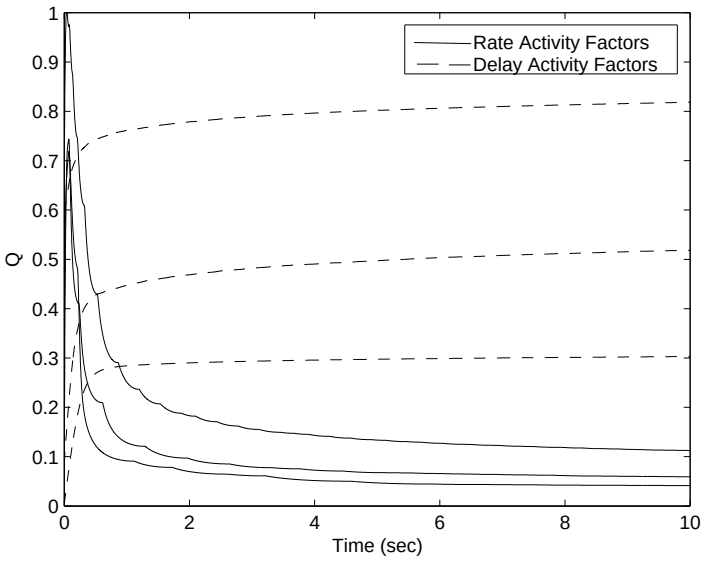

Figure 5.10: Activity probabilities for mean rates & delays

5.10 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 for them. The scheme does not require ideal conditions to be implemented and 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 from the problem of slow convergence. In order to speed up the convergence, a method where the new users jointly back-off was proposed. This scheme can be too cautious because of the issue of fairness by denying admission to users who do not cause the active users' rates to drop below threshold. Finally, the iterative method was compared with a one-shot admission control scheme. The results suggest that the scheme can achieve smaller admission error probabilities than the RLS based one-shot scheme. The main drawback of the scheme lies in the decision time for admission as the scheme requires some time to converge. In the simulations, 4.2 s was used for the time to make the decision which in radio communications can be considered too long, therefore, the scheme is suitable with non real-time traffic only and cases where the channel variation is slow. Admission errors were found to increase when a shorter decision time was used.

An opportunistic scheduler was also proposed that would provide users with requested target QoS levels while maintaining at the same time fairness between them. 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.

Part III
Feedback in Multi-Carrier Systems

Chapter 6

Feedback in Multi-Carrier Systems

Channel state information is an important part of the adaptation process in OFDMA 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 ARQ process. If 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 RBs with one feedback per RB. The RBs can further be grouped into one or several blocks with a joint CSI feedback per block. The RBs still have their own ARQ process to cope with the possible mismatch between the joint CSI and the actual RBs state. However, this raised the question of what is the best basis on which the joint CSI should be determined since there are RBs with different channel states in a block. In this chapter the impact of different decision variables for the feedback information is studied and proposes the use of rank ordering to find the RB state that maximizes the total throughput.

The chapter also provides an analysis that shows the overall system performance for perfect and imperfect channel estimation from multiple resources in a multi-carrier system. The analysis derives the probability of the correct scheduling decision by the base-station when it receives perfect channel information for one resource and imperfect information for the other.

6.1 System Model

In the general multi-carrier model described in Section 1.3 subcarriers are grouped into $L = \frac{B_W}{\Delta f_c}$ RBs having $\frac{\Omega}{L}$ subcarriers each. The base-station allocates power equally among the different subcarriers and RBs. 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. Assume that the channel is subject to Rayleigh fading and 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 a perfect CSI case, the rate assigned to RB 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 RBs 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 RB 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 is utilized, the receiver coherently combines the original code block and the retransmitted block [106]. Assume that the transport format was selected to match rate $\mu = \frac{B_W}{L}\log_2(1+\eta Z)$ 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

$$\begin{aligned}\Pr\{D_i \geq d\} = \quad \Pr \quad & \left\{ \log_2 \left(1 + \eta \sum_{t=t_0}^{t_0+d} |h_i(t)|^2 \right) \right. \\ & \left. \leq \log_2(1+\eta Z) \right\} \end{aligned} \tag{6.1}$$

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

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

$$f_\xi(x) = e^{-x} \tag{6.2}$$

and cumulative probability density function

$$\mathbb{F}_{\xi}(x) = \Pr\{\xi_i(t) < x\} = 1 - e^{-x} \tag{6.3}$$

Now (6.1) can be rewritten as

$$\Pr\{D_i \leq d\} = \Pr\left\{\sum_{t=t_0}^{t_0+d} \xi_i(t) \geq Z\right\}, \tag{6.4}$$

If the coherence time of the channel is small, then $\{\xi_i(t),\, t = t_0, t_0+1, t_0+2, \cdots\}$ become independent and identically distributed (i.i.d) random variables and the sum $\sum_{t=t_0}^{t_0+d} \xi_i(t)$ becomes Erlang-$(d+1)$ distributed. It follows that the delay (in terms of number of retransmissions) becomes Poisson distributed and conditioned on the CSI value Z.

$$\Pr\{D_i \leq d\} = \sum_{t=0}^{d} \frac{(Z)^t}{t!} e^{-Z} \tag{6.5}$$

Therefore, we get

$$\Pr\{D_i = d\} = \frac{(Z)^d}{d!} e^{-Z}, \quad d = 1, 2, 3, \cdots \tag{6.6}$$

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

$$\begin{aligned} \mathrm{E}\{x_i|Z\} &= \sum_{d=0}^{\infty} \frac{\mu(Z)}{d+1} \frac{Z^d}{d!} e^{-Z} && (6.7) \\ &= \frac{\mu(Z)}{Z}(1 - e^{-Z}) && (6.8) \end{aligned}$$

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 ($\eta \to 0$), we have $\mu(Z) = \frac{B_W}{L}(\log_2(1+\eta Z) \approx \frac{B_W}{L}\eta Z$. Therefore, we have $\mathrm{E}\{x_i|Z\} \propto (1 - e^{-Z})$ 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, Therefore,

$$\Pr\{D_i \leq d\} = \Pr\left\{\frac{Z}{\xi_i} < d\right\} \tag{6.9}$$

$$= \Pr\left\{\xi_i > \frac{Z}{d}\right\} = 1 - \mathbb{F}_\xi\left(\frac{Z}{d}\right) \tag{6.10}$$

This distribution $\mathbb{F}_\xi\left(\frac{Z}{d}\right)$ 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

$$f_D(y) = \frac{1}{y^2} e^{-\frac{Z}{y}} \tag{6.11}$$

The throughput conditioned on Z becomes

$$\mathrm{E}\{x_i|Z\} = \int_0^\infty \frac{\mu(Z)}{y+1}\frac{1}{y^2} e^{-\frac{Z}{y}} dy \tag{6.12}$$

$$= \frac{\mu(Z)}{Z}\left(1 - Z e^Z \mathcal{E}_1(Z)\right) \tag{6.13}$$

where $\mathcal{E}_1(x) = \int_1^\infty t^{-1} e^{-xt} dt$ denotes the exponential integral. For large values of Z, $\mathcal{E}_1(x) \sim \frac{e^{-Z}}{Z}$ and we get $\mathrm{E}\{x_i|Z\} \to 0$. 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 $\mu(Z)$ 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_f for a number of RBs L is small. If $L \leq b_f$, we still can use individual feedback for each RB with $\frac{b_f}{L}$ bits per RB. However, if $L > b_f$, this is not possible, since there is less than 1 feedback bit per RB. If b_f is large, say 8 to 16, then joint feedback information can be approximated with a real number.

One way to compose the joint feedback information for a block of RBs is to use the average value $Z = \frac{1}{L}\sum_{j=1} \xi_j$. This variable is correlated with ξ_i, but for large L the throughput can be approximated 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 normally distributed random variable with zero mean and unit variance), so $Z \to 1$. Hence, we have $x_i \sim \mu(1)(1 - e^{-1}) \approx 0.6321\mu(1)$ in the

fast fading case and $x_i \sim \mu(1)\,(1 - e^1 \mathcal{E}_1(1)) \approx 0.4037\mu(1)$ 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 $Z_{L,n}$ denote the n^{th} smallest RB state value in a block containing L RB $\{\xi_i(t_0), i = 1, 2, \cdots, L\}$; where $\xi_i(t_0)$ denotes the state value of RB i at time t_0. If $Z_{L,n}$ is utilized to select the utilized data rate, then $n-1$ RBs have to use retransmission while the rest can transmit the packet directly. The probability density function of $Z_{L,n}$ in the case of exponentially distributed random variables is given by

$$f_{Z_{L,n}}(x) = \frac{L!}{(L-n)!} \sum_{k=0}^{n-1} \frac{1}{(n-1-t)!t!} e^{-(k+n)x} \tag{6.14}$$

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

$$\begin{aligned} \mathbb{E}\{x_i\} &= \int_0^\infty \frac{\mu(Z)}{Z} \mathbb{F}_\xi(Z) f_{Z_{N,n}}(x) dx & (6.15) \\ &= \int_0^\infty \frac{\mu(Z)}{Z} (1 - (1 - \mathbb{F}_\xi(Z))) f_{Z_{L,n}}(x) dx \\ &= \int_0^\infty \frac{\mu(Z)}{Z} (f_{Z_{L,n}}(x) - \frac{L-n+1}{L+1} f_{Z_{L+1,n}}(x)) dx \\ &= \int_0^\infty \frac{\mu(Z)}{Z} (f_{L_{L,n}}(x) - (1 - \frac{n}{L+1}) f_{Z_{L+1,n}}(x)) dx & (6.16) \end{aligned}$$

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

$$\mathbb{E}\{x_i\} \propto \frac{n}{L+1} \tag{6.17}$$

Hence, throughput is maximized with $n = L$ which corresponds to the SINR of the best RB and so, $Z = \max_i\{\xi_i\}$ maximizes the throughput. In a high SINR domain, the term $\frac{\mu(\gamma)}{Z}$ cannot be ignored and (6.15) must be solved numerically with the help of a computer simulator as illustrated in Section 6.3.

