Helsinki University of Technology Communications Laboratory Technical Report T53
Teknillinen korkeakoulu Tietoliikennelaboratorio Raportti T53
Espoo 2005

QoS MANAGEMENT IN UMTS TERRESTRIAL RADIO ACCESS FDD NETWORKS

David Soldani

Dissertation for the degree of Doctor of Science in Technology (Doctor of Philosophy) to be presented with due permission of the Department of Electrical and Communications Engineering, Helsinki University of Technology, for public examination and debate in Auditorium S2 at Helsinki University of Technology (Espoo, Finland) on the 21st October, 2005, at 12 o'clock noon.

Helsinki University of Technology
Department of Electrical and Communications Engineering
Communications Laboratory

Teknillinen korkeakoulu
Sähkö- ja tietoliikennetekniikan osasto
Tietoliikennelaboratorio

Distributor:
Helsinki University of Technology
Communications Laboratory
P.O. Box 3000
FIN-02015 HUT
Tel. +358-9-451 2366
Fax +358-9-451 2345

ISBN 951-22-7833-2
ISBN 951-22-7834-0
ISSN 0356-5087
URL: http://lib.hut.fi/Diss/2005/isbn9512278340/

Edita Prima Ltd
P.O. Box 510, 00043 EDITA
Espoo 2005

Soldani, David, **QoS Management in UMTS Terrestrial Radio Access FDD Networks**, Helsinki University of Technology Communication Laboratory, Technical Report T53, Espoo 2005, 234 pp. ISBN 951-22-7833-2. ISBN 951-22-7834-0. ISSN 0356-5087.

Keywords: UTRA FDD, UTRAN, WCDMA, UMTS, Quality of Service, Quality of Experience, Radio Resource Management with QoS Differentiation, QoS Provisioning, QoS Monitoring, Service Driven Radio Network Planning, Service Performance, Service Optimisation

Abstract

This work investigates the role and importance of some of the key aspects of QoS planning, provisioning, monitoring and optimisation (QoS Management) for UMTS Terrestrial Radio Access (UTRA) FDD networks within the framework of the 3rd Generation Partnership Project (3GPP).

Firstly, the differences between Quality of end user Experience (QoE) and Quality of Service (QoS) are explained. This is followed by a review of 3GPP requirements for QoS concept and architecture. Then all models and the main assumptions in this dissertation are presented. Based on these, original QoS mechanisms in the radio access network domain, means and methods for QoS provisioning, planning, monitoring and "optimisation" are discussed.

Simulation results showed substantial spectral efficiency gains provided by service (or user) differentiation in UTRAN by means of priorities and differentiated parameter settings. When appropriately configured, the proposed QoS mechanisms can greatly reduce the need for bandwidth. Performance results proved also the proposed virtual time simulator to be an appropriate tool for service driven WCDMA radio interface dimensioning and detailed radio network planning.

It is also shown that measuring QoS performance by a proper classification of counters (and or gauges), based on a particular subset of radio access bearer attributes, is a promising technique for assessing performances of service applications through WCDMA networks. With this new method there is no need to trace upper layer protocols at different interfaces or dumping data in mobile terminals. The proposed metrics allow operators to measure the bandwidth required for robust statistical reliability, to assess and exploit statistical sharing of resources, to configure QoS functions effectively, and to monitor QoE. The application of the proposed technique is not limited to the WCDMA Radio Network Subsystem (RNS), yet it can be deployed in any radio access and packet core network supporting mapping of performance indicators onto a particular subset of QoS attributes.

Finally, in order to maximise the performance of the available services in UTRAN, at a given QoE, simulation results showed clear needs for the network administrator to adapt the parameter settings to diverse input application traffic conditions and the proposed genetic approach to be an appropriate solution space search algorithm for this purpose.

Preface

This thesis is based on the work carried out in Network System Research, Nokia Networks, during the period Jan. 2000 – Dec. 2004.

I wish to express my gratitude to Prof. Sven-Gustav Häggman, supervisor of this work and head of the Communication Laboratory at the Helsinki University of Technologies, for his interest and advise during this project, to Dr. Bo Hagerman, senior specialist at Ericsson Research (Sweden), and Dr. Andrea Abrardo, associate professor at University of Siena (Italy), for reviewing this monograph, as pre-examiners of this dissertation.

Also, Mr. Kari Sipilä, head of the 3G Radio System Research team in Espoo until December 2004, Mr. Lauri Oksanen, head of Network System Research department in Nokia Networks, Mr. Antti Myllykangas, director of Service Configuration in Nokia Networks Operations Solutions, and Mr. Javier Muñoz, operability system research & specifications program area manager in 2003, are acknowledged for giving me the opportunity to spend almost four years in dealing with QoE and QoS management issues in UTRA FDD networks.

I am deeply grateful to my colleagues Dr. Jaana Laiho, Mr. Kari Sipilä, Mr. Achim Wacker, Dr. Kimmo Valkealahti, Ms. Nilmini Lokuge, Mr. Renaud Cuny, Mr. Juho Pirskanen, Mr. Pekka T. Kohonen, Mr. Antti Kuurne, Mr. Markus Djupsund, Mr. Mikko Kylväjä and Dr. Man M. Li for fruitful teamwork, contributions, discussions and comments through these years.

Additionally, I would like to thank my colleague Sandro Grech for the final review of this monograph, and my internal customers Ms. Anneli Korteniemi, Mr. Kalle Tarpila, Mr. Mika Kiikkilä, Mr. Markku T. Mäkinen, Mr. Erkka Ala-Tauriala, and Ms. Outi Hiironniemi for their valuable steering, suggestions and contributions to this endeavour.

Ultimately, I would like to dedicate this dissertation to my wife Romina for her love and support during these years, and our child Martin David, who was born on Dec. the 1st 2003.

Espoo, June 2005

David Soldani

Contents

Symbols

$100(1-\alpha)$	Percentage confidence interval
α	DL code orthogonality, probability of type I error in test of hypotheses
α_{im}	Orthogonality factor for MS i_m
β	Probability of type II error in test of hypotheses
β_c	DPCCH gain factor
β_d	DPDCH gain factor
ΔP_{GB}	Estimated transmission power increase for GB traffic
ΔP_{NGB}	Estimated transmission power increase for NGB traffic
ΔP_{Tx}	Estimated downlink transmission power increase
η	Fractional loading
λ	Mean call arrival intensity; inverse Gaussian distribution parameter
μ	Mean of a population
ν_k	Activity Factor (AF) of service application k
θ	Parameter, e.g. mean or standard deviation, of a population
θ_0	Parameter taken from a sample of the population
ρ	Required E_b/N_0
ρ_c	Energy per chip per noise spectral density (CPICH E_c/N_0)
ρ_{im}	Required E_b/N_0 for MS i_m
ρ_k	required E_b/N_0 for service service k
σ	Standard deviation of a population
$\tau_{DPCH,n}$	Frame timing offset between n:th DPCH and P-CCPCH
τ_{PICH}	Frame timing offset between k:th PICH and related k:th S-CCPCH
$\tau_{S\text{-}CCPCH,k}$	Frame timing offset between k:th S-CCPCH and P-CCPCH
A_i	Mean arrival rate for service i
$B_i^{m,k}$	Delivered RLC blocks of size k by entity i, for management class m
d_i	Duration of $B_i^{m,k}$ during the sampling period s
$\bar{E}$	Mean value of quantity E
E_b	Energy per (RLC) bit
E_s	Energy per symbol
F_i^m	Number of RLC blocks to be Tx after segmentation by AM RLC entity i
G	Geometry factor
G_b	Interface between GGSN and SGSN
G_i	Interface between GGSN and External Packet Data Network
G_n	Interface between SGSN and GGSN

H_0	Null hypothesis
H_1	Alternative hypothesis
i	Others to own cell interference ratio
$\hat{I}_{or}$	Received own orthogonal interference
$I(m)$	Set of MS indices served by BS m
$i_{k,DL}$	Other to own cell interference ratio for bearer service k
i_m	Index of a MS served by BS m
I_{oth}	Received interference due to other cells
I_u	Interface between serving radio network controller and SGSN
I_{ub}	Interface between base station and radio network controller
$I_{u\text{-}BC}$	$\mathrm{I_u}$ interface for Broadcast Centre
$I_{u\text{-}CS}$	$\mathrm{I_u}$ interface for Circuit Switched domain
$I_{u\text{-}PS}$	$\mathrm{I_u}$ interface for Packet Switched domain
I_{ur}	Interface between serving and drifting radio network controller
J_i	Jitter of bearer service i
k	Inactivity power weighting factor for NGB traffic
$L_{(N)GB}$	Actual load in the cell due to (N)GB traffic
$L_{m,im}$	Pathloss from BS m to MS i_m served by BS m
$L_{n,im}$	Pathloss from BS n to MS i_m served by BS m
L_{Target}	Target load in the cell
L_{Total}	Total load in the cell, also presented as L_{DL}
M	Number of cells
m	Management class, subset of RAB attributes
m, n	Indices of BSs
n	Random sample size
$N(0,1)$	Normal distribution with mean zero and standard deviation 1
N_0	Noise spectral density
N_C	Number of Columns
$N^{C,RM}{}_{DCH}$	Number of bits per radio frame after layer 1 rate matching
$N^{C}{}_{DCH}$	Number of bits per radio frame after radio frame equalisation
N_i	Offered traffic in number of subscribers per cell for service i
N_{im}	Noise power (thermal plus equipment) of MS i_m
$N^{m}{}_{B}$	Number of Bearer services of MC m during the measurement interval S
N_R	Number of Rows
N_{SF}	Total number of DPDCH bits available per radio frame for a specific SF
$N^{U}{}_{DCH}$	Number of bits per radio frame prior to CRC attachment

$\hat{P}$	Estimator of parameter p
$Offset$	Downlink overload margin
p	Binomial parameter (proportion)
P_a	Power exploited by active traffic
P_{GB}	Downlink transmission power exploited by GB traffic
P_i	Probability of a user to make use of service i
$P_{i,DTX}$	Power reserved for bearer services in DTX
$P_{i,GB}$	Power reserved for inactive GB traffic
p_{im}	BS transmitted power for MS i_m
P_i^m	Total number of transmitted blocks during the sampling period s
P_m, P_n	Total transmit power of BS m and BS n
P_{NGB}	Downlink transmission power exploited by NGB traffic
$P^{NGB}{}_{Allowed}$	Power budget for NGB traffic, also denoted as $P_{NGBallowed}$ or P_{BP}
P_{SC}	Power scheduled for NGB traffic
P_{Target}	Target (planned) cell transmission power
P_{TxLink}	Radio link transmission power
P_{TxMax}	Maximum transmission power in the cell
$P_{TxNGBcapacity}$	Downlink capacity dedicated to NGB traffic
$P_{TxTotal}$	Total cell transmission power
q	Binomial parameter (1-p)
QE	Quality estimate, BLER after the 1st iteration.
Q^m	Service data unit error ratio for management class m
R	DCH user bit rate (RLC bit rate without header overhead)
r	DPCCH, DPDCH gain factor ratio
R	Interface between TE and MT
$R^{C,RM}{}_{DCH}$	Layer 1 encoding rate after coding and rate matching
Ref 1	Reference parameter settings for differentiated provisioning
Ref 2	Reference parameter settings for undifferentiated provisioning
$R_i(t)$	Measured average AST at the time t for the connection i
R_{im}	Bit rate used by MS i_m
$R^i{}_{Target}$	Target bit rate of connection i
R_k	Transport channel user bit rate for bearer service k
r_k	RLC PDU size k in bits without header
RM_{DCH}	Rate matching attribute for a specific DCH
RP	Reporting period for measurement computation

$RRP_i(t)$	Radio resource priority of the connection *i* at the time *t*
s	Standard deviation of a random sample, measurement sampling period
S	Observation time, measurement period
S_a	Actual user (call/session) satisfaction ratio
S_c	Computed user (call/session) satisfaction ratio
SF	Spreading factor
S_i	Share of subscription to service *i*
s_{tot}	Total satisfied users (calls/sessions) during the simulated period
s_{tr}	Satisfied users (calls/sessions) during the transient time
SW	Sliding window size for average computation
$t_{\alpha/2,\, n\text{-}1}$	Upper $\alpha/2$ percentage point of the *t* distribution with *n*-1 degree of freedom
TD_i	Transfer delay of bearer service *i*
T_i	Mean service time for service *i*; virtual time step size
T_{in}	Inactivity Timer
t^m	Downlink throughput for a particular management class *m*
T^m	Downlink cell throughput for a particular management class *m*
U_i	Average number of active bearers carrying the service *i*
U^m	Total number of RLC retransmission for the management class *m*
u_{tot}	Total call/session arrivals (users) during the simulated period
u_{tr}	Call/session arrivals during the transient time
U_u	UMTS radio interface
W	Chip rate (3.84 Mchip/s),
w_i	Differentiated power weight for bearer service *i*
x	Priority step parameter
$\bar{x}$	Mean of a random sample
x_i	Observation number *i* in a random sample of size *n*
X_i^m	SDU detected as erroneous or discarded for MC *m* and AM RCL entity *i*
Y_i^m	Total number of SDU for MC *m* and AM RCL entity *i*
Z_0	Test statistic
z_0	Computed value of test statistic Z_0