Table 6.1: System parameters for a multi-carrier system

Parameter	Value
Total Bandwidth	100 MHz
Coherence Bandwidth	333 kHz
Spreading factor	16
Number of multicodes	10
TTI duration	2 ms
Fading model	One path Rayleigh (Jakes' 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
RB spacing	320 kHz
Subcarriers/Subband	32

6.3 Numerical Analysis

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

6.3.1 Decision Variable Based on Rank Ordering

Considering a single user case, the bandwidth allocated to the user is B_W which is divided using the inverse fast Fourier transform into multiple orthogonal subcarriers with equal spacing. The subcarriers are grouped into RBs 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) which resembles WCDMA-HSDPA but with OFDM in the radio interface [107]. Adaptive Modulation and Coding is used in a transmission time interval based on the CSI report. The TTI is short enough that the channel can be assumed to be 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 chunk 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 RBs) for different decision

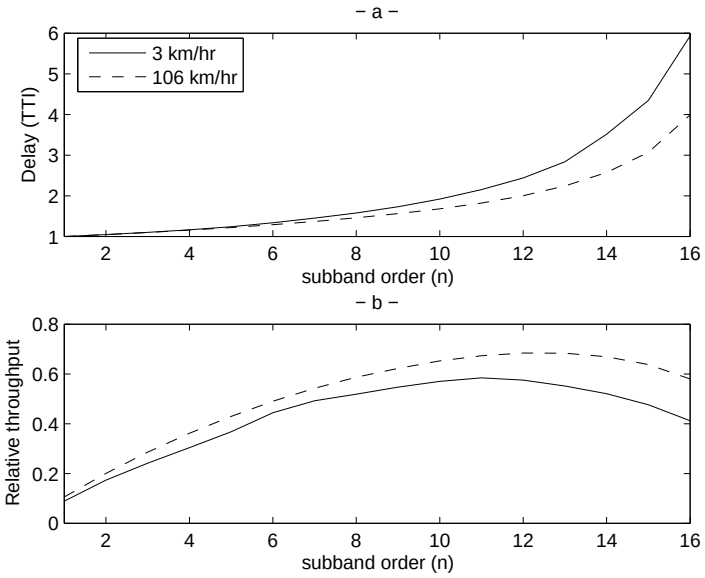

Figure 6.1: a) Delay and b) relative throughput as functions of the decision variable for different speeds (mean SINR=3 dB), (subband=RB)

variables. The variable n represents the rank order of a RB, i.e. $n = 1$ is the RB with the lowest channel state and $n = 16$ represents the highest. It is noted 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 RBs for the joint CSI. This is illustrated in Fig. 6.1-b where it can 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 chunk of blocks 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 RBs as well as the chunk average. The throughput obtained with the optimal RB (the RB 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. In a high mobility case, Fig. 6.3 shows that it is more convenient to depend on the maximum RB 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 Chunks

This section will study the effect of having multiple CSI feedback channels. the RBs are grouped into multiple chunks instead of only one chunk of blocks considered earlier. We are interested in seeing the impact of using multiple blocks each having it's own ARQ process on the total throughput. Each chunk is assumed to have the same number of feedback bits for its feedback channel. The number of quantization levels for the AMC obtained with this number is assumed to be large (e.g. 16 bits $\rightarrow$ 65536 levels).

Figure 6.4 shows that aggregating the throughput of multiple chunks of smaller sizes results in a higher relative throughput. This is mainly due to the fact that increasing the number of chunks increases the diversity gain, (subband=RB).

6.3.4 Effect of Number of Feedback Bits

This section studies the effect of the number of feedback bits allocated to the RBs and compares 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 RBs L is b_f. There are two scenarios for utilizing these bits, 1) each RB can, individually, have its own feedback channel with a capacity of b_f/L bits giving $2^{b_f/L}$ quantization levels, 2) group the RBs into K chunks of L/K RBs with one feedback channel per chunk. The number of feedback bits for this channel will be the sum of the feedback bits of the individual RBs within the chunk yielding $b_f/L{*}L/K = b_f/K$ feedback bits and consequently $2^{b_f/K}$ levels ($2^{b_f/K} > 2^{b_f/L}$). Rank ordering is then performed for the chunk to find the RB state that maximizes the total throughput and report that state back to the transmitter using the b_f/K bits. For small bit feedback channels a coarse quantization method is used [108] which finds the optimal quantization levels based on the mean SNR assuming that there is no delay in the

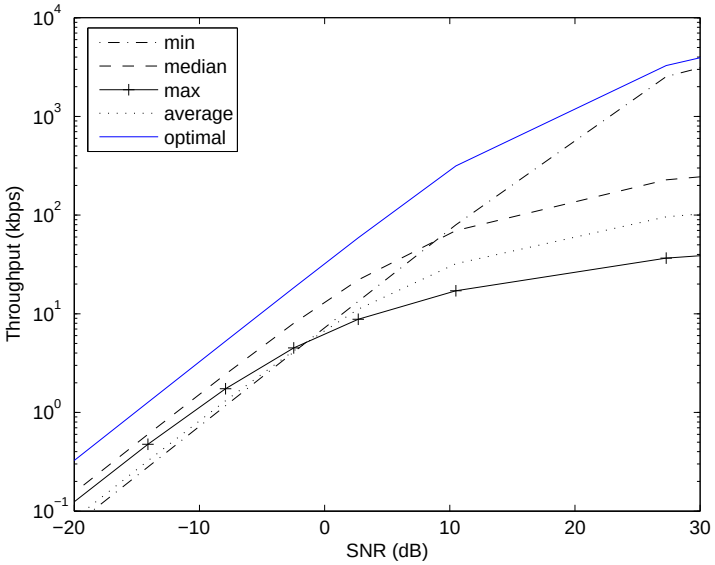

Figure 6.2: Total throughput for different decision variables (speed=3 km/hr)

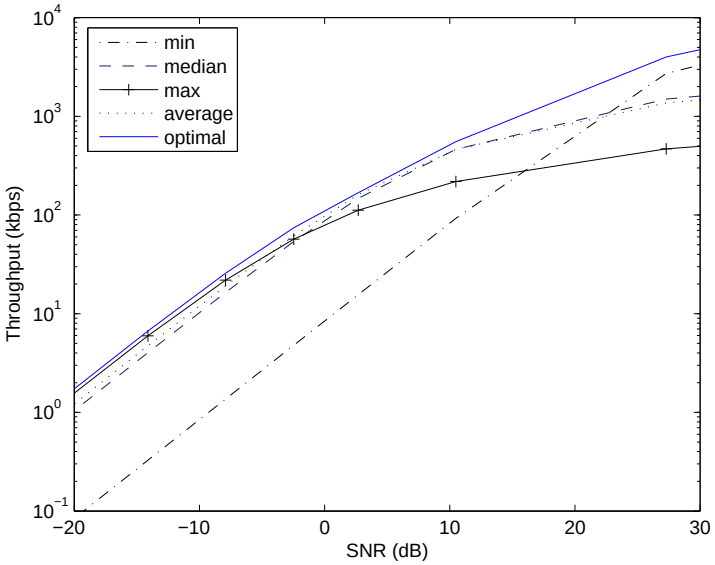

Figure 6.3: Total throughput for different decision variables (speed=106 km/hr)

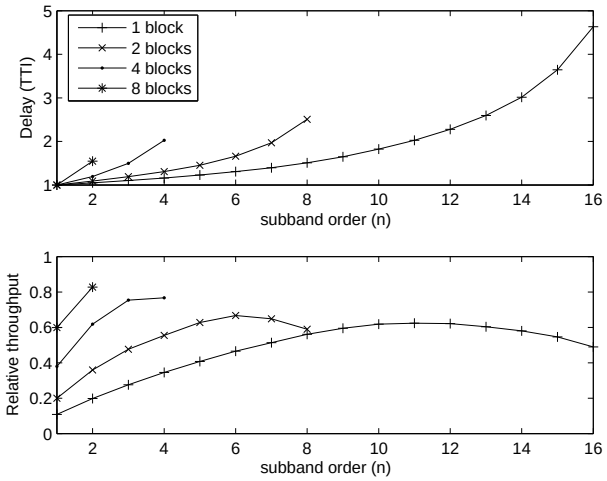

Figure 6.4: Relative throughput for different chunk sizes

feedback. That is, the quantization levels $\{\gamma_k, k = 0, 1, 2, \cdots \rho - 1\}$, where k is the level number and ρ is the total number of quantization levels are determined by solving the following optimization problem

$$\max_{\gamma_k} \sum_{k=0}^{\rho-1} \left(\mathbb{F}_\xi \left(\frac{\gamma_{k+1}}{\gamma} \right) - \mathbb{F}_\xi \left(\frac{\gamma_k}{\gamma} \right) \right) \log_2(1 + \gamma_k) \tag{6.18}$$

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

Two cases are considered. The first is for a large number of bits per RB ($\frac{b_f}{L}$) and the second is with a smaller one. $b_f = 32$ and 8 respectively, $L = 8$ RBs, $K = 1, 2, 4$ and 8 chunks. 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 RBs as shown in Table 6.3.

Table 6.2: Different bit allocation ($b_f = 32, L = 8$)

Number of chunks	Feedback bits per chunk	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

Table 6.3: Different bit allocation ($b_f = 8, L = 8$)

Number of chunks	Feedback bits per chunk	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 Impact of Feedback Information Accuracy

The estimation accuracy of the feedback information is essential for opportunistic schedulers. Erroneous information leads to significant degradation in system performance because of the fact that users are allocated resources that do not match their actual channel conditions. Consequently this increases bit error rates leading to requests for retransmission in the case of non-real time traffic causing increased latency. In the case of real-time traffic which has very limited latency tolerance, the information will simply be lost if forward error correction (FEC) methods cannot retrieve the data. In this section an analysis is made to show the impact of errors in channel estimation on the scheduling decision in a real-time traffic scenario. The baseline assumptions in this section are

- UE is transmitting without disruption using a dedicated RB which occurs after each fixed time period. This RB is called the 'first RB'.
- In addition to the applied RB there is another RB that can be made available to the UE. The UE transmits periodically a sounding signal over this RB. This RB is called the 'second RB'.
- First and second RBs are uncorrelated.
- The channel estimation related to the first RB is perfect while the result of the channel estimation related to the second RB contains an error.

- The BS carries out channel estimation corresponding to both first and second RBs. Based on the result of the estimation a single-bit feedback is sent to the UE over a dedicated control channel. The UE then selects the RB for data transmission based on the received control bit.

Let us briefly go through the validity of the assumptions. In VoIP a lot of small speech packets are transmitted making the fully dynamic scheduling an exhaustive task due to large control overhead. Therefore, it is reasonable to use fixed allocations as much as possible keeping in mind that the proposed scheduling represents a trade-off between fixed and dynamic scheduling. While the location of the second RB in frequency span can be changed after each time period we are able to seek a better RB for VoIP data transmission. In practice RBs may admit some correlation but since VoIP packets are small, a certain minimum distance can be used between first and second RBs with respect to frequency.

The channel estimation related to the first RB is assumed to be perfect because longer pilot sequences are used corresponding to the data packets and, furthermore, it is possible to filter the channel estimation over a few consecutive RBs if changes do not occur too often. In the case of the second RB, only a sounding signal can be used for channel estimation and thus the result contains an error that is in the following analysis assumed to be complex Gaussian.

Finally, since only two RBs are used, the feedback from the BS consists of a single bit that needs to be convoyed to the UE. In fact, it is not difficult to model the impact of feedback errors in the control channel provided that there is a constant feedback bit error probability, say P. This is the case when accurate power control is applied in the feedback channel. The aim is to model the uplink signal distribution when the scheduling mentioned above is applied. The bit-error-probability is also computed.

In the signal model the complex channels H_1 and H_2 resulting from channel estimation are given by

$$\begin{cases} H_1 = h_1, & \mathbb{E}\{|H_1|^2\} = 2\sigma_1, \\ H_2 = h_2 + \varepsilon, & \mathbb{E}\{|H_2|^2\} = 2\sigma_2 + 2\epsilon^2. \end{cases} \tag{6.19}$$

Here h_1 and h_2 are the true channels and ε is the error from channel estimation. All these variables are assumed to be complex zero-mean Gaussian. Parameters σ_1 and σ_2 are the standard deviations of the underlying Gaussian distribution. Since true channels are mutually uncorrelated we have

$$\mathbb{E}\{h_1^* h_2\} = 0, \quad \mathbb{E}\{H_1^* H_2\} = 0. \tag{6.20}$$

It is assumed that normalization $\mathbb{E}\{|H_1|^2\} = \mathbb{E}\{|H_2|^2\}$ has been used.

6.4.1 General Form for SNR Distribution

In general the PDF of the SNR distribution in the uplink BS receiver after the scheduling is of the form

$$f(\eta) = (1 - P) \cdot f_{\max}(\eta) + P \cdot f_{\min}(\eta), \tag{6.21}$$

where P is the probability that a wrong scheduling decision is made in the BS. It is not difficult to see that (6.21) can be written in the form

$$f(\eta) = (1 - 2P) \cdot f_{\min}(\eta) + 2P \cdot f_{\mathrm{S}}(\eta), \tag{6.22}$$

with

$$f_{\max}(\eta) = 2 \cdot f_{\mathrm{S}}(\eta) - f_{\min}(\eta) \tag{6.23}$$

where f_{S} is the SNR distribution when scheduling is not applied or scheduling decisions are fully random. It is noted that in this case

$$f_{\mathrm{S}}(\eta) = \frac{1}{\bar{\eta}} e^{-\eta/\bar{\eta}}, \quad f_{\max}(\eta) = \frac{2}{\bar{\eta}} e^{-\eta/\bar{\eta}} \big(1 - e^{-\eta/\bar{\eta}}\big)$$

According to (6.22) we can write for BEP

$$\mathrm{BEP}(\bar{\eta}) = (1 - 2P) \cdot \mathrm{BEP}_{\min}(\bar{\eta}) + 2P \cdot \mathrm{BEP}_{\mathrm{S}}(\bar{\eta}) \tag{6.24}$$

where $\bar{\eta}$ is the mean SNR. It is known that in the asymptotic region we have

$$\mathrm{BEP}_{\min}(\bar{\eta}) \sim C_{\max}/\bar{\eta}^2, \quad \bar{\eta} >> 1$$
$$\mathrm{BEP}_{\mathrm{S}}(\bar{\eta}) \sim C_{\mathrm{S}}/\bar{\eta}, \quad \bar{\eta} >> 1$$

Hence, we have in the asymptotic region

$$\mathrm{BEP}(\bar{\eta}) \sim 2P \cdot C_{\mathrm{S}}/\bar{\eta} \tag{6.25}$$

which means that the slope of the BEP is the same as for single antenna transmission. Factor $2P$ defines the gain from selection in the asymptotic SNR region.

6.4.2 Probability of a Correct Scheduling Decision

Let us consider the computation of P. This probability is the key to system performance. The following events are first considered:

A = 'SNR of the first estimated channel H_1 is larger than the SNR of the second estimated channel H_2'.
B = 'SNR of the first true channel h_1 is larger than the SNR of the second true channel h_2'.

Then the probability of the correct decision P^C can be written in the form

$$P^C = \mathrm{Q}(A, B) + \mathrm{Q}(A^c, B^c) \tag{6.26}$$

We note that joint decisions (A, B), (A^c, B^c), (A^c, B) and (A, B^c) are mutually exclusive. Hence, we need to compute the joint probabilities of (6.26). In computations it is denoted that $r_m = |h_m|$ and $R_m = |H_m|$. Consequently, we have

$$R_1 = r_1, \quad R_2 = |r_2 + \epsilon \cdot e^{j\phi}| \tag{6.27}$$

where ϕ is the argument between the true channel h_2 and error ε. Using these notations we find that

$$\mathrm{Q}(A, B) = \mathrm{Q}(R_1 > R_2, r_1/\sigma_1 > r_2/\sigma_2) \tag{6.28}$$

where r_1 and r_2 are scaled to have the same mean power. Without this scaling, the scheduling decision would favor either one of the RBs depending on the power allocation. The mean power of R_1 and R_2 are the same by definition. The corresponding integral formulation for joint probability of (6.28) is of the form

$$\mathrm{Q}(A, B) = \int_0^{\infty} \int_0^{r_1} \int_0^{\sigma_1 r_1/\sigma_2} g(r_1, R_2, r_2) dr_2 dR_2 dr_1 \tag{6.29}$$

where g is the joint distribution of r_1, R_2 and r_2. Similarly we obtain

$$\begin{aligned} \mathrm{Q}(A^c, B^c) &= \mathrm{Q}(R_1 < R_2, r_1/\sigma_1 < r_2/\sigma_2) \\ &= \int_0^\infty \int_{r_1}^\infty \int_{\sigma_1 r_1/\sigma_2}^\infty g(r_1, R_2, r_2) dr_2 dR_2 dr_1 \end{aligned} \tag{6.30}$$

Let us compute (6.30) first. The result of (6.30) can be then deduced easily based on analogous computational procedures. In (6.30) variables r_1 and R_2 are uncorrelated but R_2 depends on r_2. Furthermore, there holds that

$$g(r_1, R_2, r_2) = g(r_1, R_2|r_2)g(r_2) = g(R_2|r_2)g(r_1)g(r_2) \tag{6.31}$$

where distributions of r_1 and r_2 are Rayleigh,

$$g(r_m) = \frac{r_m}{\sigma_m^2} e^{-r_m^2/2\sigma_m^2} \tag{6.32}$$

but R_2, conditioned by r_2, follows the Rice distribution,

$$g(R_2|r_2) = \frac{R_2}{\epsilon^2} e^{-(R_2^2+r_2^2)/2\epsilon^2} I_0\Big(\frac{R_2 r_2}{\epsilon^2}\Big) \tag{6.33}$$

Here I_0 is the modified Bessel function of order zero. After combining (6.31)- (6.33) we obtain

$$\mathrm{Q}(A^c, B^c) = \int_0^\infty \frac{r_1}{\sigma_1^2} e^{-r_1^2/2\sigma_1^2} \int_{r_1}^\infty \frac{R_2}{\epsilon^2} e^{-R_2^2/2\epsilon^2} \mathcal{I}(r_1, R_2) dR_2 dr_1 \tag{6.34}$$

where we have used the notation

$$\mathcal{I}(r_1, R_2) = \int_{\sigma_1 r_1/\sigma_2}^\infty \frac{r_2}{\sigma_2^2} e^{-r_2^2\left(\frac{1}{2\sigma_2^2}+\frac{1}{2\epsilon^2}\right)} I_0\Big(\frac{R_2 r_2}{\epsilon^2}\Big) dr_2 \tag{6.35}$$

The next task is to compute the integral (6.35). For that purpose the following expression are used

$$I_0(z) = \sum_{k=0}^\infty \Big(\frac{1}{k!}\Big)^2 \Big(\frac{z}{2}\Big)^{2k} \tag{6.36}$$

After some elementary manipulations we find that

$$\mathrm{Q}(A^c, B^c) = \sum_{k=0}^\infty \Big(\frac{1}{k!}\Big)^2 (1-\nu^2)^k \nu^2 \int_0^\infty e^{-t} \Gamma(k+1, t/\nu^2)^2 dt \tag{6.37}$$

where $\Gamma(\cdot,\cdot)$ refers to the complementary incomplete gamma function and $\nu^2 = \epsilon^2/\sigma_1^2 = \epsilon^2/(\sigma_2^2+\epsilon^2)$ is the power of the estimation error divided by the total mean received power of an RB. Similarly we obtain

$$\mathrm{Q}(A,B) = \sum_{k=0}^{\infty}\left(\frac{1}{k!}\right)^2(1-\nu^2)^k\nu^2\int_0^{\infty} e^{-t}\gamma(k+1,t/\nu^2)^2dt \tag{6.38}$$

where $\gamma(\cdot,\cdot)$ refers to the incomplete gamma function.

After summing up results of (6.37) and (6.38) and applying some elementary manipulations we find that

$$P^C = 2 - \sum_{k=0}^{\infty}\sum_{m=0}^{k}\frac{(m+k)!}{m!k!}\frac{4\nu^2(1-\nu^2)}{(2+\nu^2)^{m+k+1}} \tag{6.39}$$

Fig. 6.5 demonstrates the theory and simulation BEP curves for a user in a one user case. In a high SNR region, the presence of a slight estimation error introduces a significant loss in gain which grows proportional to the power of the estimation error.

6.4.3 Two-User Case

In this case, two users each utilizing the available two bands are considered. Therefore, in order for a user to switch to the other channel, the second user must also want to switch at the same time. Thus, the following scenarios materialize

- ◇ $[A_1, B_1; A_2, B_2]$: Both users make a correct decision and keep their channels fixed resulting in $\mathrm{BEP}_{\min}$ for both users.
- ◇ $[A_1, B_1; A_2^c, B_2^c]$: User 1 correctly keeps its channel fixed and experiences $\mathrm{BEP}_{\min}$ and user 2 correctly wishes to switch but cannot, resulting in $\mathrm{BEP}_{\max}$.
- ◇ $[A_1^c, B_1^c; A_2, B_2]$: User 1 correctly wishes to switch and user 2 correctly keeps its channel: user 1 experiences $\mathrm{BEP}_{\max}$ and user 2; $\mathrm{BEP}_{\min}$.
- ◇ $[A_1^c, B_1; A_2, B_2]$: User 1 erroneously wishes to switch while user 2 correctly keeps its channel: both users experience $\mathrm{BEP}_{\min}$.
- ◇ $[A_1, B_1; A_2^c, B_2]$: User 2 erroneously wishes to switch while user 1 correctly keeps its channel: both users experience $\mathrm{BEP}_{\min}$.

- $\diamond$ $[A_1^c, B_1; A_2^c, B_2]$: Both users erroneously wish to switch leading to $\text{BEP}_{\max}$ for both.

- $\diamond$ $[A_1, B_1^c; A_2, B_2]$: User 1 erroneously keeps its channel and user 2 correctly keeps its channel: user 1 experiences $\text{BEP}_{\max}$ and user 2; $\text{BEP}_{\min}$.

- $\diamond$ $[A_1, B_1; A_2, B_2^c]$: User 2 erroneously keeps its channel and user 1 correctly keeps its channel: user 1 experiences $\text{BEP}_{\min}$ and user 2; $\text{BEP}_{\max}$.

- $\diamond$ $[A_1, B_1^c; A_2, B_2^c]$: Both users erroneously keep their channels leading to $\text{BEP}_{\max}$ for both.