Abbreviations

0-9	
1G	First Generation
2G	Second Generation
3G	Third Generation
3GPP	3rd Generation Partnership Project
A	
AC	Admission Control
AF	Activity Factor
AICH	Acquisition Indication Channel
ALCAP	Access Link Control Application Part protocol
AM	Acknowledge Mode
AMR	Adaptive Multi Rate speech codec
APN	Access Point Name
ARP	Allocation Retention Priority
ARQ	Automatic Repeat Request
AS	Access Stratum
ASC	Access Service Class
AST	Active Session Throughput
AuC	Authentication Centre (Register)
B	
BB	Backbone
BCCH	Broadcast Control Channel
BCH	Broadcast Channel
BER	Bit Error Rate (Ratio)
BG	Border Gateway
BLER	Block Error Ratio
BMC	Broadcast/Multicast Control
BS	Bearer Service, Base Station (or Node B)
BTFD	Blind Transport Format Detection
C	
CAPEX	CAPital EXpenditure
CBR	Call Block Ratio
CC	Call Control, Convolutional Coding
CCCH	Common Control Channel
CCH	Common Channel (s)
CCTrCH	Code Composite Transport Channel
CDR	Call Drop Ratio
CFN	Connection Frame Number
CID	Cell Identifier
CM	Configuration or Connection Management
CN	Core Network
CP	Control Plane
CPCH	Common Packet Channel
CPICH	Common Pilot Channel (Perch Channel)
CR	Capacity Request
CRC	Cyclic Redundancy Check
CRNC	Controlling RNC

CRRR	Capacity Request Rejection Ratio
CS	Circuit Switched
CTCH	Common Traffic Channel
D	
DBR	Delay Buffering Ratio
DCCH	Dedicated Control Channel
DCH	Dedicated Channel
DCS	Digital Communication System
DECT	Digital Enhanced Cordless Telephone
DHO	Diversity Handover
DL	Down Link
DNS	Domain Name Server
DPCCH	Dedicated Physical Control Channel
DPCH	Dedicated Physical Channel
DPDCH	Dedicated Physical Data Channel
DRNC	Drifting RNC
DRX	Discontinuous Reception
DSCH	Downlink Shared Channel
DTCH	Dedicated Traffic Channel
DTX	Discontinuous Transmission
E	
EDF	Earliest Detect First
EDGE	Enhanced Data Rates for GSM Evolution
EGPRS	Enhanced GPRS
EIR	Equipment Identity Register
ETSI	European Telecommunications Standards Institute
F	
FACH	Forward Access Channel
FDD	Frequency Division Duplex
FER	Frame Erasure Ratio
FIFO	First In First Out
FP	Frame Protocol
FR	Fair Resources
FT	Fair Throughput
FTP	File Transfer Protocol
G	
GB	Guaranteed Bit rate
GERA	GSM/Edge Radio Access
GERAN	GSM/Edge Radio Access Network
GGSN	Gateway GPRS Support Node
GMM	GPRS Mobility Management
GMSC	Gateway MSC
GPRS	General Packet Radio Service
GSM	Global System for Mobile Communications
GSN	GPRS Support Node
GTP	GPRS Tunnelling Protocol
GW	Gateway
H	
HC	Handover Control
HLR	Home Location Register
HMA	Hybrid Multiple Access

HO	Handover
HS	High Speed
HSDPA	High Speed Downlink Packet Access
HS-DPCCH	High Speed Dedicated Physical Control Channel
HS-DSCH	High Speed DSCH
HS-PDSCH	High Speed Physical DSCH
HS-SCCH	High Speed Shared Control Channel
HTML	Hypertext Markup Language
HTTP	Hyper Text Transfer Protocol
I	
IEEE	Institute of Electrical and Electronic Engineering
IETF	Internet Engineering Task Force
IMS	IP Multimedia Subsystem
IMSI	International Mobile Subscriber Identity
IMT	International Mobile Telephony
IP	Internet Protocol
IPsec	IP security
ISCP	Interference Signal Code Power
ISDN	Integrated Services Digital Network
ITU	International Telecommunications Union
K	
KPI	Key Performance Indicator
L	
L 1, 2, 3	Layer 1, 2, 3
LA	Link Adaptation
LAC	Link Access Control
LC	Load Control, Congestion Control, Location Area
LoCH	Logical Channel
M	
MAC	Medium Access Control
MC	Management Class
MCC	Mobile Connection Control unit
MCU	Main Control Unit
MDC	Macro diversity Combiner
MM	Mobility Management
MMS	Multimedia Messaging Service
MS	Mobile Station
MSC	Mobile Switching Centre
MSS	Mobile Satellite Spectrum
MT	Mobile Terminal
N	
NAS	Non Access Stratum
NB	Node B
NBAP	Node B Application Part protocol
NE	Network Element
NGB	Non Guaranteed Bit rate
NMS	Network Management System
NPM1	Non Prioritised Mapping 1
NRT	Non Real Time

NSAPI	Network layer Service Access Point Identifier
NW	Network
O	
OH	Overhead
OLPC	Outer Loop Power Control
OLPCE	Outer Loop Power Control Entity
OPEX	OPEration and Management EXpenditure
P	
PC	Power Control
PCCH	Paging Control Channel
P-CCPCH	Primary Common Control Physical Channel
PCH	Paging Channel
PCPCH	Physical Common Packet Channel
PCS	Personal Communication System
PD	Packet or Protocol Data
PDC	Personal Digital Cellular (2G system in Japan)
PDCP	Packet Data Convergence Protocol
PDFRM1	Prioritised Differentiated Fair Resources Mapping 1
PDFTM1	Prioritised Differentiated Fair Throughput Mapping 1
PDM1	Prioritised Differentiated parameters Mapping 1
PDM2	Prioritised Differentiated parameters Mapping 2
PDN	Packet Data Network
PDP	Packet Data Protocol, e.g., IP
PDSCH	Physical Downlink Shared Channel
PDU	Protocol Data Unit
PHS	Personal Handy-phone System
PI	Performance Indicator
PICH	Paging Indicator Channel
PLMN	Public Land Mobile Network
PM	Performance Monitoring or Management
PM1	Prioritised Mapping 1
PoC	Push to talk over Cellular
PPP	Point-to-Point Protocol
PRACH	Physical Random Access Channel
PrC	Process Call function
PS	Packet Switched, Packet Scheduler
PSTN	Public Switched Telephone Network
PU	Payload Unit
Q	
QoE	Quality of end user Experience
QoS	Quality of Service
R	
RA	Routing Area
RAB	Radio Access Bearer
RACH	Uplink Random Access Channel
RAI	Routing Area Identity
RAN	Radio Access Network
RANAP	RAN Application Part protocol
RB	Radio Bearer
RBR	Re-buffering Ratio
RF	Radio Frequency

RL	Radio Link
RLC	Radio Link Control
RM	Rate Marching
RMA	Rate Matching Attribute
RNAS	RAN Access Server
RNC	Radio Network Controller
RNF	Radio Network Feedback
RNL	Radio Network Layer
RNP	Radio Network Planning
RNS	Radio Network Subsystem
RNSAP	Radio Network Subsystem Application Part protocol
RNTI	Radio Network Temporary Identity
RR	Round Robin, Resource Request
RRC	Radio Resource Control
RRI	Radio Resource Indication
RRM	Radio Resource Manager
RRP	Radio Resource Priority
RSCP	Received Signal Code Power
RT	Real Time
RTP	Real Time Protocol
RTVS	Real Time Video Sharing
S	
SAP	Service Access Point
S-CCPCH	Secondary Common Control Physical Channel
SCH	Synchronisation Channel
SDU	Service Data Unit
SE	Spectral Efficiency
SF	Spreading Factor
SFN	System Frame Number
SGSN	Serving GPRS Support Node
SHO	Soft Handover
SIM	Subscriber Identity Module
SIR	Signal to Interference Ratio
SLA	Service Level Agreement
SLS	Service Level Specification
SM	Session Management
SMS	Short Message Service
SMSC	Serving Mobile Switching Centre
SRB	Signalling Radio Bearer
SRNC	Serving RNC
SS	Supplementary Service
SU	Satisfied Users
SWIS	See What I See
T	
TB	Transport Block
TBS	Transport Block Size
TBSS	Transport Block Set Size
TC	Traffic Class, QoS Class
TCP	Transmission Control Protocol

TDD	Time Division Duplex
TDMA	Time Division Multiple Access
TE	Terminal Equipment, Traffic Engineering
TEID	Tunnel Endpoint IDentifier
TF	Transport Format
TFC	Transport Format Combination
TFCI	Transport Format Combination Indicator
TFCS	Transport Format Combination Set
TFI	Transport Format Indicator
TFRI	Transport Format Resource Indication
TFS	Transport Format Set
TG	Traffic Generator
THP	Traffic Handling Priority
TI	Transaction Identifier
TID	Tunnel Identifier
TM	Transparent Mode
TMSI	Temporary Mobile Subscriber Identity
TNL	Transport Network Layer
TrBk	Transport Block
TrCH	Transport Channel
TrF	Tariff
TTI	Transmission Time Interval
U	
UARFCN	Absolute Radio Frequency Channel Number
UDP	User Datagram Protocol
UE	User Equipment
UEP	Unequal Error Protection
UL	Up Link
UM	Unacknowledged Mode
UMTS	Universal Mobile Telecommunication System
UP	User Plane
URA	UTRAN Registration Area
USIM	UMTS Subscriber Identity Module
UTRA	UMTS Terrestrial Radio Access
UTRAN	UMTS Terrestrial Radio Access Network
V	
VAS	Value Added Service
VBR	Variable Bit Rate
VLR	Visitor Location Register
W	
WAP	Wireless Application Protocol
WARC	World Administrative Radio Conference
WCDMA	Wideband Code Division Multiple Access
WLAN	Wireless Local Area Network
WRC	World Radiocommunication Conference

1

Introduction

UMTS has been designed to support a wide range of applications with different quality of service requirements. The system is intended for long time duration and the modular approach adopted in 3GPP (3rd Generation Partnership Project) provides the necessary flexibility for operators to offer new services to their potential and existing customers. In 3GPP, Quality of Service (QoS) generically refers to the *quality of a requested service as perceived by the user of the service* [1]. However, two aspects need to be considered, namely: The ability of a network to provide such a service with an assured service level, and what the end user really perceives, i.e. how satisfied he or she is with the service, in terms of usability, accessibility, retainability and integrity of the service. (Service *integrity* concerns throughput, delay, delay variation (or jitter) and data loss during the user data transmission; service *accessibility* relates to unavailability, security (authentication and authorisation), activation, access, coverage, blocking, and set-up time of the related bearer service; service *retainability*, in general, characterises the connection losses.) The latter topic, denoted by Quality of Experience (QoE), reflects the collective effect of service performances that determines the degree of satisfaction of the end user. The former aspect, denoted as QoS, relates to the network ability to comply with a service level specification (SLS) resulting from a negotiation between a customer (consumer) and a service provider in a service level agreement (SLA) [2]. In general, a SLA is a formal negotiated contract between two parties that establishes committed levels of network service performance and responsiveness [3]. The two parties may be a consumer and an operator, or two operators where one takes the customer-role buying services from another service provider.

This main goal of this work is QoS management, starting with the broad aim of meeting administrators' requirements for the mechanisms to control access to radio resources in UTRAN (UMTS Terrestrial Radio Access Network) FDD (frequency division duplex). The thesis proposes a comprehensive framework for QoS management that encompasses service planning, provisioning, monitoring and performance improvement across WCDMA radio access networks. This includes tools for service driven radio network dimensioning (initial planning) and detailed radio network planning, QoS management functions in the radio access network elements (algorithms for radio resource management (RRM) with QoS differentiation) and methods for deriving QoS and QoE performance statistics. As a part of this framework, the thesis presents dynamic management solutions for QoS performance improvement, which adapt to changes in traffic conditions and/or operator constraints.

This chapter introduces the challenge of QoS management in UTRAN and explains why each of the covered aspects is important. Besides this, the chapter provides a review of what has been done before in the area (previous work), a description of the new problems solved in this thesis, as well as a summary of the original contributions.

1.1 High Level Problem Definition and Motivation

3GPP defines only a layered bearer service architecture and QoS attributes and leaves the implementation of the actual QoS management functions needed to handle bearer services with specific QoS to vendors' and operators' choice [1]. Besides this, the QoS framework specified in [2] provides only principles and high-level requirements and little work is available in the literature that addresses and resolves the QoS management problem.