- $\diamond$ $[A_1^c, B_1^c; A_2^c, B_2^c]$: Both users correctly want to switch leading to $\text{BEP}_{\min}$ for both.

Based on the above scenarios, the probability that a user will have $\text{BEP}_{\min}$ is

$$\begin{aligned}(P_i)^{(min)} &= Q_i(A,B) + Q_i(A^c,B^c)Q_j(A^c,B^c) \\ &+Q_i(A^c,B)[Q_j(A,B) + Q_j(A,B^c)] \quad i \neq j\end{aligned} \tag{6.40}$$

with

$$\begin{aligned}\text{Q}(A,B) &= 1 - \sum_{k=0}^{\infty}\sum_{m=0}^{k} \frac{(m+k)!}{n!k!}\frac{2\nu^2(1-\nu^2)}{(2+\nu^2)^{m+k+1}} \\ &= \text{Q}(A^c,B^c)\end{aligned} \tag{6.41}$$

Since (A,B), (A^c,B^c), (A^c,B) and (A,B^c) are mutually exclusive, then it is fair to say that $(A^c,B) = (A,B^c)$ which was also verified through simulations. We end up with

$$\text{Q}(A,B) + \text{Q}(A^c,B) + \text{Q}(A,B^c) + \text{Q}(A^c,B^c) = 1$$

So;

$$\text{Q}(A^c,B) = \text{Q}(A,B^c) = \frac{1}{2}\cdot[1 - 2\text{Q}(A,B)] \tag{6.42}$$

Now the BEP for any of the users can be found from

$$BEP_i = P_i^{(min)}\text{BEP}_{\min} + (1 - P_i^{(max)})\text{BEP}_{\max} \tag{6.43}$$

Fig. 6.6 demonstrates the theory and simulation BEP curves for a user in a two-user case. The BEP grows significantly due to the fact that both users have to be in accord in order to change their channel. We also notice that the gain in different estimation error powers becomes smaller when compared to Fig. 6.5. In Fig. 6.8 we can see the spectral efficiency with one and two users in the system.

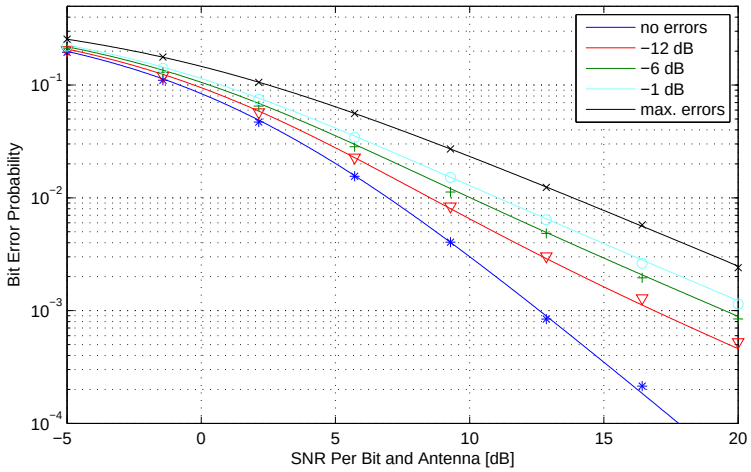

Figure 6.5: Theory and simulation (marked) BEP of a user in a one user case for different values of power ratio ν^2

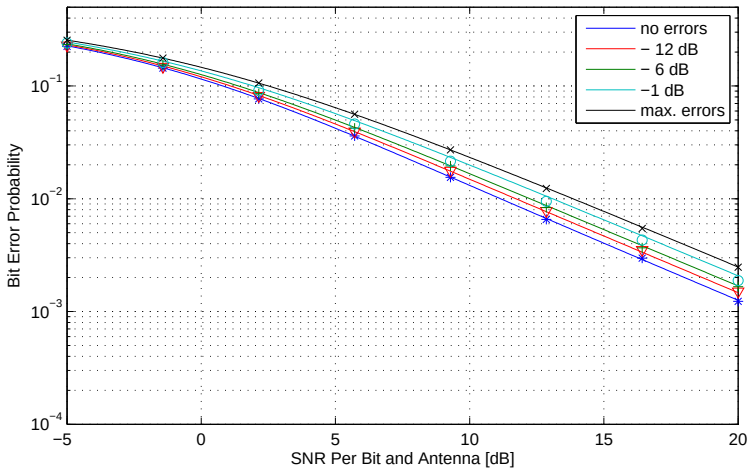

Figure 6.6: Theory and simulation (marked) BEP of a user in a two-user case for different values of power ratio ν^2

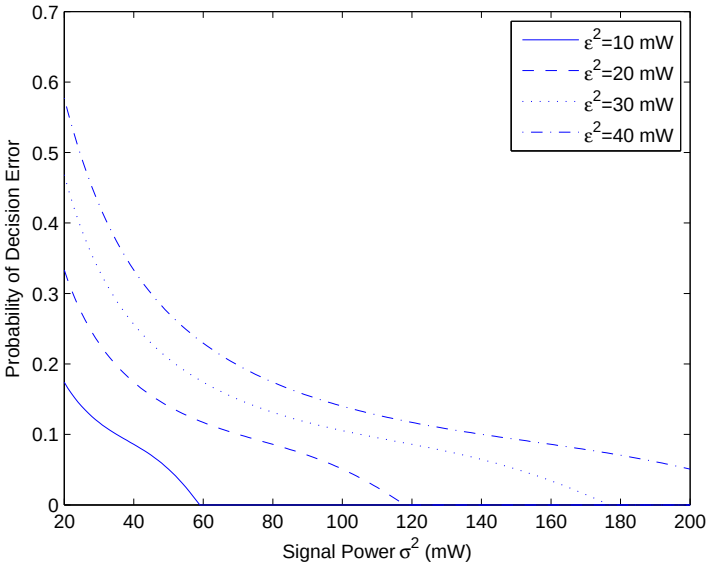

Figure 6.7: Probability of a decision error versus the signal power for fixed values of estimation error

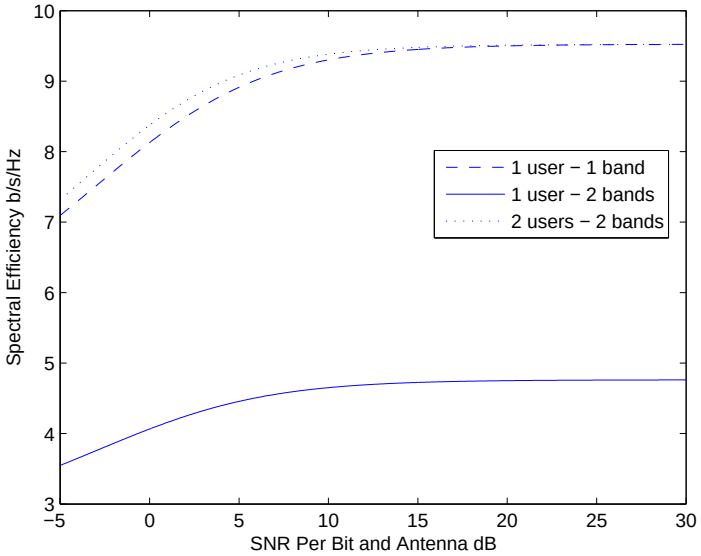

Figure 6.8: System Spectral Efficiency

6.5 Concluding Remarks

Rank ordering is a useful method in 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 the case of linear SINR-rate mapping, using the highest received SINR as a feedback would be optimal. However, in cases where the SINR-rate mapping has a logarithmic shape, some $n < N$ order statistic is 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 also found that dividing the RBs into multiple small sized chunks would further increase the aggregate throughput. Rank ordering is a good choice when the number of feedback bits allocated to a single RB is small. It was found that grouping the RBs and using the sum of their feedback bits for the joint feedback outperforms the case when each RB used an individual feedback channel when the number of total bits was less or equal to the number of RBs.

An analysis was made in this chapter to examine the impact of the presence of error in the estimation of the feedback data on the scheduling decision. System performance was evaluated and theoretical results were verified with practical simulations. The probability of decision error was numerically evaluated for different degrees of estimation error. Therefore, this analysis has also provided a good insight into the impact of having perfect and imperfect channel information available at the scheduler for separate resources. The results can, for example, be utilized in admission control where the controller can compute the BEP of connections based on the estimation error and deny them admission if the BEP exceeded a certain threshold.

Chapter 7

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 heuristic computing methods were utilized in this research. Nevertheless, optimization problems with different degrees of computational complexity were introduced to evaluate the performance of the suggested heuristics. 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.

A number of utility-based scheduling algorithms for multi-carrier systems were proposed in this thesis. Utility-based scheduling is able to provide high system performance while maintaining certain degrees of fairness. A simple single-cell scheduler was first introduced that uses a heuristic in its allocation decision and takes into account the constraints of the access technique. The result of this allocation was compared with the optimal solution and the difference was found to be fairly good. Due to the allocation constraints in certain multi-carrier systems there was the problem of allocating excessive resources to users. For this reason, a procedure was proposed that would solve, or at least minimize the waste in resources. It is worth noting that the work carried out in this part was one of the early studies in channel adaptive scheduling for the uplink of LTE systems. In general, static traffic is assumed in the models. However, it was necessary to show the impact of dynamic buffer traffic assumptions. In dynamic buffer traffic, a report about the buffer occupancy has to be reported to the scheduler which uses this information

in its decision. The impact of the delays in that report was studied as well as the amount of information that should be reported.

Inter-cell interference coordination via coordination of transmission times was proposed in a way that would mitigate interference. The scheduler therefore provides utilization of the whole frequency spectrum for all cells but separates them in time. The separation guarantees that users who interfere with each other do not transmit at the same time. The heuristic solution provided a relatively good result when compared with the optimal solution which was also formulated. Coordination was also conducted in the frequency domain where cell-edge users of neighboring cells would be allocated a different number of resources. For this task Nash bargaining was proposed that allowed cells to enter the bargaining process with a minimum QoS and leave the process if bargaining cannot provide a better service. In an attempt to increase performance, the bargaining process was combined with another coordination scheme where users are handed over to neighboring cells that have lesser loads.

The topic of activity control was discussed in this thesis in terms of an opportunistic admission controller that assigns a back-off probability in its scheduling to new users that would limit their activity and consequently the impact on ongoing calls. This approach is very useful in the sense that a user is gradually admitted into the system instead of making an instantaneous and negative decision that may harm the new user or a positive decision that may harm other users. Different aspects of this task are studied such as: i) different values for the back-off probability, ii) different admission times, iii) dynamic buffer traffic, iv) possibilities of erroneous admission decisions, v) the impact of adding multiple users at the same time. In the same context, the quality constrained scheduling was proposed where the scheduler controls the activity of the users to reach requested QoS levels. The scheduler provides a safety net whereby in case the requested QoS level cannot be realized then the user is guaranteed a fair share of resources.

Since the algorithms deal with multi-carrier systems, it was important to provide a study about the feedback aspect of these systems. For this part, there was a need to show the impact of feedback in a multi-carrier systems with real-time traffic because of the sensitivity of this traffic. Different criteria were examined such as the type of information that should be reported and the amount of that informa-

tion. Reporting the state of every subcarrier naturally causes considerable overhead therefore rank ordering was suggested to find the least and most suitable information to send. It was found that the way to send the information also has an impact on performance. For this purpose the effect of different chunk sizes with different numbers of feedback bits was studied and it was found that having multiple chunks of RBs yielded higher throughput. The quality of the feedback information is also an important issue. Having estimation errors in the feedback reports leads to faulty scheduling decisions which in turn leads to system degradation. An analysis was made in this task and showed the impact of estimation error on system performance.

Discussion

The work generally assumes fixed power allocation. Therefore it would be interesting to see the effect of power control especially in inter-cell interference coordination schemes. The main issue is that the complexity of the algorithms will grow when the power vector is included in the optimization problems.

In transmission inter-cell interference coordination, ideal assumptions were made and therefore, potential future work can be to develop a more practical scheduler by, for example, taking into account the signaling constraints and imperfect channel estimation as well as considering power control and real-time data. Another suggestion for this part would be to include a threshold that classifies the users who need to be coordinated since users are distributed uniformly and suffer different degrees of interference. In bargaining inter-cell interference coordination, users were divided into center and edge-cell users. Edge users were defined according to a threshold difference in the beacon power a user receives from its base-station and the neighboring ones. This threshold consequently defines the cell split between users which in turn has a significant impact on system throughput i.e. the more users defined as edge users the lower the cell throughput, this was observed in a study by Laakso *et al.* [109]. Another observation in their study is the impact of intra-cell scheduling techniques and whether different scheduling algorithms can add additional gain to the interference coordination gain. It was found that for low split cases (enough center users) the scheduler didn't add that much gain. This observation can help lower the complexity of our algorithm by using simpler intra-cell scheduling methods without fear of affecting overall performance. In general, ICIC requires signaling between

base-stations that need to exchange information. Since this exchange takes place periodically in Nash bargaining and is done over the X2 interface then it is possible to assume that such an ICIC method can be implemented in practice especially that ICIC is not standardized but part of an implementation-specific strategy [110]. However, based on the complexity of the algorithm, the core network size plays an important role as bargaining becomes more complicated in larger networks.

In utility based scheduling, the author's approach in scheduling the RBs was among the initial suggestions for LTE uplink opportunistic schedulers. Papers that came afterwards proposed similar approaches as well introducing derivatives that provided increased gains, one example can be found in [111].

In the admission control part, 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, the algorithm is more sensitive to channel condition variations than others. The suggestion of a level crossing counter to verify a drop or rise in channel state may not be a very good choice. Further studies can look at a more efficient way to verify these channel states for the controller.

In the quality controlled scheduler, it was seen that in the case of the rate control algorithm the capacity region was smaller compared to a more generic quality control algorithm such as the CDF-based scheduler (CS). The structure of the RCA is designed such that the algorithm falls back to the proportional fair scheduler in the event the requested QoS is not possible. Therefore, it is possible to combine the RCA with CS instead of PF scheduling to achieve the convex hull of the CS rates.

Bibliography

[1] H. Holma, A. Toskala. WCDMA for UMTS, Radio access for third generation mobile communications. Third edition, Wiley, 2005.

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

[3] J. Bergman, M. Ericson, D. Gerstenberger, B. Gransson, J. Peisa and S. Wager. HSPA evolution - boosting the performance of mobile broadband access. Ericsson Review, vol. 85(2008):1, pp. 32-37.

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

[5] J. Zander, S.-L. Kim. Radio resource management for wireless networks, Artech House, 2001.

[6] C. E. Shannon. A mathematical theory of communication. *Bell Systems Technology Journal*, vol. 27, pp. 379-423, 623-656, 1948.

[7] S. Haykin and M. Moher. Modern wireless communications. Pearson Prentice Hall, 2005.

[8] TR 25.943. Technical specification group radio access networks; deployment aspects. Release 7, 2007.

[9] L. Hanzo, M. Muenster, B. Choi and T. Keller. OFDM and MC-CDMA for broadband multi-user communications, WLANs and broadcasting. New York: Wiley, 2003.

[10] W.C. Jakes. Microwave mobile communications. Wiley, NY, Section 1.7.2, 2001.

[11] 3GPP-RP-060907. Sounding channel for UL channel-dependent scheduling. Nortel, May 2006.

[12] R. Knopp and P. Humblet. Capacity and power control in single cell multiuser communications. In *Proc. International Conference on Communications*, Seattle, WA, June 1995.

[13] D. Tse. Optimal power allocation over parallel Gaussian channels. In *Proc. of IEEEInternational Symposium on Information Theory*, Ulm, Germany, June 1997.

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

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

[16] V. Vukadinovic and E. Drogou. Opportunistic scheduling in wireless networks. Project report, Royal Institute of Technology, 2006.

[17] D. Tse and S. Hanly. Multiaccess fading channels-part I: polymatroid structure, optimal resource allocation and throughput capacities. *IEEE Transactions on Information Theory*, No. 7, 1998.

[18] A. Jalali, R. Padovan and R. Pankaj. 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.

[19] F. Berggren and R. Jäntti. Asymptotically fair transmission scheduling over fading channels. *IEEE Transactions on Wireless Communications*, 3, 326-336, 2004.

[20] S. Shakkottai and A.L. Stolyar. Scheduling for multiple flows sharing a time varying channel: the exponential rule. Bell Labaratories, Tech. Rep. 2000.

[21] S. Shakkottai and A.L. Stolyar. Scheduling algorithms for a mixture of real-time and non-real-time data in HDR. *Proc. Int. Teletrafic Congress*, 793-804, 2001.

[22] K. Chang and Han Y. QoS-based adaptive scheduling for a mixed service in HDR systems. *Proc. IEEE PIMRC 2002*, 4, 1914-1918.

[23] A.L. Stolyar. On the asymptotic pptimality of the gradient scheduling algorithm for multiuser throughput allocation. *Operations Research, Informs*, vol. 53, No. 1, pp 12-25, 2005.

[24] T. Bonald and A. Proutire. Insensitive bandwidth sharing in data networks. *Queueing Systems*, vol 44, pp 69-100, 2003.

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

[26] R. Agrawal and V. Subramanian. Optimality of certain channel aware scheduling policies. In *Proc. of 2002 Allerton Conference on Communication, Control and Computing*, Oct. 2002.

[27] H.J. Kushner and P.A. Whiting. Asymptotic properties of proportional-fair sharing algorithms. *Allerton Conference on Communication, Control and Computing*, Oct. 2002.

[28] J.A. Van Mieghem. Dynamic scheduling with convex delay costs: the generalized cμ rule. *Annals of Applied Probability*, 5(3), 1995.

[29] A. Mandelbaum and A.L. Stoylar. Gcμ scheduling of flexible servers: Asymptotic optimality in heavy traffic. *Proc. of 2002 Allerton Conference on Communication, Control and Computing*, Oct. 2002.

[30] P. Liu, R. Berry and M. Honig. A Fluid analysis of a utility-based wireless scheduling policy. *IEEE Trans. on Information Theory*, 2006.

[31] A.L. Stolyar. MaxWeight scheduling in a generalized switch: state space collapse and equivalent workload minimization in Heavy Traffic. *Annals of Applied probability*, 14(1):1-53, 2004.

[32] L. Tassiulas and A. Ephremides. Dynamic server allocation to parallel queue with randomly varying connectivity. In *IEEE Transactions on Information Theory*, vol. 39, pp. 466-478, March 1993.

[33] M. Andrews, L. Qian , A.L. Stolyar. Optimal utility based multi-user throughput allocation subject to throughput constraints. *INFOCOM'2005*, Miami, March 13-17.

[34] F. Long and G. Feng. Efficient rate guaranteed opportunistic scheduling for wireless networks. *IEEE Transactions on Vehicular Technology*, 2007.

[35] M. Assad and D. Zeghlache. Opportunistic scheduling for streaming services in HSDPA. In *Proc. 17th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC '06)*, 2006.

[36] H. G. Myung, K. Oh, J. Lim, and D. J. Goodman. Channel-dependent scheduling of an uplink SC-FDMA system with imperfect channel information. *IEEE Wireless Communications and Networking Conference (WCNC) 2008*, Las Vegas, USA, Mar. 2008.

[37] H. Holma, A. Toskala. LTE for UMTS-OFDMA and SC-FDMA based radio access. John Wiley and Sons, 2009.

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

[39] K. Jersenius. Uplink channel dependent scheduling for future cellular systems. Master thesis, Linköping university, 2007.

[40] R. Agrawal, A. Bedekar, R.J. La and V. Subramanian. Class and channel condition based weighted proportional fair scheduler. *Teletraffic Engineering in the internet*, in Proc. ITC-17, pp. 553-565, 2001.

[41] S. Borst, P. Whiting. Dynamic rate control algorithmsfor HDR throughput optimization. In *Proc. Infocom*, pp. 976-985, 2001.

[42] X. Liu, E.K. Chong, N.B. Shroff. A framework for opportunistic scheduling in wireless networks. In *Proc. Computer Networks* 41, 457-474, 2003.

[43] M. Andrews, K. Kumaran , K. Ramanan, A.L. Stolyar, R. Vijayakumar and P. Whiting. Scheduling in a queueing system with asychronously varying service rates. *Prob. Eng Inf. Sc.* 18, 197-217, 2004.

[44] M. Neely, E. Mondiano and C. Li. Fairness and optimal stochastic control for heterogenous networks. In *Proc. Infocom*, 2005.

[45] M. Neely, E. Mondiano and C. Rohrs. Power and server allocation in a multi-beam satellite with time-varying channels. In *Proc. Infocom*, 1451-1460 2006.

[46] L. Tassiulas and A. Ephremides. Stability properties of constrained queueing systems and scheduling policies for maximum throughputin multihop radio networks. In *IEEE Transactions on Aut. Control*, vol. 37, pp. 1936-1948, 1992.

[47] S. Borst, M. Jonckheere. Flow-level stability of channel-aware scheduling algorithms. In *Proc.* IEEE *WiOpt*, pp. 272-277, 2006.

[48] S. Borst. User-level performance of channel-aware scheduling algorithms in wireless data networks. In *Proc. Infocom*, 2003.

[49] S. Aalto, P. Lassila. Flow-level stability and performance of channel-aware priority-based schedulers. In *Proc. 6th Euro-NF conference on Next Generation Internet (NGI 2010)*, 2010.

[50] 3GPP TSG RAN WG1 R1-060667. Interference coordination for E-UTRA uplink. Alcatel, 2006.

[51] 3GPP TSG RAN WG1 R1-070099. Frequency domain channel-dependent scheduling considering interference to neighbouring cell for E-UTRA uplink. NTT DoCoMo, 2007.