Since the design of radio resource management functions has a direct influence on the overall service quality and capital expenditures (CAPEX), the deployed algorithms and related management functions in UTRAN will certainly play a key role in the mature 3G scenarios. Furthermore, due to the system complexity, any practical realisation of the aforementioned functions and radio network plan needs to be validated a priori by means of static or dynamic simulations, depending on the desired level of time resolution and accuracy. In the reviewed literature (see Section 1.2), none of the published resource allocation schemes, tools and theoretical approaches for WCDMA radio network planning showed enough flexibility for an efficient and effective QoS provisioning. The benefit of service differentiation based on the negotiated QoS profile was hardly ever taken into account. Resources (radio spectrum or wireless bandwidth) for packet switched real time services, as streaming or video calls, were always reserved and/or explicitly managed in order to meet application requirements. The presented radio network planning tools were based on circuit switched communications, analysed snapshots of the system status and did not take into consideration the possibility of handling radio resources according to the offered traffic mix characteristics, priorities and QoS. The described dynamic system level simulators offered far too high time resolution and thus lengthy simulation periods for evaluating QoS management functions. This work tries to give more insight into the problem of QoS provisioning by means of service differentiation and peers (users) discrimination across WCDMA radio access networks and provides new models for estimating system performance by means of simulations.

UMTS networks are designed to support service applications with diverse performance requirements. Packet switched person-to-person services, content to person and business services are becoming common public utilities, and as such, must be monitored and provisioned separately to provide quality of service guarantees and ensure that the agreed QoS is sustained. It is neither sufficient nor efficient to reserve network (radio) resources, since QoS degradation is often unavoidable. Any fault, change in traffic mix or volume, or incorrect parameter setting of a network element, may result in the deterioration of a contracted QoS in a SLA or expected performance from a user perspective. Thus, QoS monitoring is required to track the ongoing QoS, detect possible quality deterioration, compare the monitored metrics against the contacted performance and, accordingly, enable operators to fine tune network resources to sustain the contracted QoS. Although delivered service performance is becoming a major aspect of differentiation of service providers, little material, either theoretical or experimental, is available in the literature to describe how essential performance metrics can be measured for a proactive management of radio resource utilisation and service performance. Only the challenges involved in providing QoS distribution monitoring are discussed in the published works (see Section 1.2), and the proposed approaches to meet these challenges are far from any practical cost-effective service assurance solution. In the proposed methods for service assurance, only management models and architecture descriptions of service performance management systems and related capabilities are introduced. Other adopted QoS monitoring solutions are based on drive/walk tests. Furthermore, standardisation forums define QoS attributes and specify what to measure, but not how to monitor and analyse the actual performance of user and/or control plane protocols. This thesis proposes a simple approach that maps service applications with different requirements onto distinct QoS profiles defined by a subset of the bearer attributes, i.e. bit rates, priorities and traffic classes, and derives useful metrics to assess performances of distinct services within the radio access network without any visibility of the content carried by upper layer protocols. The proposed indicators may be also used to estimate what network resources are needed for the traffic to achieve the quality it requires, and, given the existing resources in the network, how much traffic can be carried before the resulting quality degrades excessively.

Ultimately, operators need to have means to improve the performance of the network within a number of constraints and requirements, e.g. expenditures, QoE, changes in traffic and/or service portfolio, revenues, low complexity, effectiveness, efficiency, and resource utilisation. Several methods are available in the literature to improve system performances, yet some meaningful results on QoS optimisation have been hardly ever achieved. The thesis presents a customised genetic approach that performs best in terms of accuracy and speed of convergence. The algorithm is applied to optimise RRM parameters that are differentiated for the different bearer services, but the same for all cells of the analysed network, where statistics are collected. This reduces the computation complexity of evaluating a solution and provides a method to optimise one cell at a time.

1.2 Review of Previous Work

This section summarises major works that paved the way of QoS management in UTRA FDD. Further details on the related literature are given at the beginning of each chapter of the thesis.

1.2.1 QoS Management Framework

QoS requirements for each Traffic Class (TC) in wireless environment and problems in implementing the QoS requirements in UMTS networks were discussed in [4]. In this work, some suggestions to realise QoS requirements in Medium Access Control (MAC) and Link Access Control (LAC) sub layer along with restrictions imposed by WCDMA air interface were presented.

Problems about QoS management in 3G networks, such as architecture of service differentiation, mapping of UMTS QoS to external packet data networks, and the complex characteristics of the wireless "last hop" of the IP network, were addressed in [5]. Methods for improving upper layer protocols performance over the wireless link were also presented in this paper.

1.2.2 Approaches to Modelling Performance in UTRAN

A number of ways for estimating UMTS system performance though simulations were reviewed in [6]. The paper describes different approaches to the embedding of physical layer performance data into wireless WCDMA system simulators. This includes analytical methods, basic and multidimensional approaches for static simulators, long- and short-term models for dynamic simulators, and physical layer models for short-term dynamic models.

Analytical approaches, as e.g. presented in [7]-[9], are invaluable but also very limited in their scope of application. Such techniques can be extremely useful in dimensioning a WCDMA system for circuit switched (CS) traffic, but do not satisfactorily address mixed QoS classes (voice and variable rate packet data). The complexity of packet switched (PS) traffic has effectively prevented any model based approach from being able to describe packet network behaviour accurately. Many models have been proposed and analysed, but there is an inherent trade-off between the predictive power of a model and the ease with which it can be analysed. If a model is simple enough to be solved in the way that Erlang solved the telephony traffic model, it will fail to predict some important features of real packet traffic. If it is rich enough to capture the phenomenology of IP traffic behaviour, it will be impossible to solve the model and use it to calculate the resource requirements of the traffic.

Static simulators (see e.g. [10] and [11]) are based on Monte Carlo approaches and essentially work by "dropping" mobiles in a predefined network layout, and thereafter using some algorithm to decide which proportion of the mobiles would have been correctly served. Then the process is repeated with a number of other drops, where in each case the mobile spatial distribution and numbers correspond to a realisation of a global statistical model of network load. The performance indicators will eventually converge, after which the process can be re run for another set of parameters (e.g. different network load, system layout, etc…). The tools have no concept of time, and therefore cannot directly handle system functions such as admission control (AC) or handover control (HC), though it might be possible to extrapolate useful data related to those functions. In the *basic approach* users in other cells are ignored, or a set of users is simply defined in a given (and well defined) coverage region. The main performance measure is the base station received noise rise, and the interference from other cells has a fixed relationship to that caused by own cell users (given by the well known *i* factor [11]). Because the system has no concept of time-varying effects (such as fading), the required E_b/N_0, for each bearer service type, is an average number for a given set of fading conditions, channel model type, etc. In reality the target E_b/N_0 requirements can be quite different. In the *multidimensional approach* the model is further extended to multi-cells and layer 1 scenarios, which are simulated offline using a link level (physical layer) simulator (for a comprehensive description of the implementation of an UTRA FDD simulator see e.g. [12]). This approach can account for variations in channel multipath, in-cell and out-of-cell interference variation and different mobile speeds.

In dynamic simulators, a population of moving mobiles is typically generated according to some traffic statistics and mobility models. Call activity is generated according to another set of models, and statistics are collected during the simulation. A run could be quite long, but a priori there is no need to start again from new initial conditions.

Long-term dynamic models can be seen as an evolution of the static models based on Monte Carlo procedures. Once such a model is constructed, it can be extended by considering a single large drop of mobiles, where each mobile has two important time-related characteristics, movement and call activity, each requiring some modelling. Long-term dynamic models operate in a relatively large time steps, from one up to ten seconds. Just as in the static models, there is no explicit signal level fading and therefore no explicit power control. Every time step, an iterative power-balancing algorithm needs to be triggered. Enhanced versions of the simulator described in [10], were presented in [13] and [14], where a time-driven simulation engine with nested loops was added. This engine requires the recalculation of the mobile positions, generating call originating and clearing events, recalculating path losses and cell ownership, and determining the E_b/N_0 achieved by means of iterations, at the regular time intervals.

Short-term dynamic models have a fundamentally different split between physical layer and system level functions. The inner (fast) and outer loop power control is run for each mobile station. The interference environment is changing on a slot-by-slot basis, due to high dynamic processes, such as: Own signal and own cell signal fading, other cell / other mobile fading, change of fingers being used in the RAKE process, and instantaneous changes in cell loading due to scheduling and voice/data activity spurts. Such a model provides a much better platform to assess in detail the dynamic performance of the system, at the cost of high computational complexity. The physical layer is modelled by mapping series of E_b/N_0 values over the transmission time interval (TTI or interleaving period) onto the expected block error rate (BLER). The look-up tables are usually built using physical layer simulations (see e.g. [15]). An example of advanced dynamic simulator (denoted as *Wallu* in this thesis) for an effective analysis of radio resource management algorithms in UTRAN was described in [16]. The evaluation of the feasibility and accuracy of such a tool, as well as a discussion thereof, can be found in [17].

1.2.3 Functions in UTRAN for QoS Provisioning

The QoS management functions for UMTS bearer service in the control plane support the establishment and modification of a UMTS bearer service, which includes all aspects to enable the provision of a contracted QoS (required service characteristics). The QoS management functions in the user plane ensure the provision of the QoS negotiated for a UMTS bearer service, and maintain the data transfer characteristics according to the commitments (or bearer service attributes) established by the UMTS bearer service control functions [1].

Various schemes (RRM algorithms) for an effective radio resource control in WCDMA networks to "optimally" support multimedia services can be found in the literature.

A comprehensive resource allocation framework based on differentiated QoS classes was presented in [18]-[21]. Prioritised resource control is based on assigning different priority levels to differentiate the traffic classes according to their QoS requirements. Class priority is thus enforced at call admission control and during scheduling of non real time (NRT) data. New connection requests from the service with highest priority (conversational RT service, on Conversational class) are admitted onto the system if there is enough power remaining in the base station power budget to compensate for the estimated path loss by the mobile unit on an open loop basis, and the individual dedicated traffic channel power limit is not violated. The second priority service (streaming RT service, on Streaming class) must meet the same constraints but the notion of link quality is added in the admission process. Active connections, in the scheduling queue, are served on the basis of a Round Robin (RR) policy with interactive users (on Interactive traffic class) having preference over the Background class users. The users are scheduled only if their link quality is sufficient and the impact on higher order services link quality is acceptable. The analysis indicates that service prioritisation provides enhanced flexibility and control in meeting QoS constraints and gives better performance in the examined mixed services scenarios. In [21], it is also shown that rate adaptation on the downlink shared channel (DSCH), combined with frame based round robin scheduling, results in significant performance improvement not only for NRT service classes, but also benefits RT applications. Capacity gains close to 50% can be achieved while providing system stability (otherwise missing in the case of fixed rate) and much flexibility in the design of WCDMA systems aiming at a multimedia service offer.

A priority based dynamic radio resource allocation algorithm that reduces the call drop ratio during handover in UTRAN was proposed in [22]. The proposed scheme increases the system utilisation and decreases the call drop rate at the expense of the service quality of lower priority users.

Several quality based AC policies and scheduling schemes for WCMDA radio access networks were investigated in [23]-[25]. In [23], to differentiate traffic classes, and deal with them correspondingly, a priority ranking is decided for each class. The admission control is then based on the assigned rank and throughput estimates assuming perfect power control. For delay tolerable classes of traffic, a QoS renegotiation procedure is introduced, resulting in a lower blocking probability for overall traffic. The QoS renegotiation is based on the assumption that if the users cannot acquire the necessary resources in order to obtain their highest QoS level, they are willing to accept an admission at a lower service level, rather then being blocked. In [24], two adaptive admission control algorithms that aim at offering the requested QoS in downlink by acting on transmission power thresholds are proposed. The goal of the adaptive AC schemes is to provide the desired downlink BLER at least to a percentage of real time services, through the assignment of suitable radio link powers. In the study, it is assumed that by limiting the number of active radio links (both for RT and NRT traffic) the required RT quality can be granted to an arbitrary percentage of users. Similar algorithms could be applied to provide the desired quality to NRT traffic by controlling the number of admitted sessions. In [25], several resource allocation issues were considered: Hybrid Multiple Access (HMA) transmission strategy for RT and NRT traffic, power allocation for users with different QoS classes, an admission control policy based on QoS fulfilment, and a scheduling scheme based on QoS requirements. The first issue is residing in the physical layer, the second and third issues are working on the radio resource control (RRC) layer, and the fourth issue is considered for the MAC layer. In the AC policy four first in first out (FIFO) queues are considered, i.e. for RT handoff calls, RT new calls, NRT handoff calls and NRT new arrival calls. The packet scheduling (PS) consists of an earliest deadline first (EDF) based scheduler for RT traffic, and a priority queue based scheduler for NRT traffic. Three queues are implemented: EDF high priority queue for RT packets, medium priority queue for Interactive class packets, and low priority queue for Background class. Within each NRT queue, packets are sorted based on the corresponding bit error rate (BER) requirements. A decision scheme and a cost function for choosing uni-access or multi-access transmission are proposed. Multi-access is the scheme that all users can transmit on the channel at the same time and uni-access is the scheme that only one user is permitted to transmit on the channel at one time [26].