[52] X. Mao, A. Maaref, K.H. Teo. Adaptive soft frequency reuse for inter-cell interference coordination in SC-FDMA based 3GPP LTE uplinks. *IEEE GLOBECOM 2008*. page(s):1 - 6 , 2008.

[53] N. Reider. Inter-cell interference coordination techniques in mobile networks. M.Sc. Thesis, Budapest University of Technology and Economics, 2007.

[54] J. Neel. How does game theory apply to radio resource management? Ph.D. Thesis, Virginia Tech.

[55] J.F. Nash. The bargaining problem. *Econometrica*, 18, 155-162, 1950.

[56] Z. Han, Z. Ji, and K. J. R. Liu. Fair multiuser channel allocation for OFDMA networks using Nash bargaining and coalitions. *IEEE Transactions on Communications*, vol.53, no.8, pp.1366-1376, Aug. 2005.

[57] D. Wu. QoS provisioning in wireless networks. *Wireless Communications and Mobile Computing*, vol. 5, pp. 957–969, Wiley, 2005.

[58] Z. Liu and M.E. Zarki. 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.

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

[60] M. Dianati, R. Tafazolli, X. Shen, and S. Naik. Call admission control with opportunistic scheduling scheme. *Wireless Communications and Mobile Computing*, Wiley, vol. 10, issue 3, pages 372 - 382, 2009.

[61] J. Gomes, H.-A. Choi, J.-H. Kim, J. K. Sohn, and H. I. Choi. Integrating admission control and packet scheduling for quality controlled streaming services in HSDPA networks. *Proc. of IEEE Broadnets 2007*, Raleigh, NC, Sep 10-14.

[62] H.-W. Lee and S. Chong. Combined packet scheduling and call admission control with minimum throughput guarantee in wireless networks. *IEEE Transactions on Wireless Communications*, vol. 6, No. 8, pp. 3080-3089, August 2007.

[63] X. Liu, E.K. Chong and N.B. Shroff. Opportunistic transmission scheduling with resource-sharing constraints in wireless networks. In *Proc. IEEE Journal on Selected Areas in Communications*, 2001a.

[64] H. Lee and S. Chong. Combined QoS scheduling and call admission control algorithm in cellular networks. In *Proc. of IEEE WiOpt* 2006, pp. 114-123, 2006.

[65] M. Hu, J. Zhang. Two novel schemes for opportunistic multiuser communications. *Special session on wireless networks in IEEE Multimedia Signal Processing Workshop*, Dec 2002.

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

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

[68] 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.

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

[70] Z. Zhang, Y. He and E.K.P Chong. Opportunistic downlink scheduling for multiuser OFDM systems. In *Proc. IEEE Networking Conference 2004*, vol. 2 pp. 1206 - 1212, 2004.

[71] T.J. willink and P.H. Wittke. Optimization and performance evaluation of multicarrier transmission. *IEEE Trans. Inform. Theory*, vol. 43, pp. 426-440, Mar. 1997.

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

[73] M. Kobayashi and G. Caire. A practical approach for the weighted sum rate maximization in MIMO-OFDM BC. Asilomar 2007.

[74] K. Fazel and G. Fettewis. Multi-carrier spread-spectrum. Norwell, MA: Kluwer, p. 260, 1997.

[75] A. Czylwik. Adaptive OFDM for wideband radio channels. In *Proc. IEEE Global Telecommunications Conference* (*GLOBECOM '96)*, vol. 1, pp. 713-718, 1996.

[76] L. van der Perre, S. Theon, P. Vandenameele, B. Gyselinckx and M. Engels. Adaptive loading strategy for a high speed OFDM-based WLAN. In *Proc. IEEE Global Telecommunications Conference (GLOBECOM '98)*, vol. 4, pp. 1936-1940, 1998.

[77] A.N. Barreto and S. Furrer. 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.

[78] D. Dardari. 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.

[79] 3GPP-RP-040461. Proposed study item on evolved utra and utran. Technical Report.

[80] C. Y. Wong, R. S. Cheng, K. B. Letaief and R. D. Murch. Multiuser OFDM with adaptive subcarrier, bit and power allocation. *IEEE Journal on Selected Areas in Communications*, vol. 17, no. 10, Oct. 1999.

[81] J. Jang and K. B. Lee. Transmit power adaptation for multiuser OFDM system. *IEEE Journal on Selected Areas in Communications*, vol. 21, no. 2, pp. 171-178, Feb. 2003.

[82] Y. J. Zhang and K. B. Letaief. Adaptive resource allocation and scheduling for multiuser packet-based OFDM networks. In *Proc. of IEEE ICC*, pp. 2949-2953, June 2004.

[83] T. Keller and L. Hanzo. 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.

[84] D. J. Love, R. W. Heath, V. K. N. Lau, D. Gesbert, B. D. Rao, and M. Andrews. An overview of limited feedback in wireless communication systems. *IEEE Journal on Selected Areas in Communications*, vol. 26, No.8, pages: 1341-1365, Oct. 2008.

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

[86] H. Zhang, Y. Li, V. Stolpman, and N. Van Waes. A reduced CSI feedback approach for precoded MIMO-OFDM systems. *IEEE Trans. Wireless Comm.*, vol. 6, no. 1, pp. 55-58, Jan. 2007.

[87] D. Gesbert and M. Alouini. Selective multi-user diversity. In Proc.*IEEE Int. Symp. Signal Proc. Info. Theory*, Dec. 2003, pp. 162- 165.

[88] D. Gesbert and M. Alouini. How much feedback is multi-user diversity really worth? In *Proc. IEEE Int. Conf. on Commun.*, vol. 1, June 2004, pp. 234-238.

[89] V. Hassel, D. Gesbert, M. Alouini, and G. E. Oien. A threshold-based channel state feedback algorithm for modern cellular systems. *IEEE Trans. Wireless Comm.*, vol. 6, no. 7, pp. 2422-2426, July 2007.

[90] S. Sanayei, D. J. Love, and A. Nosratinia. Opportunistic downlink transmission with limited feedback. *IEEE Trans. Info. Th.*, vol. 53, no. 11, pp. 4363-4372, Nov. 2007.

[91] I. C. Wong and B. L. Evans. OFDMA downlink resource allocation for ergodic capacity maximization with imperfect chanel knowledge. *IEEE GLOBECOM*, pp. 3729-3733, Nov. 2007.

[92] R. Agrawal, R. Berry, J. Huang, and V. Subramanian. Optimal scheduling for OFDMA systems. *Proc. of 40th Annual Asilomar Conference on Signals, Systems, and Computers*, (invited paper) Pacific Grove, CA, October 29 - November 1, 2006.

[93] F. Kelly. Charging and rate control for elastic traffic. *European Transactions on Telecommunications*, vol 8, pp 33-37. 1997.

[94] E. Kalai, M. Smorodinsky. Other solutions to Nash's bargaining problem. *Econometrica*, 43, 513-518, 1975.

[95] N. Dagan, O. volij, E. Winter. A characterization of the Nash bargaining solution. *Social Choice and Welfare*, 19, 811-823, Springer 2002.

[96] W. Thompson. Cooperative models of bargaining. *Handbook of Game Theory*, vol.2, chapter 35, Elsevier Science 1994.

[97] D. Pisinger. Algorithms for Knapsack problems. Ph.D. thesis, University of Copenhagen, 1995.

[98] T. Casey, N. Veselinovic, R. Jäntti. Base station controlled load balancing with handovers in mobile WiMAX. In *Proc. IEEE Personal Indoor Mobile Radio Communications PIMRC 2008.*

[99] S. Chae, P. Haidhues. A group bargaining solution. *Mathematical Social Sciences*, Elsevier, 48, 37-53, 2004.

[100] X. Zhang. A note on the group bargaing solution. *Mathematical Social Sciences*, Elsevier, 57, 155-160, 2009.

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

[102] K. Gribanova. Packet scheduling for video streaming data in the HDR system. Master's thesis, telecommunication software and multimedia laboratory, Helsinki University of Technology, 2004.

[103] Pop and Beaulieu. Limitations of sum of sinsoids fading simulator simulators. *IEEE Transactions on Communications*,49, 669-708, 2001.

[104] D. Park, H. Seo, H. Kwon, B. Gi Lee. Wireless packet scheduling based on the cumulative distribution function of user transmission rates. *IEEE Trans. On Communications*, vol. 53, Nov. 2005, pp 1919-1929.

[105] R. Jäntti. Radio resource management methods. Lecture notes, communications and networking department, Aalto university.

[106] S. Sanayei, A. Nosratinia and N. Aldhahir. Code Combining: A maximum-likelihood decoding approach for combining an arbitrary number of noisy packets. *IEEE Transactions on Communications 1985*, vol. 33, pp. 593-607, 1985.

[107] 3GPP-RP-25.892. Feasibility study for OFDM for UTRAN enhancement. Version 1.2.0, May 2004.

[108] T. Kim. Quantized feedback for slow fading channels. Licentiate thesis, KTH, paper A, 2006.

[109] F. Laakso, K. Aho, T. Chapman and T. Ristaniemi. Applicability of interference coordination in highly loaded HSUPA Network. In *Proc. IEEE VTC2001 Spring*, pp. 1-5, 2010.

[110] E. Dahlman, S. Parkvall, J. Sköld and P. Beming. 3G Evolution HSPA and LTE for mobile broadband. Second edition, Elsevier, 2008.

[111] L. Temino, G. Berardinelli, S. Frattasi and P. Mogensen. Channel-aware scheuduling algorithms for sc-fdma in LTE uplink, Personal, In *Proc. IEEE PIMRC 2008*, pp. 1-6, 2008.

[112] 3GPP-RP-010586. HSDPA system performance with variable TTI. Lucent Technologies, May 2006.

Appendix A Validity of models

The models considered in this work are simplifications of real systems. The main goal of the models was to capture the basic characteristics of these real systems. In this appendix an attempt is made 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 the models of this work and the performance of other models from the literature under the same assumptions and with the same parameters. The comparison is merely to give an insight of the credibility of the results. The two main system models considered in this thesis are addressed. The first representing the HSDPA system and the second is the LTE system.

HSDPA

The reference in this section is the HSDPA model used in a 3GPP TSG RAN document [112]. The document mainly discusses HSDPA throughputs with fixed and variable sized TTIs. In the thesis model only fixed TTIs are considered. Therefore, the comparison will be made only with the fixed TTI result. The same simulation parameters used in the paper are considered and applied to our simulator and the cumulative distribution function was computed for packet call throughputs of 37 UEs.

LTE

In this part the comparison will be with a related work model. The reference is the study of Jersenius [39]. Her work mainly suggests a number of scheduling approaches for LTE systems. We will apply some of the scheduling methods proposed in the reference to this work's model using the same scenarios, simulation assumptions and parameters.The following scheduling algorithms will be considered:

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 blocks.