One of the goals of 3G mobile communication systems is the delivery of multimedia services to the mobile user. The use of several different services (parallel data flows) at the same time raises the demands for mechanisms to guarantee QoS for the used applications. In [27] and [28], the mapping of logical channels to appropriate combination of transport channels (TrCHs), which are then multiplexed on a code composite transport channel (CCTrCH) at the physical layer, is achieved by using a MAC scheduling algorithm with dynamic channel type switching and Transport Format (TF) selection in accordance to the service requirements.

A new method to adapt the QoS in WCDMA networks, called radio network feedback (RNF), was presented in [29]. The proposed concept is in general applicable to all the services requiring a minimum guaranteed quality (i.e., non best effort). It is shown how RNF makes it possible for a streaming server to adapt its source bit rate to a WCDMA radio link, whose bandwidth may vary in time, for example, due to decongestion/ congestion situations over the radio interface or to handover. The bandwidth (i.e., the quality) is increased when possible and decreased (instead of just dropping the service) when needed. RNF was compared with client-based adaptation solutions. In [29], simulation results show that RNF is fast and accurate and performs better than client-based adaptation.

1.2.4 QoS Monitoring in UTRAN

QoS monitoring is required to measure the service quality, compare the measured metrics against the contacted performance and, accordingly, enable operators to fine tune parameter settings to sustain the delivered QoS.

Measurements for service accessibility and retainability are based on the success/failure of procedures needed to setup, modify or maintaining a certain bearer service or signalling connection. Some measurement types that are specific to UMTS or combined UMTS/GSM networks are defined in [30].

Definitions and procedures to be used for statistical calculations that are related to QoE measurements in mobile communications networks, especially GSM and 3G networks, are described in [31]-[37] (denoted by *Part 1-7* in this paragraph). QoS measurements and related post-processing are only marginally covered in these recommendations. All the defined quality of service parameters and their computations are based on field measurements. That indicates that the measurements were made from customers' point of view (full end-to-end perspective, taking into account the needs of testing). In particular, *Part 1* identifies QoE aspects for popular services in GSM/EGPRS and 3G networks. For each service chosen, QoE indicators are listed. They are considered to be suitable for the quantitative characterisation of the dominant technical QoE aspects as experienced from the end-customer perspective. *Part 2* defines QoE parameters and their computation for popular services in GSM and 3G networks. The technical QoE indicators, listed in [31], are the basis for the parameter set chosen. The parameter definition is split into two parts: Abstract definition and the generic description of the measurement method with the respective trigger points. Only measurement methods not dependent on any infrastructure provided are described in [32]. The harmonised definitions given in [32] are considered as the prerequisites for comparison of QoE measurements and measurement results. *Part 3* describes typical procedures used for QoE measurements over GSM, along with settings and parameters for such measurements. *Part 4* defines the minimum requirements of QoE measurement equipment for GSM and 3G networks in the way that the values and trigger-points needed to compute the QoE parameter as defined in [32] can be measured following the procedures defined in [33]. Test-equipment fulfilling the specified minimum requirements, will allow performing the proposed measurements in a reliable and reproducible way. *Part 5* specifies test profiles that are required to enable benchmarking of different GSM or 3G networks both within and outside national boundaries. *Part 6* describes procedures to be used for statistical calculations in the field of QoE measurement of GSM and 3G network using probing systems. *Part 7* describes the field measurement method procedures used for QoE measurements over GSM where the results are obtained applying inferential statistics.

Some insights into the QoS performance monitoring issues were presented in [38]-[40]. In [38], the challenges involved in providing QoS distribution monitoring were discussed, and some methods to meet these challenges were proposed. A realistic approach to end-to-end service assurance on the mobile Internet access was presented in [39]. As a part of this framework, a management model and architecture of the service performance management system and related capabilities were introduced. In [40], an interesting performance evaluation of GPRS network accomplished through drive tests was presented.

1.2.5 QoS "Optimisation"

In this work, the service optimisation is presented as a process to optimise the spectral efficiency at a given QoE (percentage of satisfied users of the offered services) by choosing the best values for a selected set of crucial parameters. Traditionally, technical experts, who have access to live networks, perform the network optimisation. In order to atomise the optimisation process, several methods are available in the literature to search the solution within a part of the possible combinations of network configurations; some of those were presented in [41]-[50].

In [41]-[43], the issues that arise in the definition and deployment of neighbouring cells for handovers in UMTS networks were examined. The papers provide a description of the algorithm adopted for automatic generation of neighbour cell lists and related performance monitoring and configuration management aspects. Performance results show the proposed approaches to be feasible solutions for cellular network troubleshooting and "optimisation". The system performance was improved and the average length of the neighbour cell lists shortened. Costs related to drive or walk tests were considerably reduced.

In [44] and [45], the focus is on the radio coverage problem of UMTS networks, i.e. to cover a maximum surface of a given geographical region at low costs. This combinatorial optimisation problem was solved with a bio-inspired genetic algorithm. The experiments and simulations exhibit promising results and the ability to adapt to different problems and criteria of genetic algorithms.

Ultimately, in [46]-[50], several methods for assessing the performance of user data transfer and parameters tuning of RRM functions in UTRAN by means of a cost function were proposed. Simulation results show the described approaches to improve the overall network performance in comparison to default parameter values. The attained capacity gains were up to 20% with acceptable quality of experience.

1.3 Detailed Problem Definition

The new problems solved in this monograph, in close connection to current research within the cellular industry, are presented with a list of questions covering the most crucial aspects of QoS management in UTRAN.

- *QoS provisioning*: Using dedicated transport channels, how is it possible to provide diverse treatment to distinct bearers or users of the same service in UTRAN, according to service specific characteristics and performance requirements? How the basic wide band power based RRM functions presented in e.g. [11] and [17], such as AC, PS, LC and HC, can be enhanced to support such discrimination, in order to preserve, maintain and control the QoE of the most demanding users or applications? How a robust, effective (capable of exploiting the desired power budget), efficient (fast in bit rate scheduling), yet simple (easy to implement in a real network), DCH allocation scheme with QoS differentiation, fair in bit rates or resources allocation, may be designed to preserve the quality of RT services with no bit rate guarantees?
- *QoS planning*: It is know that the accuracy of the modelling of a wireless WCDMA system depends heavily on the choices made for the physical layer model. The solution to the problem has to satisfy two contradictory requirements of being of reasonable accuracy and complexity. The meaning of the term "reasonable" depends on the use of the data delivered by the system simulations, and the computational time (power) of the intended user. What approach to the embedding of physical layer performance data into WCDMA system level simulators would make it possible to implement QoS management functions for UTRAN and specific characteristics of the offered services, without achieving the complexity of short-term dynamic models? Such a solution should be feasible for service driven radio network planning and for predicting the performance of RRM algorithms before their deployment in real networks. Beside this, how such a prediction method could be simplified for just solving radio interface dimensioning issues that may arise during the deployment of new services in WCDMA networks in mixed QoS classes of voice and data?
- *QoS monitoring*: How may be made accessibility, retainability and integrity measures in UTRAN to determine the degree of satisfaction of the users of the deployed services, without making use of drive/walk tests or protocol analysers at the different interfaces? How may be measured the spectral efficiency of the radio interface in UTRAN? In the uplink direction, how may be determined cell based values for the E_b/N_0, BLER and BER to a selected DCH multiplexed with more transport channels to a dedicated physical channel (DPCH), which is influenced by signals transmitted on all active DCHs multiplexed to the DPCH?
- *QoS "optimisation"*: Is there any need (business opportunity, added value for the network operator) to adapt the UTRAN configuration to diverse traffic scenarios, in order to maximise the system performance at a given QoE? In other words, do good default parameter settings need to be changed for different traffic mixes? In the case there is a need to tune several parameters simultaneously, what is a practically

feasible methodology for improving the spectral efficiency (system load at a given QoE)? In turn, how this could be then put in practice, i.e. in real networks? What tool can be efficiently used for testing the proposed optimisation process (solution) before deploying it in real networks?

1.4 Original Contributions

The overall goal of the thesis is to provide new models for service driven radio interface dimensioning, and service driven detailed radio network planning. The tools are also used for estimating the performance of RRM functions in UTRA FDD networks. Beside this, the thesis describes new algorithms for differentiated service treatment and measure metrics for QoS and QoE monitoring in UTRAN. As a part of this framework a genetic approach to spectral efficiency improvement is also proposed. In particular, the major contributions of this work are the following:

- Plain methods and a basic approach (simulator) for the first and most rapid estimation of the number of users of a set of services a cell can serve with a predefined level of quality. The proposed solution, based on the knowledge of downlink orthogonality and interference, target load, E_b/N_0 values, activity factors, and traffic mix characteristics, provides quantitative answers to how many users of mixed QoS classes a WCDMA cell can satisfactorily accommodate. Besides this, the proposed approach can be used for addressing the radio interface dimensioning issue that arises from the introduction of novel services in WCDMA networks. The simulator (based on throughput estimates and snapshots of the system status) supports essential radio resource management functions, as admission control and packet scheduler with QoS differentiation. The tool presented in this thesis for dedicated channels can be enhanced to support all other transport channels, in uplink and downlink directions. Some of the performance results collected in this dissertation were published in [51].
- A virtual time simulator for service driven radio network planning. The tool allows network planners to find a good trade-off between quality constraints, capacity and coverage criteria for all services in operators' service portfolio. The proposed approach overcomes the limitations and complexity of static and dynamic system simulators. Besides this, the described solution provides an appropriate platform to assess the performance of QoS management functions in UTRAN before their deployment in real networks. The simulator described in this dissertation for dedicated channels and static terminals can be enhanced to support mobility and high-speed packet access concepts. A part of the results collected in this dissertation were presented in [52]-[53].
- Radio resource management functions in UTRAN, such as AC, LC, PS and HC, with bearer services (dedicated transport channels) differentiation. The proposed algorithms enable operators to handle the requested radio resources according to the offered service characteristics and priorities. In the AC, PS, LC and HC queues, resource requests are served based on priorities and at a give priority based on their arrival time. For guaranteed bit rate (GB) traffic, the AC algorithm determines for each bearer service request (or modification) whether the required resources can be provided and these resources are reserved if allocated to the bearer service in question. For non-guaranteed bit rate (NGB) services, the provision and maintenance of the user data transfer characteristics (according to the specific QoS requirements) is provided by

differentiated parameters settings and capacity (bit rate) requests prioritisation based on experienced throughput and delays. As a part of this framework, two scheduling algorithms, Fair Throughput (FT) and Fair Resources (FR), for MAC or RRC layer, are proposed. Some of the results reported in this dissertation were presented in [54]-[57].

- This thesis describes how QoS accessibility, retainability and integrity measurements, such as throughput, erroneous data blocks, transfer delays and delay variations can be measured in the radio network subsystem (RNS). In the uplink, this includes a solution that avoids the influence of the static rate matching in the DCH measured performance. It is also shown how to assess performance of service applications through the radio interface (without disclosing upper layer protocols carrying content) by means of an appropriate classification of counters based on a sub set of bearer service attributes. The target of this contribution is to provide means and methods for operators to monitor separately, and in a cost-effective way, service applications, without tracing or explicitly disclosing the characteristics of upper layer protocols by means of any tool or protocol entity. The proposed performance indicators provide essential inputs to ensure quality compliance to service layer management commitments. The described performance metrics can be also used to measure how much bandwidth packet traffic needs to meet a statistical service guarantee, how well packet traffic multiplex (statistical multiplexing gain derivation), and how QoS mechanisms can be configured to provide application quality efficiently. A part of the performance measures proposed in this dissertation were successfully implemented in EGPRS and UTRA FDD Networks [58]-[59].
- A genetic approach to QoS optimisation for WCDMA mobile networks. The solution space search algorithm and corresponding QoE fitness function are proposed for tuning simultaneously, within a reasonable time interval, several parameters affecting the QoS. Some of the results reported in this dissertation were published in [60]-[61].

Ultimately, the adopted research methodologies are not limited to UTRA FDD networks and the use cases presented in this work. Yet, the proposed approaches can be applied to a vast range of access networks supporting QoS and similar stacks of protocols.

1.5 Outline of the Thesis

The rest of the dissertation is organised as follows. Chapter 2 reviews the UMTS concepts, architectures, interfaces and layered functions. Chapter 3 reports the main assumptions and research methodologies adopted in this monograph. In this chapter, the simulators for studying QoS management functions and for service driven WCDMA radio interface dimensioning and detailed radio network planning are described. In Chapter 4, the radio resource management functions with QoS differentiation are described and validated by means of simulations. In Chapter 5, the potential and feasibility of the proposed tools are shown by means of several case studies. Chapter 6 defines a cost-effective method for QoS monitoring in WCDMA networks and data analysis. The chapter provides key performance metrics for assessing service integrity, accessibility and retainability in UL and DL directions. Chapter 7 introduces the challenge of QoS “optimisation” and presents a genetic approach to find “optimal” settings of the differentiated parameters introduced in Chapter 3 and 4. Chapter 8 concludes this work with a summary of the key issues and outlines some directions for further research in WCDMA Evolved Radio Access Networks.

2

UMTS Overview

The content of this chapter is prevalently based on the 3GPP Release 5 specifications, and it is aimed at giving an overview of the UMTS architecture, interfaces and system functions. The UMTS QoS concept and architecture are also presented. As a part of this framework, the basic concepts of Access Stratum (AS) and Non-Access Stratum (NAS) are introduced in the following sections. In addition, the roles of UTRAN and radio interface protocols are pointed out.

The high level functional grouping of layer 1-3 protocols allows the reader to get a clear view of the protocol architecture and the transfer of a specific type of information over the air (radio) interface on common, shared or dedicated resources.

In this thesis, the models and algorithms presented in Chapter 3, 4 and corresponding use cases investigated in Chapter 5 apply to dedicated channels, whereas the scope of the proposed performance measures in Section 3.6 and in Chapter 6, as well as the genetic approach to spectral efficiency improvement described in Chapter 7 is wider. The proposed solutions are valid also for common and (high-speed) shared channels (see Section 2.4).

2.1 Introduction

Analogue cellular networks are commonly referred to first generation systems. The digital system currently in use, such as GSM, PDC, cdmaOne (IS-95), and US-TDMA (IS-136), are second generation systems. These networks have enabled voice communications to go wireless in many of the leading markets, and customers are increasingly finding value also in other services, such as text messages and access to data networks, which are growing rapidly.

Third generation systems are designed for multimedia communications. With them person-to-person communication can be enhanced with high quality images and video, and access to information and services on public and private networks will be enhanced by the higher data rates and new flexible communication capabilities of 3rd generation systems.

This, together with the continuous evolution of the 2G systems, will create new business opportunities not only for network infrastructure vendors and operators, but also for the providers of content and applications using these networks.

In the standardisation forums, WCDMA technology has emerged as the most widely adopted 3rd generation radio interface. Its specification has been created in 3GPP, which is the joint project of the standardisation bodies from Europe, Japan, Korea, USA and China. Within 3GPP, WCDMA is called Universal Terrestrial Radio Access (UTRA) Frequency Division Duplex (FDD) and Time Division Duplex (TDD), the name WCDMA being used to cover both FDD and TDD operations. Through this thesis, the term WCDMA relates only to UTRA FDD, since TDD is not within the scope of this work.

The spectrum allocation in Europe, Japan and USA is shown in Figure 2.1. In Europe the International Mobile Telephony IMT-2000 (or Word Administrative Radio Conference WARC-92) bands of 2 x 60 MHz (1920-1980 MHz plus 2110-2170 MHz) are available for WCDMA FDD. FDD systems use different frequency bands for uplink and for downlink, separated by the duplex distance, while TDD systems utilise the same frequency for both directions.

At the ITU-R WRC-2000 in May 2000 other frequency bands were also identified for IMS-2000. The main new spectrum in Europe for IMS-2000 will be 2500-2690 MHz. The duplex arrangement of that spectrum is still under discussion.

In Europe and Japan, the actual number of 3G operators per country is between three and six, and the number of FDD carriers (2 x 5 MHz) per operator is from two to three.

The remainder of this chapter is organised as follows: Section 2.2 introduces the UMTS architecture. UMTS protocols thereof are described in Section 2.3. Transport and physical channels are defined in Section 2.4. Section 2.5 describes the QoS concept and architecture as defined in 3GPP R5 specifications. Requirements on QoS management functions in the network are summarised in Section 2.6. Section 2.7 presents the UMTS and corresponding radio access bearer service attributes. The packet data transfer across UMTS networks is explained in Section 2.8.

2.2 UMTS Architecture

In 3GPP the UMTS architecture is described in terms of its entities – User Equipment (UE), UTRAN (UMTS Terrestrial Radio Access Network) and Core Network (CN). The radio interface (U_u) and the CN-UTRAN interface (I_u) are the reference points between the subsystems.

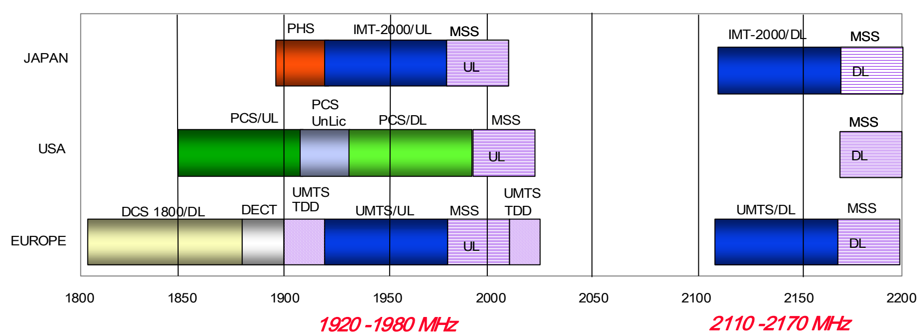


Figure 2.1. 2GHz band spectrum allocation in Europe, Japan and USA (MSS = mobile satellite spectrum).

The protocols over U_u and I_u interfaces are divided into two structures – User Plane (UP) protocols, i.e. the protocols implementing the actual Radio Access Bearer (RAB) service; and Control Plane (CP) protocols, i.e. the protocols for controlling radio access bearers and the connection between the UE and CN. Both Radio and I_u protocols provide transparent transfer of Non-Access Stratum (NAS) messages between the UE and CN [62].

The high-level functional grouping into Access Stratum (AS) and Non-Access Stratum defined in [63] is depicted in Figure 2.2. The Access Stratum is the functional grouping of protocols specific to the access technique. It includes protocols for supporting transfer of radio-related information, for coordinating the use of radio resources between UE and access network, and protocols for supporting the access from the serving network to the resources provided by the access network. The Access Stratum offers services through Service Access Points (SAP) to the Non-Access Stratum (CN related signalling and services), i.e. provides the Access Link between UE and CN, which consists of one or more independent and simultaneous UE-CN radio access bearer services, and only one signalling connection between the upper layer entities of UE and CN. The signalling connection consists of two parts: Radio Resource Control (RRC) connection and the I_u connection, which expands the RRC signalling connection towards the CN. The Non-Access Stratum is the functional grouping of protocols aimed at – Call Control (CC), for circuit switched (CS) voice and data; Session Management (SM), for packet switched (PS) data; Mobility Management (MM, GMM), for circuit switched and packet switched domains; Short Message Services (SMS), for packet and circuit switched domain; Supplementary Services (SS) and RAB management, for re-establishment of RAB(s) which still have active PDP (Packet Data Protocol) contexts.

The radio access bearer (RAB) is a service provided by the Access Stratum to the Non-Access Stratum in order to transfer user data between UE and CN; a bearer is described by a set of parameters (attributes), which define that particular traffic aspect or Quality of Service profile of that particular application (or service). The QoS concept and architecture used in UMTS, i.e. the list of attributes applicable to UMTS Bearer Service and Radio Access Bearer Service are discussed in Section 2.7.

The UMTS logical architecture is depicted in Figure 2.3. UTRAN consists of one or more Radio Network Subsystems (RNS). The RNS includes one Radio Network Controller (RNC) and multiple base stations (or Node B). The Controlling RNC (CRNC) controls one Node B, i.e. load and congestion control of its own cells, executes the admission control and code allocation (RM) for new radio links to be established in those cells. There is one CRNC for each Node B. The Serving RNC (SRNC) for one UE terminates the I_u link for the transport of user data, the corresponding Radio Access Network Application Protocol (RANAP) signalling to/from the CN and the RRC signalling UE/UTRAN. It performs the L2 processing to/from the radio interface and the basic RRM operations. One UE connected to UTRAN has only one SRNC and one RRC connection. The Drift RNC (DRNC) is any RNC other than SRNC that controls cells used by mobile stations. If needed the DRNC may perform macro-diversity combining and splitting. Except when the UE is using a common or shared transport channel (TrCH), the DRNC does not perform L2 processing of the user data, but it routes the data transparently between the I_{ub} and I_{ur} interfaces. One UE may have several DRNCs. The Node B (NB) performs the air interface L1 processing (channel coding and interleaving, rate adaptation, TrCHs multiplexing, spreading, scrambling, etc…).

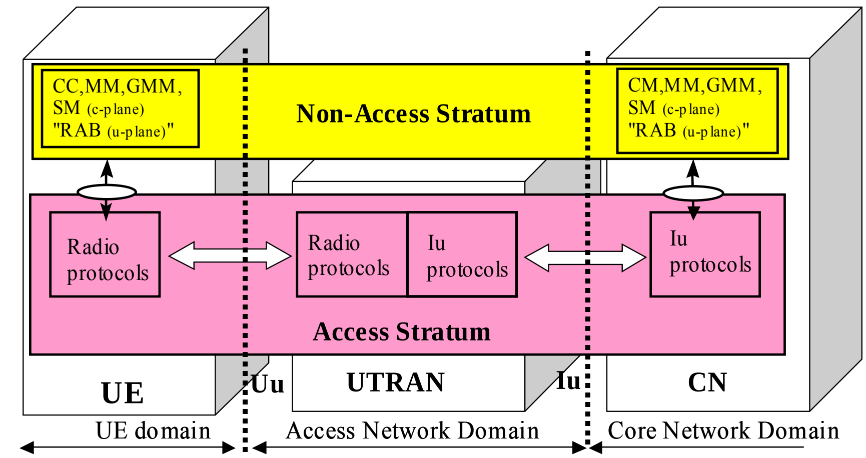


Figure 2.2. High-level functional grouping into Access Stratum and Non-Access Stratum.

A Node B can have multiple cells (from one to six, normally); a "cell" is defined by a cell identification (CID), timing delay (T_Cell), UTRA Absolute Radio Frequency Channel Number (UARFCN), maximum transmission power, and primary scrambling code.

The Core Network (CN) maps the E2E Quality of Service requirements to the UMTS bearer service. The QoS requirements are also mapped onto the available external bearer service by the Gateway (GMSC, GGSN) of the UMTS CN. The Serving MSC/VLR is responsible for CS connection management (CM = call control, CC) activities, mobility management (MM) related issues, such as location update, location registration, paging and security activities. It contains the transcoders required for speech coding conversion (in 3G is a part of the CN). The Gateway MSC takes care of the incoming/outgoing connections to/from other networks. From the CM point of view, the GMSC establishes a call path towards the serving MSC/VLR under which the addressed subscriber is to be found. From the MM point of view, the GMSC initiates a location info retrieval procedure, whose aim is to find the correct serving MSC/VLR for call path connection. In the PS domain packet connection are called sessions and they are established and managed by a function called Session Management (SM). The Serving GPRS Support Node (SGSN) supports packet communication towards the access network. The SGSN is mainly responsible for SM and MM related issues, such as routing area update, location registration, packet paging and controlling the security mechanisms related to the packet communication. The Gateway GPRS Support Node (GGSN) maintains the connection towards the other packet switched networks such as Internet. From the CN point of view, this node is responsible for the MM related issues like the GMSC in the CN CS domain. The session management responsibility is also located in the GGSN.

The IP Backbone (IP-BB) is the transport network connecting GPRS Support Nodes (GSN) together, which can be regarded as a private "intranet". This is why the IP backbone is actually separated from other networks by firewall functionality. For IP backbone routing the PS domain must contain a Domain Name Server (DNS). With this node SGSNs and GGSNs are able to perform routing and actually the GGSN and SGSN may belong to different networks in this respect.

The Registers contains the Home Location Register (HLR), Authentication Register (AuC) and Equipment Identity Register (EIR). This part of the CN does not deliver traffic. Instead it contains the addressing and identity information for both the CS and PS domains, which is required, for instance, for MM related procedures. The HLR contains permanent data of the subscribers. One subscriber can only be registered into one and only one HLR. The HLR is responsible for MM related procedures. In 3G the Visitor Location Register (VLR) is considered to be an integral part of the serving MSC. The VLR participates in MM related procedures like location update, location registration, paging and security activities. The VLR database contains temporary copies of the active subscribers, which have performed the location update in the VLR area.

The Network Management System (NMS) controls and monitors the entire network. In general, network management can be perceived as a service that employs a variety of methods and tools, applications and devices to enable the network operator to monitor and maintain the entire network.

More information on the UMTS architecture can be found in [64].

2.3 UMTS Protocols

This section gives an interface-centric view of the UMTS network by focusing on system protocols. UMTS protocols are used to control the execution of network functions in a coordinated manner across system interfaces. Particular attention will be paid on protocol reference models of UTRAN and radio interface. For CN related issues see e.g. [64].