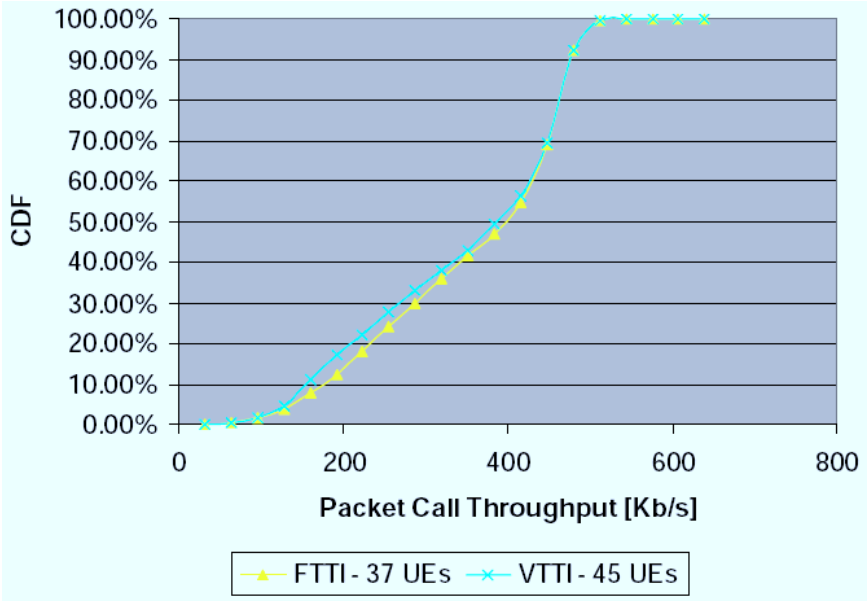

Figure 7.1: CDF of packet call throughput from reference HSDPA model [112]

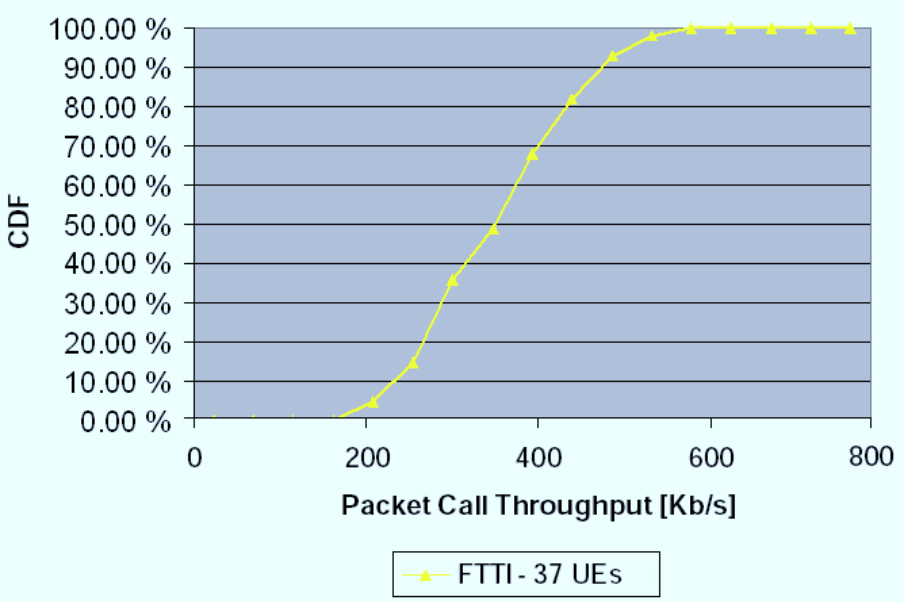

Figure 7.2: CDF of packet call throughput from our model

2. Channel Dependent Frequency and Time Domain Scheduling (CDFT): Resource blocks 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.

Validation

Comparing the reference Figures 7.1 and 7.3 with the Figures produced with our models using the same scenarios and parameters 7.2 and 7.4 reveals that our simulator provides a roughly close result or at least the same order of magnitude to the ones obtained with the simulators considered in the references. It is possible to say that the result obtained from this work's HSDPA and LTE models can be considered reliable since similar performance was produced when similar simulation environments were assumed.

Accuracy of results is an essential part of any study. For this reason, the author resorted to parallel programming where simulations are carried out on multiple processors and the data gathered and analyzed. This method provides accuracy and consistency of results but at a high cost in computing resources. The standard deviation of the data was one benchmark used to measure the reliability of the results.

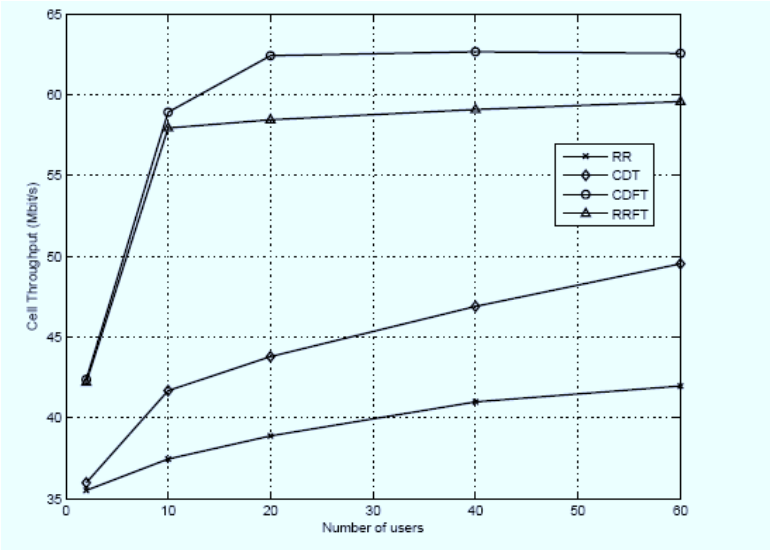

Figure 7.3: Cell throughput for 2, 10, 20, 40 and 60 users from reference model [39]

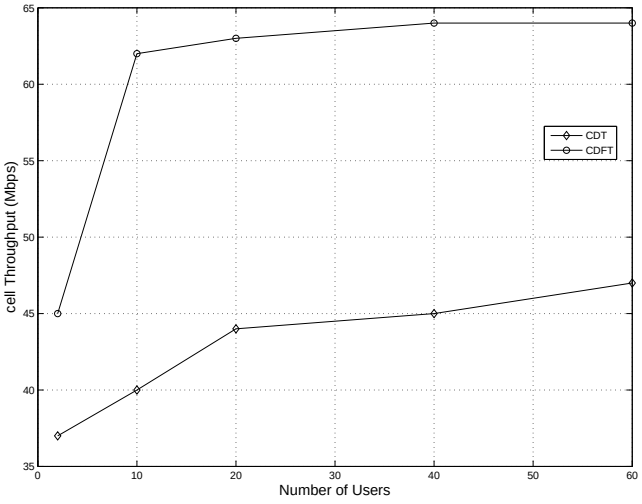

Figure 7.4: Cell throughput for 2, 10, 20, 40 and 60 users from our model

Appendix B Proofs

Proof of Proposition 4.1

The Lagrangian of the problem can be written as:

$L_b[\bar{x}, \phi, \varphi, \eta, \varrho, c] = \sum_{i\in\mathcal{N}_{c,b}\cup\mathcal{N}_{e,b}} u_\alpha[\bar{x}_i] -$
$\sum_{i\in\mathcal{N}_{c,b}} \eta_i \left(\bar{x}_i - \mu_i(l\phi_i + (L-l)\varphi_i) - \sum_{i\in\mathcal{N}_{e,b}} \eta_i(\bar{x}_i - \mu_j l\phi_i)\right.$
$+ \varrho_1 \left(1 - \sum_{i\in\mathcal{N}_{c,b}\cup\mathcal{N}_{e,b}} \phi_i - c_1\right) + \varrho_2 \left(1 - \sum_{i\in\mathcal{N}_{c,b}} \varphi_i - c_2\right)$

where η and ϱ are the lagrangian parameters. c is the slack variable.

$$\frac{d}{dx_i} L_b[\bar{x}, \phi, \varphi, \eta, \varrho, c] = 0 \Leftrightarrow \frac{1}{\bar{x}_i^\alpha} - \eta_i = 0 \Rightarrow \bar{x}_i = \frac{1}{\eta_i^{\frac{1}{\alpha}}}, \forall i \tag{A-1}$$

$$\frac{d}{d\phi_i} L_b[\bar{x}, \phi, \varphi, \eta, \varrho, c] = 0 \Leftrightarrow 0 = \eta_j r_j l - \varrho_1 = 0 \Rightarrow \eta_i = \frac{\varrho_1}{\mu_j l}, i \in \mathcal{N}_{e,b} \tag{A-2}$$

$$\begin{aligned}\frac{d}{d\varphi_i} L_b[\bar{x}, \phi, \varphi, \eta, \varrho, c] = 0 \Leftrightarrow 0 = \eta_j \mu_i (L-l) - \varrho_2 = 0 \\ \Rightarrow \eta_i = \frac{\varrho_2}{\mu_i(L-l)}, i \in \mathcal{N}_{c,b}\end{aligned} \tag{A-3}$$

Either the slack variable c_κ is zero or the lagrangian ϱ_k is zero, $\kappa = 1, 2$. By studying the above conditions, we can conclude that the slack variables must be zero and thus the constraints are binding. From (A-1), (A-2) and (A-3) we can conclude that

$$\bar{x}_i = \begin{cases} \left(\frac{\mu_j l}{\varrho_1}\right)^{\frac{1}{\alpha}}, & i \in \mathcal{N}_{e,b} \\ \left(\frac{\mu_i (L-l)}{\varrho_2}\right)^{\frac{1}{\alpha}}, & i \in \mathcal{N}_{c,b} \end{cases} \tag{A-4}$$

Furthermore, either $\phi_i = 0$ or $\varrho_2 = \frac{L-l}{l}$. Let us first consider the case, in which $\eta_i > 0$. It follows from (A-1) that the optimal throughputs of the users can be expressed as:

$$\bar{x}_i = \begin{cases} \mu_j l\phi_i \Rightarrow \phi_i = \frac{(\mu_j l)^{\frac{1}{\alpha}-1}}{\varrho_1^{\frac{1}{\alpha}}}, & i \in \mathcal{N}_{e,b} \\ \mu_i(l\phi_i + (L-l)\varphi_i) \Rightarrow \phi_i = \frac{(\mu_j l)^{\frac{1}{\alpha}-1}}{\varrho_1^{\frac{1}{\alpha}}} - \frac{(L-l)}{l}\varphi_i & i \in \mathcal{N}_{c,b} \end{cases}$$

Using the fact that the constraints are binding, b.e.
$\sum_{i\in\mathcal{N}_{c,b}\cup\mathcal{N}_{e,b}}\phi_i = 1$ and $\sum_{i\in\mathcal{N}_{c,b}}\varphi_i = 1$, we can solve the lagrangian $\varrho_1^{\frac{1}{\alpha}}$.