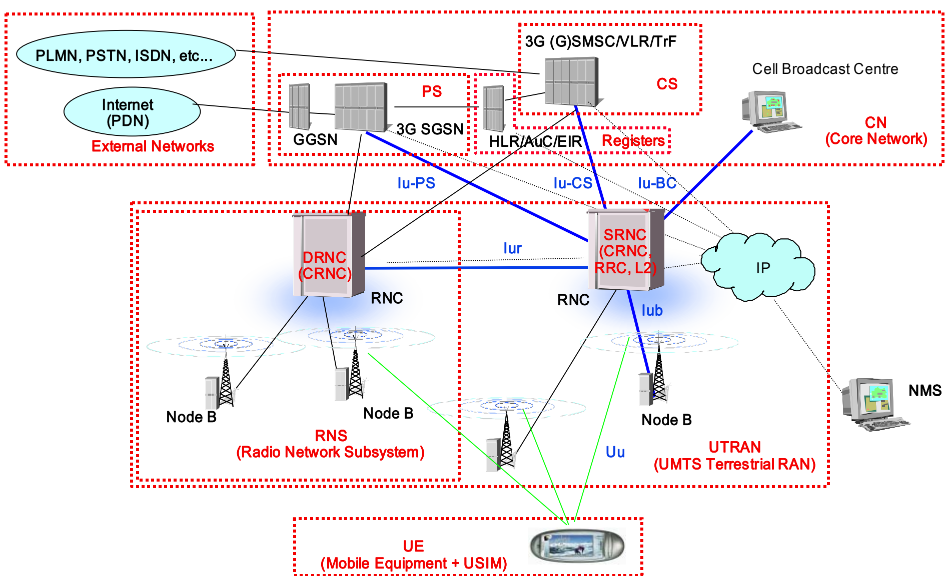


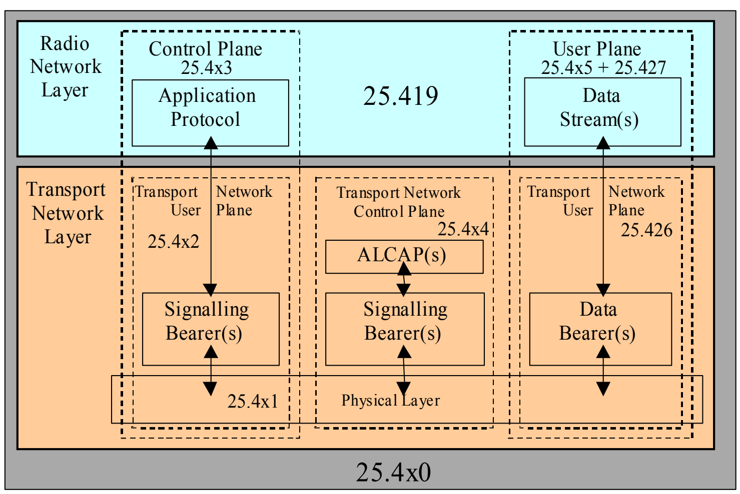
Figure 2.3. UMTS Logical Architecture.

2.3.1 UTRAN Protocol Reference Model

The general protocol model for UTRAN interfaces is depicted in Figure 2.4. The structure is based on the principle that the layers and planes are logically independent of each other, so that when required the standardisation body can easily alter protocol stacks and planes to fit future requirements. The model is generic in the sense that the architectural elements – though, not the specific protocols – are the same for all the UTRAN interfaces: I_u, I_{ub}, and I_{ur}. The specific protocols are referred to in Figure 2.4 only using their 3GPP specification numbers, which also underline the idea of sharing common protocol architecture.

The Protocol Structure consists of two main layers, denoted as Radio Network Layer (RNL), and Transport Network Layer (TNL). All UTRAN related issues, such as control, management and utilisation of radio access bearers across the various UTRAN interfaces, consistently with the radio interface protocol model described in Section 2.3.2, are handled only in the Radio Network Layer. The Transport Network Layer represents standard transport technology, which can be selected without any UTRAN specific requirements.

In the Radio Network Layer, the Control Plane includes the Application Protocol (or Application Part, AP), i.e. RANAP, RNSAP or NBAP, and the Signalling Bearer for transporting the Application Protocol messages. Among other things, the Application Protocol is used for setting up bearers for (i.e. Radio Access Bearer or Radio Link) in the Radio Network Layer. The Signalling Bearer for the Application Protocol may or may not be of the same type as the Signalling Protocol for the Access Link Control Application Part (ALCAP). The Signalling Bearer is set up by O&M actions. The User Plane includes the Data Stream(s) and the Data Bearer(s) for the Data Stream(s). The Data Stream(s) is/are characterised by one or more frame protocols (FPs) specified for that interface.



Legend 3GPP Specifications
x = 1 Iu protocols
x = 2 Iur protocols
x = 3 Iub protocols

Figure 2.4. General Protocol Model for UTRAN Interfaces.

The UTRAN AP and the frame protocols together form the UMTS access stratum, which cover all those communication aspects that are dependent on the selected radio access technology. Within generic UMTS radio access bearers; NAS protocols are then used for direct transfer of signalling and transparent flow of user data frames between UEs and the CN. This is achieved by encapsulation of the upper layer payload into UTRAN protocol messages.

The Transport Network Control Plane does not include any Radio Network Layer information, and is completely in the Transport Layer. It includes the ALCAP protocol(s) that is/are needed to set up the transport bearers (Data Bearer) for the User Plane. It also includes the appropriate Signalling Bearer(s) needed for the ALCAP protocol(s). The Transport Network Control Plane is a plane that acts between the Control Plane and the User Plane. The introduction of Transport Network Control Plane is performed in a way that the Application Protocol in the Radio Network Control Plane is kept completely independent of the technology selected for Data Bearer in the User Plane. ALCAP might not be used for all types Data Bearers. If there is no ALCAP signalling transaction, the Transport Network Control Plane is not needed at all. This is the case when pre-configured Data Bearers are used or when the IP UTRAN option is used between two IP UTRAN nodes or between an IP UTRAN node and an IP CN node. When Transport Network Control Plane is used, the transport bearers for the Data Bearer in the User Plane are set up through a signalling transaction by the Application Protocol in the Control Plane, which triggers the set up of the Data Bearer by the ALCAP protocol that is specific for the User Plane technology. The Data Bearer(s) in the User Plane, and the Signalling Bearer(s) for Application Protocol, belong also to the Transport Network User Plane. The Data Bearers in Transport Network User Plane are directly controlled by Transport Network Control Plane during real time operation, but the control actions required for setting up the Signalling Bearer(s) for Application Protocol are O&M actions.

The UTRAN protocol architecture is discussed in more detail in [17], [62], and [64].

2.3.2 Radio Interface Protocol Architecture and Logical Channels

The radio interface protocols are needed to set up, reconfigure and release the Radio Bearer services. The radio interface consists of three protocol layers – the physical layer (L1); the data link layer (L2); and the network layer (L3). Layer 2 contains the following sub layers – Medium Access Control (MAC), Radio Link Control (RLC), Packet Data Convergence Protocol (PDCP) and Broadcast/Multicast Control (BMC). RLC is divided into Control (C) and User (U) planes, whilst PDCP and BMC exist only in the U-plane. Layer 3 consists of one protocol, denoted Radio Resource Control (RRC), which belongs to the C-plane [63].

2.3.2.1 Radio Interface Protocols Architecture

The radio interface protocol architecture and the connections between protocols are shown in Figure 2.5. Each block represents an instance of the corresponding protocol. The dashed lines represent the control interfaces through which the RRC protocol controls and configures the lower layers. The service access points between MAC and physical layer and between RLC and MAC sub layers provide the *transport channels* (TrCHs) and the *logical channels* (LoCHs), respectively. The TrCHs are specified for data transport between physical layer and Layer 2 peer entities, whereas logical channels define the transfer of a specific type of information over the radio interface.

2.3.2.2 Medium Access Control Protocol

MAC is responsible for mapping of logical channels onto appropriate transport channels. MAC provides an efficient use of transport channels; based on the instantaneous source rate(s), it selects the appropriate transport format (TF) within an assigned transport format set (TFS) for each active TrCH. The transport format is selected based on the transport format combination set (TFCS), which is assigned by the RRC protocol and produced by the admission control in the RNC for each connection, when a radio access bearer is set up or modified. The functionality of the MAC layer includes – priority handling between data flows of one connection (selection of a TFC for which high priority data is mapped onto L1 with a "high bit rate" TF); priority handling between UE(s) by means of dynamic scheduling (MAC realises priority handling on common and shared TrCHs); identification of UEs on common transport channels (in band identification). MAC provides: Multiplexing and de-multiplexing function of RLC PDUs into and from TBs delivered to and from the physical layer on common channels (services or better LoCHs multiplexing for CCHs); multiplexing and de-multiplexing function of RLC PDUs into and from TBSs delivered to and from the physical layer on dedicated channels (services or better LoCHs multiplexing for DCHs). MAC is further responsible for: Traffic volume monitoring (measurement of traffic volume on logical channels and reporting to RRC, based on which the RRC performs TrCH switching decisions); dynamic TrCH type switching (execution of the switching between common and dedicated transport channels); Ciphering (for transparent RLC mode); Access Service Class (ASC) selection for uplink common channels transmission; and HARQ functionality for HS-DSCH transmission, in-sequence delivery and assembly-disassembly of higher layer PDUs on HS-DSCH [65].

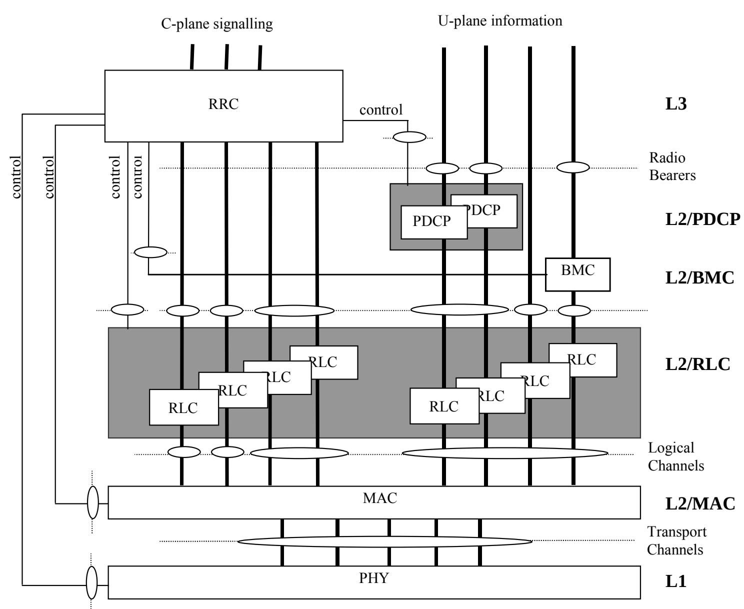


Figure 2.5. UTRA FDD Radio Interface protocol architecture.

The UTRAN side MAC architecture is shown in Figure 2.6. The MAC sublayer is made up of several different MAC entities, MAC-b, MAC-d, MAC-c/sh and MAC-hs.

MAC-b handles the broadcast channel (BCH). There is one MAC-b entity in each UE and one MAC-b in the UTRAN (located in the Node B) for each cell.

MAC-c-sh handles the common channels and shared channels – Paging Channel (PCH), Forward link Access Channel (FACH), Random Access Channel (RACH), uplink Common Packet Channel (CPCH) and Downlink Shared Channel (DSCH). There is one MAC c-sh entity in each UE that is using shared channels and one MAC-c/sh in the UTRAN (located in the controlling RNC) for each cell.

MAC-d is responsible for handling Dedicated Channels (DCH) allocated to a UE and one MAC-d entity in the UTRAN (in the serving RNC) for each UE.

The MAC-hs entity is responsible for establishing the HARQ entity in accordance with the higher layer configuration and handling all the tasks required to perform HARQ functionality. This functionality ensures delivery between peer entities by use of the ACK and NACK signalling between the peer entities. The transmitting MAC-hs entity assembles the data block payload for the MAC-hs PDUs from the delivered MAC-d PDUs. The MAC-d PDUs that are assembled in any one MAC-hs PDU are the same priority, and from the same MAC-d flow. The receiving MAC-hs entity is then responsible for the reordering of the received data blocks according to the received TSN, per priority and MAC-d flow, and then disassembling the data block into MAC-d PDUs for in-sequence delivery to the higher layers.

The data transfer services of the MAC layer are provided on logical channels. The type of information transferred defines each logical channel type. A general classification of logical channels is presented in [63], where the logical channels are divided into two groups – Control Channels and Traffic Channels. The control channels are used for transfer of control plane information and the traffic channels are used for the transfer of user plane information only.

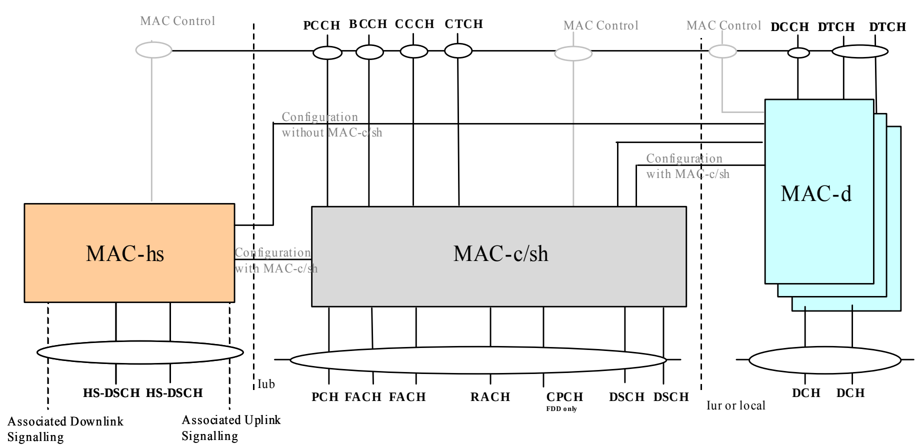


Figure 2.6. UTRAN side MAC architecture.

The **Control Channels** are:

- *Broadcast Control Channel* (BCCH), for broadcasting system control information in the downlink.
- *Paging Control Channel* (PCCH), for transferring paging information in the downlink (used when the network does not know the cell location of the UE, or, the UE is in cell-connected state).
- *Common Control Channel* (CCCH), for transmitting control information between the network and UEs in both directions (commonly used by UEs having no RRC connection with the network and by UEs using common transport channels when accessing a new cell after cell reselection).
- *Dedicated Control Channel* (DCCH). Point-to-point bi-directional channel for transmitting dedicated control information between the network and a UE (established through RRC connection set up procedure).

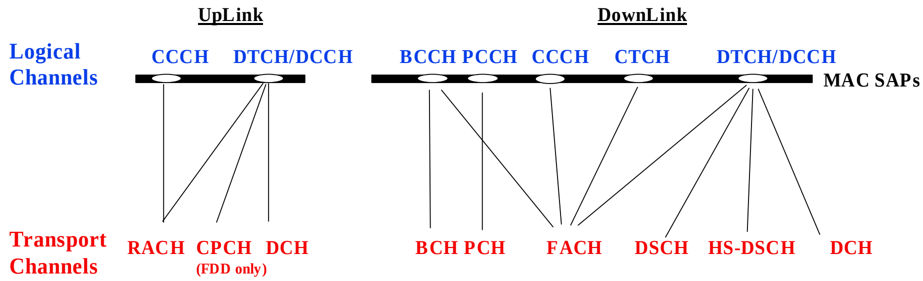
The **Traffic Channels** are:

- *Dedicated Traffic Channel* (DTCH). Point-to-point channel, dedicated to one UE for the transfer of user information (a DTCH can exist in both uplink and downlink directions).
- *Common Traffic Channel* (CTCH). Point-to-multipoint unidirectional channel for transfer of dedicated user information for all or a group of specified UEs.

The mapping between logical channels and transport channels is depicted in Figure 2.7.

2.3.2.3 Radio Link Control (RLC) Protocol

The radio link control protocol provides segmentation/reassembly (payloads units, PU) and retransmission services for both user (Radio Bearer) and control data (Signalling Radio Bearer) [66]. Each RLC instance is configured by RRC to operate in one of the three modes – Transparent mode (TM), where no protocol overhead is added to higher layer data; Unacknowledged Mode (UM), where no retransmission protocol is in use and data delivery is not guaranteed; and Acknowledged Mode (AM), where the Automatic Repeat reQuest (ARQ) mechanism is used for error correction. For all RLC modes, the CRC error detection is performed on the physical layer and the results of the CRC is delivered to the RLC together with the actual data. Some of the most important functions of the RLC protocol are – segmentation and re-assembly of variable length higher layer PDUs into/from smaller RLC Payload Units; error correction, by means of retransmission in the acknowledge data transfer mode; in-sequence delivery of higher layer PDUs; flow control, i.e. rate control at which the peer RLC transmitting entity may send information; protocol error detection and recovery; and ciphering. As shown in Table 2.1, the RLC transfer mode indicates the data transfer mode supported by the RLC entity configured for that particular Radio Bearer. The transfer mode for a RB is the same in both uplink and downlink directions; and it is determined by the admission control in the SRNC from the RAB attributes and CN domain information. The RLC transfer mode affects the configuration parameters of outer loop power control in the RNC and the user bit rate. The quality target is not affected if TM or UM RLC is used, whilst the number of retransmissions should be taken into account if AM RLC is employed. The user bit rate is affected by the transfer mode of the RLC protocol, since the length of the layer 2 headers is – 16 bits for AM; 8 bits for UM and 0 bits for TM. Hence, the user bit rate for network dimensioning is given by the layer 1 bit rate reduced by the layer 2-header bit rate.



BCCH	**Broadcast Control Channel**	**DCH**	**Dedicated Channel**
BCH	**Broadcast Channel**	**DSCH**	**Downlink Shared Channel**
CCCH	**Common Control Channel**	**HS-DSCH**	**High Speed-DSCH**
CCH	**Control Channel**	**DTCH**	**Dedicated Traffic Channel**
CPCH	**Common Packet Channel**	**FACH**	**Forward Access Channel**
CTCH	**Common Traffic Channel**	**PCCH**	**Paging Control Channel**
DCCH	**Dedicated Control Channel**	**PCH**	**Paging Channel**
		RACH	**Random Access Channel**

Figure 2.7. Mapping between logical channels and transport channels, in the uplink and downlink directions.

Table 2.1 RLC transfer modes for UMTS QoS classes [66].

UMTS QoS Class*	**Domain**	**Source statistics descriptor**	**Service type**	**RLC transfer mode**
Conversational	CS	Speech	CS speech	TM
		Unknown	CS T-data	TM
	PS	Speech	PS speech	UM
		Unknown	PS RT data	UM
Streaming	CS	Speech	CS speech	N/A
		Unknown	CS NT-data	TM
	PS	Speech	PS speech	N/A
		Unknown	PS RT data	AM or UM**
Interactive	CS	N/A	-	N/A
	PS	N/A	PS NRT data	AM
Background	CS	N/A	-	N/A
	PS	N/A	PS NRT data	AM

* Type of application for which the UMTS bearer service is optimised [1].

**Transfer mode depends on the value of RAB attribute Transfer delay.

2.3.2.4 Packet Data Convergence Protocol (PDCP)

This protocol exists only in the U-plane and only for services from the PS domain. The main PDCP functions are – compression of redundant protocol control information (e.g. TCP/IP and RTP/UDP/IP headers) at the transmitting entity and decompression at the receiving entity; transfer of user data, i.e. receiving a PDCP SDU form NAS and forwarding it to the appropriate RLC entity and vice versa; multiplexing of radio bearers into one RLC entity [67].

2.3.2.5 Broadcast Multicast Control (BMC) Protocol

Like the PDCP, the BMC protocol exists only in the user plane. This protocol provides a broadcast/multicast transmission service on the radio interface for common user data in transparent or unacknowledged mode. It utilises UM RLC using CTCH logical channel mapped onto FACH. The CTCH has to be configured and the TrCH used by the network has to be indicated to all UEs via RRC system information broadcast on the BCH [68].

2.3.2.6 Radio Resource Control (RRC) Protocol

The RRC signalling is used to control the mobility of the user equipment in the connected mode; to broadcast the information related to the NAS and AS; and to establish, reconfigure and release Radio Bearers. The RRC protocol is further used for setting up and controlling the UE measurement reporting criteria and the downlink outer loop power control. Also paging, control of ciphering, initial cell selection and cell re-selection are a part of the RRC connection management procedures. The RRC messages carry all parameters required to set-up, modify and release layer 2 and layer 1 protocol entities [69].

After power on, terminals stay in Idle Mode until a request to establish an RRC Connection is transmitted to the network. In Idle Mode the connection of the UE is closed on all layers of the Access Stratum. In Idle Mode the UE is identified by NAS identities such as International Mobile Subscriber Identity (IMSI), Temporary Mobile Subscriber Identity (TMSI) and Packet- TMSI. The RNC has no information about any individual UE, and it can only address for example all UEs in a cell or all UEs monitoring a paging occasion [69].

The transitions between Idle Mode and UTRA Connected Mode are illustrated in Figure 2.8. The UTRAN Connected Mode is entered when an RRC Connection is established. The RRC connection is defined as a point-to-point bi-directional connection between RRC peer entities in the UE and UTRAN. A UE has either zero or one RRC connection. The RRC connection establishment procedure can only be initiated by the UE sending an RRC connection request message to the radio access network. The event is triggered either by a paging request from the network or by a request from upper layers in the UE. When the RRC connection is established, the UE is assigned a Radio Network Temporary Identity (RNTI) to be used as own identity on common transport channels. When the network releases the RRC connection, the signalling link and all RBs between the UE and the UTRAN are released [69].

As depicted in Figure 2.8, the RRC states are (a description of the physical channels can be found in Section 2.4):

- *Cell_DCH*; in this state the dedicated physical channel (DPCH), plus eventually the physical downlink shared channel (PDSCH) or HS-PDSCH, is allocated to the UE. This state is entered from the idle mode or by establishing a dedicated transport channel (DCH) from Cell_FACH state. In this state the UE performs measurements according to the RRC: MEASUREMENT CONTROL message. The transition from Cell_DCH to Cell_FACH can occur either through the expiration of an inactivity timer or via explicit signalling.
- *Cell_FACH*; in this state no DPCH is allocated to the UE, the random access transport channel (RACH) and the forward transport channel (FACH) are used instead, for transmitting signalling and small amount of user data. The UE listens to the BCH system information and moves to Cell_PCH sub-state via explicit signalling when the inactivity timer on FACH expires.
- *Cell_PCH*; in this state the UE location is known by the SRNC on a cell level, but it can be reached only via a paging message. This state allows low battery consumption. The UE may use Discontinuous Reception (DRX), reads the BCH to acquire valid system information and moves to Cell_FACH if paged by the network or through any uplink access, e.g. initiated by the terminal for cell reselection (cell update procedure).
- *URA_PCH*; this state is similar to Cell_PCH, except that the UE executes the cell update procedure only if the UTRAN Registration Area (URA) is changed. One cell can belong to one or several URAs in order to avoid ping-pong effects. When the number of cell updates exceeds a certain limit, the UE may be moved to URA_PCH state via explicit signalling. The DCCH cannot be used in this state and any activity can be initiated by the network via a paging request on PCCH or through an uplink access by the terminal using RACH.

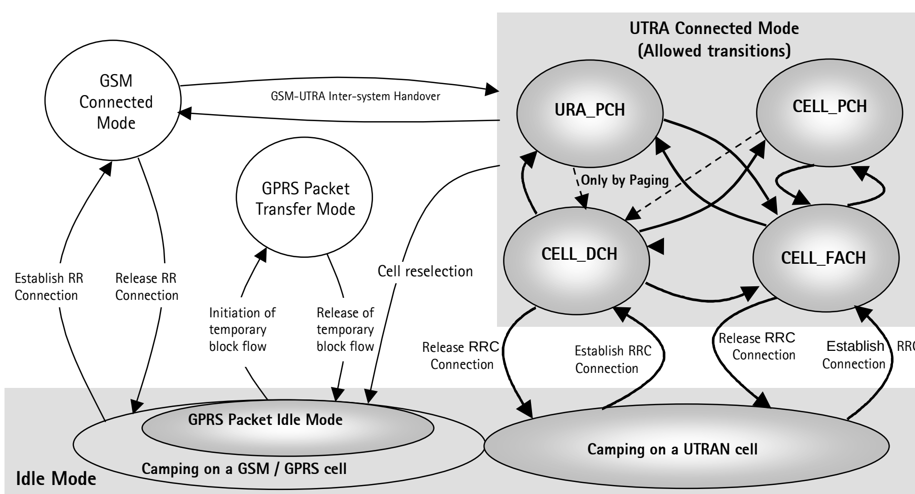


Figure 2.8. RRC States and State Transitions, including GSM connected mode for PSTN/ISDN domain services and GSM/GPRS packet modes for IP domain services.

2.4 Transport and Physical Channels

In UTRAN the data generated at higher layers is carried over the air interface using transport channels mapped onto different physical channels. The physical layer has been designed to support variable bit rate transport channels, to offer bandwidth-on-demand services, and to be able to multiplex several services within the same RRC connection. The single output data stream from the coding and multiplexing unit is denoted Coded Composite Transport Channel (CCTrCH). A CCTrCH is carried by one physical control channel and one or more physical data channels. In general there can be more than one CCTrCH, but only one physical control channel is transmitted on a given connection [70].