$$\sum_{i\in\mathcal{N}_{c,b}\cup\mathcal{N}_{e,b}}\phi_i = \frac{1}{\varrho_1^{\frac{1}{\alpha}}}\sum_{i\in\mathcal{N}_{c,b}\cup\mathcal{N}_{e,b}}(\mu_j l)^{\frac{1}{\alpha}-1} - (\frac{L-l}{l})\sum_{i\in\mathcal{N}_{c,b}}\varphi_i = 1 \tag{A-5}$$

$$\Rightarrow \varrho_1^{\frac{1}{\alpha}} = \frac{l}{L}\sum_{i\in\mathcal{N}_{c,b}\cup\mathcal{N}_{e,b}}(\mu_j l)^{\frac{1}{\alpha}-1} \tag{A-6}$$

Substituting (A-6) into (A-6) and simplifying the equation gives the solution b). If $\eta_i = 0$ for all $i \in \mathcal{N}_{c,b}$, we can solve $\varrho_1^{\frac{1}{\alpha}}$ with $\sum_{i\in\mathcal{N}_{e,b}} = 1$ and $\varrho_2^{\frac{1}{\alpha}}$ with $\sum_{i\in\mathcal{N}_{c,b}} = 1$ which gives the solution ii). Assume now that $l < L$, $\eta_i > 0$ for $i' \in \mathcal{N}_{c,b}$ and $\eta_i = 0$ for all $i \neq i'$, $i \in \mathcal{N}_{c,b} \neq \emptyset$. Now we must have $\varrho_2 = \frac{L-l}{l}$ and $\bar{x}_{i'} = \left(\frac{\mu_j l}{\varrho_1}\right)^{\frac{1}{\alpha}}$. Let us define $M = \sum_{i\in\{i'\}\cup\mathcal{N}_{e,b}}(\mu_j l)^{\frac{1}{\alpha}-1}$ and $K = \sum_{i\in\mathcal{N}_{c,b},i\neq i'}(\mu_j l)^{\frac{1}{\alpha}-1}$. Now (A-5) becomes

$$\sum_{i\in\mathcal{N}_{c,b}\cup\mathcal{N}_{e,b}}\phi_i = \frac{1}{\varrho_1^{\frac{1}{\alpha}}}M - \left(\frac{L-l}{l}\right)\phi_{i'} = 1$$

$$\Rightarrow \frac{1}{\varrho_1^{\frac{1}{\alpha}}} = \frac{1-\left(\frac{L-l}{l}\right)\phi_i}{M} \tag{A-7}$$

For $i \neq i'$, $i \in \mathcal{N}_{c,b}$, we have

$$\varphi_i = \frac{((L-l)\mu_i)^{1-\frac{1}{\alpha}}}{\varrho_2^{\frac{1}{\alpha}}} = \frac{l}{L-l}\frac{((L-l)\mu_i)^{1-\frac{1}{\alpha}}}{\varrho_1^{\frac{1}{\alpha}}} \tag{A-8}$$

The allocations must sum up to 1. Hence,

$$\sum_{i\in\mathcal{N}_{c,b}}\varphi_i = \frac{l}{L-l}\frac{1}{\varrho_1^{\frac{1}{\alpha}}}K - \varphi_{i'} = 1 \Rightarrow \frac{1}{\varrho_1^{\frac{1}{\alpha}}} = \frac{L-l}{l}\frac{1-\varphi_{i'}}{K} \tag{A-9}$$

Combining (A-7) and (A-9),we get $\varphi_{i'} = 1$. This implies that for all $i \neq i'$, $i \in \mathcal{N}_{c,b} \neq \emptyset$, we get φ_i which implies $\bar{x}_i = \left(\frac{\mu_i(L-l)}{\varrho_2}\right)^{\frac{1}{\alpha}}$ which is true only if $l = L$ which contradicts our requirement $l < L$. Hence, we can conclude that there are no mixed solutions in which only some of the center users would be allocated to LI-RBs. It can be noted that for edge users the solution ii) would give higher throughput than

b) if

$$l > \frac{\sum_{i\in\mathcal{N}_{e,b}} \mu_i^{\frac{1}{\alpha}} - 1}{\sum_{i\in\mathcal{N}_{c,b}\cup\mathcal{N}_{e,b}} \mu_i^{\frac{1}{\alpha}-1}} L \tag{A-10}$$

But since solution b) corresponds also to the optimal solution when $l = L$, b.e. all users have access to all resource blocks, we can conclude that solution ii) is only used when (A-10) does not hold. This concludes the proof. ■

Proof of Proposition 4.2

Assume the case that all users are using minimum rates. For $i \in \mathcal{N}_{e,b}$, we get

$$\sum_{i\in\mathcal{N}_{e,b}} \phi_i = \sum_{i\in\mathcal{N}_{e,b}} \frac{\bar{x}_{min,i}}{\mu_j l} \leq 1 \Rightarrow l \geq \frac{1}{\sum_{i\in\mathcal{N}_{e,b}} \frac{\bar{x}_{min,i}}{\mu_j l}} \tag{A-11}$$

and for $i \in \mathcal{N}_{c,b}$

$$\sum_{i\in\mathcal{N}_{c,b}} \varphi_i = \sum_{i\in\mathcal{N}_{e,b}} \frac{\bar{x}_{min,i}}{\mu_i (L-l)} \leq 1 \Rightarrow l \leq L - \frac{1}{\sum_{i\in\mathcal{N}_{c,b}} \frac{\bar{x}_{min,i}}{\mu_j l}} \tag{A-12}$$

Combining (A-11) and (A-12), we get

$$\frac{1}{\sum_{i\in\mathcal{N}_{c,b}} \frac{\bar{x}_{min,i}}{\mu_i}} + \frac{1}{\sum_{i\in\mathcal{N}_{e,b}} \frac{\bar{x}_{min,i}}{\mu_i}} \leq L$$

This concludes the proof. ■

Proof of Proposition 4.3

In case the problem does not have a feasible solution, the system is overloaded and some means of congestion control needs to be applied. That is, either some users need to be removed or their minimum rate constraints need to be relaxed. Subsequently, it will be assumed that both methods are utilized and the cell coordination problem has a feasible solution. ■

Proof of Proposition 4.4

Algorithm 4.1 is a special case of Algorithm 7.1 in the study by Pisinger that solves the linear programming (LP) relaxed MCKP in polynomial time [97]. The LP relaxation happens to give integer solutions when $m_{b,k} - m_{b,(k-1)} = 1$. ■

Proof of Lemma 5.1

Assume at time-slot $k = 0$, the system with N users was in steady state. Assume further that an admission control rule is used, 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_b^k. Define

$$\bar{x}_i[\mathcal{A}_N; \mathcal{A}_{N+1}] = E\{\mu_i(k)\chi_i(\mu_i(k), \bar{x}_i(\mathcal{A}_{N+1}), \mathcal{A}_N)\} \tag{A-13}$$

and

$$\begin{aligned} \bar{x}_{lower,i}(k) &= p_b^k \bar{x}_i[\mathcal{A}_N; \mathcal{A}_{N+1}] + (1 - p_b^k)\bar{x}_i[\mathcal{A}_{N+1}; \mathcal{A}_{N+1}] & \text{(A-14)} \\ \bar{x}_{upper,i}(k) &= p_b^k \bar{x}_i[\mathcal{A}_{N+1}; \mathcal{A}_N] + (1 - p_b^k)\bar{x}_i[\mathcal{A}_N; \mathcal{A}_N] & \text{(A-15)} \end{aligned}$$

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

$$\lim_{k\to\infty} \tilde{x}_i(k) = \lim_{k\to\infty} \bar{x}_{lower,i}(k) = \bar{x}_i(Z_{N+1}) \tag{A-16}$$

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

$$\begin{aligned} \bar{x}_{lower,i}(k) &= p_B \bar{x}_{lower,i}(k-1) + (1 - p_B)\bar{x}_i[\mathcal{A}_{N+1}; \mathcal{A}_{N+1}] & \text{(A-17)} \\ \bar{x}_{upper,i}(k) &= p_B \bar{x}_{upper,i}(k-1) + (1 - p_B)\bar{x}_i[\mathcal{A}_N; \mathcal{A}_N] & \text{(A-18)} \end{aligned}$$

Based on (A-16) we can write

$$\begin{aligned} \tilde{x}_i(k) &\geq p_B \bar{x}_{lower,i}(k-1) + (1 - p_B)\bar{x}_i[\mathcal{A}_{N+1}, \mathcal{A}_{N+1}] & \text{(A-19)} \\ \tilde{x}_i(k) - \bar{x}_{lower,i}(k-1) &\geq -(1 - p_B)[\bar{x}_{lower,i}(k-1) - \bar{x}_i[\mathcal{A}_{N+1}, \mathcal{A}_{N+1}]] & \text{(A-20)} \end{aligned}$$

■

Proof of Lemma 5.2

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

Proof of Proposition 5.1

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:

$$g_i(t) = tq_{i1} + (1-t)q_{i2} - \beta_i \bar{s}_i(tQ_1 + (1-t)Q_2) + \beta_i \bar{s}_i^{req} \tag{A-21}$$

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

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

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

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

which establishes the contraction property. ■

Proof of Proposition 5.2

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

$$q_i^* = \mathcal{T}_i(Q^*, B) \tag{A-22}$$

$$q_i^* = q_i^* + \beta_i \left(\bar{x}_i^{req} - q_i^* \sum_{k\in\mathcal{K}_i} \bar{x}_i(\mathcal{A}_k)\pi_{k,i}^* \right) \tag{A-23}$$

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

$$\bar{x}_i^{req} = q_i^* \sum_{k\in\mathcal{K}_i} \bar{x}_i(\mathcal{A}_k)\pi_{k,i}^* \tag{A-24}$$

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. ■

Proof of Proposition 5.3

The assumption is that at least one user has $\bar{x}_i^{req}$ larger than what can be supported by the system meaning that 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

$$\bar{x}_i^* = q_i^* \sum_{k \in \mathcal{K}_i} \bar{x}_i(\mathcal{A}_k) \pi_{k,i}^* \tag{A-25}$$

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

$$\pi_{k,i}^* = \prod_{\substack{j \in \mathcal{A}_k \\ j \neq i}} q_j^* \prod_{z \in \mathcal{N} \setminus \mathcal{A}_k} (1 - q_z)^* \quad k \in \mathcal{K}_i \tag{A-26}$$

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

$$\bar{x}_i^* = q_i^* \sum_{k \in \mathcal{K}_i} \bar{x}_i(\mathcal{A}_k) \pi_{k,i}^* > \bar{x}_i(\mathcal{A}_S) \quad | \quad \exists j \quad s.t. \quad \bar{x}_j^* < \bar{x}_j(\mathcal{A}_S) \tag{A-27}$$

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

ISBN 978-952-60-3374-7
ISBN 978-952-60-3375-4 (PDF)
ISSN 1795-2239
ISSN 1795-4584 (PDF)