In 3GPP all TrCHs are defined as unidirectional, i.e. uplink, downlink, or relay-link. Depending on services and state, the UE can have simultaneously one or several TrCHs in the downlink, and one or more TrCHs in the uplink. As shown in Figure 2.9, for each TrCH, at any Transmission Time Interval (TTI) the physical layer receives from higher layers a set of Transport Blocks and the corresponding Transport Format Indicator (TFI). Then Layer 1 combines the TFI information received from different TrCHs to the Transport Formant Combination Indicator (TFCI). The TFCI is transmitted in the physical control channel to inform the receiver about what TrCHs are active in the current radio frame. In the downlink, in case of limited TFCSs the TFCI signalling may be omitted and Blind Transport Format Detection (BTFD) can be employed, where the TrCHs decoding can be done verifying in which position of the output block is matched with the CRC results [70].

Two types of transport channels exist – dedicated channels and common channels. A common channel is a resource divided between all or a group of users in a cell, whereas a dedicated channel is by definition reserved for a single user. The connections and mapping between transport channels and physical channels are depicted in Figure 2.10.

2.4.1 Dedicated Transport Channels

The only dedicated transport channel specified in 3GPP is the dedicated channel (DCH), which supports variable bit rate and services multiplexing. It carriers all user information coming from higher layers, including data for the actual service (speech frames, data, etc.) and control information (measurement control commands, UE measurement reports, etc.). It is mapped on the Dedicated Physical Data Channel (DPDCH). The DPCH is characterised by inner loop PC and fast data rate change on a frame-by-frame basis; it can be transmitted to part of the cell and supports soft/softer handover [63].

2.4.2 Common Transport Channels

The common transport channels are resource divided between all or a group of users in a cell (an in band identifier is needed). They do not support soft/softer HO, but some of them can have fast PC, e.g. Common Packet Channel (CPCH) and Downlink Shared Channel (DSCH). As depicted in Figure 2.7 and Figure 2.10, the common transport channels are [63]:

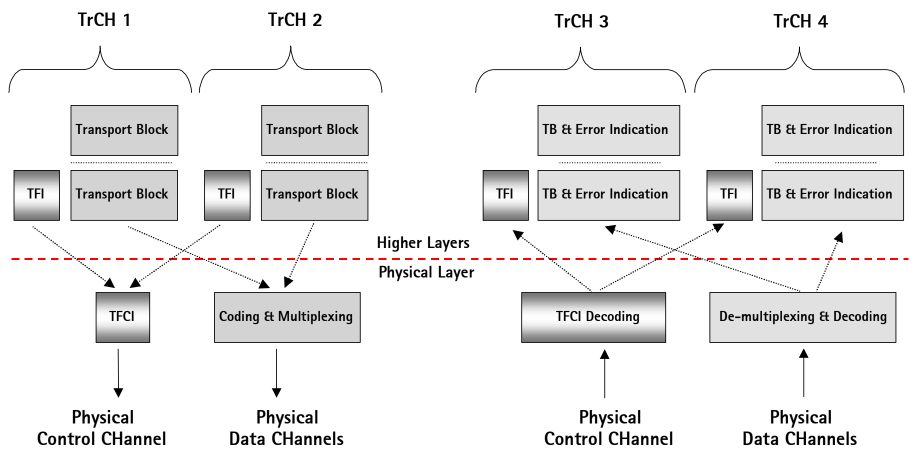


Figure 2.9. Interface between higher layers and the physical layer.

- *Broadcast Channel* (BCH). The BCH is used to transmit information (e.g. random access codes, cell access slots, cell type transmit diversity methods, etc.) specific to the UTRA network or to a given cell; is mapped onto the Primary Common Control Physical Channel (P-CCPCH), which is a downlink data channel, only.
- *Forward Access Channel* (FACH). The FACH carries downlink control information to terminals known to be located in the given cell. It is further used to transmit a small amount of downlink packet data. There can be more than one FACH in a cell, even multiplexed onto the same Secondary Common Control Physical Channel (S-CCPCH). The S-CCPCH may use different offsets between the control and data field at different symbol rate and may support slow PC.
- *Paging Channel* (PCH). The PCH carries data relevant to the paging procedure. The paging message can be transmitted in a single or several cells, according to the system configuration. It is mapped onto the S-CCPCH.
- *Random Access Channel* (RACH). The RACH carries uplink control information, such as a request to set up an RRC connection. It is further used to send small amounts of uplink packet data. It is mapped onto the Physical Random Access Channel (PRACH).
- *Uplink Common Packet Channel* (CPCH). The CPCH carries uplink packet-based user data. It supports uplink inner loop PC, with the aid of a downlink Dedicated Physical Control Channel (DPCCH). Its transmission may span over several radio frames and it is mapped onto the Physical Common Packet Channel (PCPCH).
- *Downlink Shared Channel* (DSCH). The DSCH carries dedicated user data and/or control information and it can be shared in time between several users. It is as a pure data channel always associated with a downlink DCH. It supports the use of downlink inner loop PC, based on the associated uplink DPCCH. It is mapped onto the Physical Downlink Shared Channel (PDSCH).
- *High-Speed Downlink Shared Channel (HS-DSCH).* The HS-DSCH is a downlink transport channel shared by several UEs. It has no fast power control, but uses link adaptation by varying the modulation, coding and transmission power. It is always associated with a DPCH and one or several Shared Control Channels (HS-SCCH). The

HS-DSCH is transmitted over the entire cell or over only part of the cell using e.g. beam-forming antennas.

The common transport channels needed for the basic cell operation are RACH, FACH and PCH, while the use of the DSCH, HS-DSCH and CPCH may or may not be used by the operator.

2.4.3 Physical Channels

In this section the dedicated physical channel and the main characteristics of the physical common channels are described. Further explanation can be found in [72].

A physical channel is identified by a specific carrier frequency, scrambling code, channelisation code (optional), duration and, on the uplink, relative phase (0 or $\pi/2$). In UMTS the modulation chip-rate is 3.84 Mcps and the transmission of a physical channel in normal mode is continuous. It is stopped in compressed mode to allow the UE to monitor cells on other FDD frequencies and radio access technologies (TDD and GSM).

- *Uplink Dedicated Physical Channel (UL DPCH).* It consists of one Dedicated Physical Control Channel (DPCCH) and one or more Dedicated Physical Data Channels (DPDCH). Dedicated higher layer information, including user data and signalling, is carried by the DPDCH, and the control information generated at Layer 1 is mapped on the DPCCH. The DPCCH comprises predefined pilot symbols (used for channel estimation and coherent detection/averaging), transmit power control (TPC) commands, feedback information (FBI) for closed loop mode Transmit Diversity and Site Selection Diversity Technique (SSDT), and optionally a transport-format combination indicator (TFCI). There can be zero, one, or several uplink DPDCHs on each radio link, but only one uplink DPCCH is transmitted. DPDCH(s) and DPCCH are I/Q code multiplexed with complex scrambling. Further, the uplink DPDCH can have spreading factor from 256 (=15 ks/s) down to 4 (=960 ks/s), whereas the uplink DPCCH is always transmitted with spreading factor 256 (=15 ks/s).

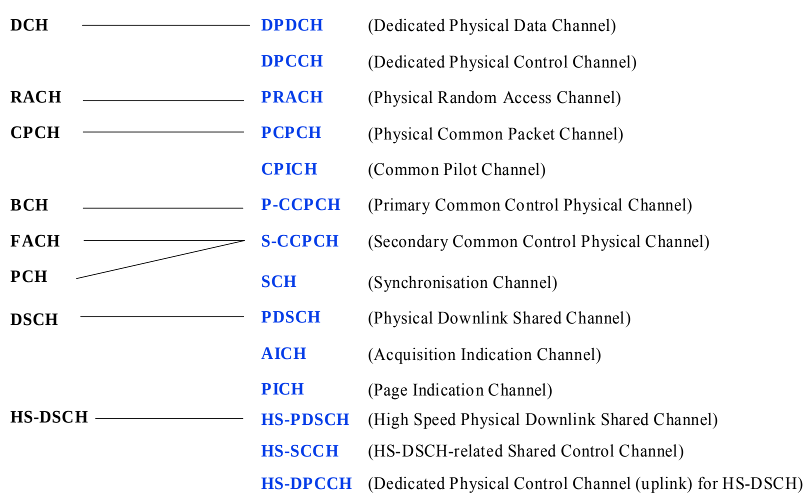


Figure 2.10. Mapping of transport channels onto physical channels.

- *Downlink Dedicated Physical Channel (DL DPCH).* It consists of a downlink DPDCH and a downlink DPCCH time multiplexed with complex scrambling. The dedicated data generated at higher layers carried on DPDCH is therefore time-multiplex with pilot bits, TPC commands, and TFCI bits (optional) generated by the physical layer. The DPCH may or may not include the TFCI; if the TFCI bits are not transmitted, then the DTX is used in corresponding field. The I/Q branches have equal power and the spreading factors range from 512 (7.5 ksps) down to 4 (960 ksps).
- *Common uplink physical channels*. They are the Physical Random Access Channel (PRACH) and the Physical Common Packet Channel (PCPCH), which are respectively used to carry RACH and CPCH. The RACH is supposed to be received from the entire cell and it is transmitted using open loop power control. The CPCH is transmitted using inner loop PC and it is always associated with a downlink DPCCH carrying TPC commands.
- *Common Pilot Channels (CPICH).* There are two types of common pilot channels, the Primary and Secondary CPICH. They are transmitted at fixed rate (15 kb/s, SF=256) and carry a pre-defined symbol sequence. The Primary Common Pilot Channel (P-CPICH) is characterised by a fixed channelisation code and it is always scrambled using a primary scrambling code. There is one P-CPICH per cell and it is broadcast over the entire cell. The P-CPICH is the phase reference for SCH, P-CCPCH, AICH, PICH and the S-CCPCH carrying PCH. The Secondary Common Pilot Channel (S-CPICH) is characterised by an arbitrary channelisation code of SF = 256 and it is scrambled by either a primary or a secondary scrambling code. In a cell there may be zero, one, or several S-CPICH. Each S-CPICH may be transmitted over the entire cell or only over a part of the cell.
- *Primary Common Control Physical Channel (P-CCPCH).* The P-CCPCH is a fixed rate (15 ks/s, SF = 256) downlink physical channel used to carry the BCH. It is a pure data channel characterised by a fixed channelisation code. The P-CCPCH is broadcast over the entire cell and it is not transmitted during the first 256 chips of each slot, where Primary SCH and Secondary SCH are transmitted instead
- *Secondary Common Control Physical Channel (S-CCPCH).* The S-CCPCH is used to carry FACH and PCH, which can be mapped onto the same S-CCPCH (same frame) or onto separate S-CCPCHs. The S-CCPCH spreading factor ranges from 256 (15 ks/s) down to 4 (960 ks/s). Fast power control is not allowed, but the power of the S-CCPCH carrying FACH only may be slowly controlled by the RNC. The S-CCPCH supports multiple transport format combinations (variable rate) using TFCI. It is on air only when there is data to transmit (available) and it may be transmitted in a narrow lobe in the same way as a dedicated physical channel, if the PCH is not mapped onto the same S-CCPCH.
- *Synchronisation Channel (SCH)*. The SCH is a pure physical channel used in the cell search procedure. It consists of two sub-channels transmitted in parallel, the Primary SCH and the Secondary SCH. The Primary SCH is transmitted once every slot; it allows the downlink slot synchronisation in the cell and is identical in every cell of the system. The Secondary SCH allows the downlink frame synchronisation and indicates which of the code groups the downlink primary scrambling code belongs to.
- *Physical Downlink Shared Channel (PDSCH)*. It carries the Downlink Shared Channel (DSCH). A PDSCH is allocated on a radio frame basis to a single UE. Within one radio

frame, different PDSCHs may be allocated under the same PDSCH root channelisation code to different UEs based on code multiplexing. Within the same radio frame, multiple parallel PDSCHs, with the same spreading factor, may be allocated to a single UE. This is a special case of multicode transmission. All the PDSCHs are operated with radio frame synchronisation. PDSCHs allocated to the same UE on different radio frames may have different spreading factors. For each radio frame, each PDSCH is associated with one downlink DPCH. The PDSCH and associated DPCH do not necessarily have the same spreading factors and are not necessarily frame aligned. All relevant Layer 1 control information is transmitted on the DPCCH part of the associated DPCH, i.e. the PDSCH does not carry Layer 1 information. To indicate for UE that there is data to decode on the DSCH, the TFCI field of the associated DPCH is used. The TFCI informs the UE of the instantaneous transport format parameters related to the PDSCH as well as the channelisation code of the PDSCH. For PDSCH the allowed spreading factors may vary from 256 to 4.

- *Acquisition Indicator Channel (AICH)*. The AICH is a fixed rate physical channel (SF = 256) used to indicate in a cell that the base station has received PRACH preambles (signatures). Once the base station has received a preamble, the same signature that has been detected on the PRACH preamble is then sent back to the UE using this channel. Higher layers are not involved in this procedure: a response from the RNC would be too slow in order to acknowledge a PRACH preamble. The AICH consists of a repeated sequence of 15 consecutive access slots (AS) of length 5120 chips.
- *Paging Indicator Channel (PICH).* The PICH is a physical channel used to carry paging indicators (PIs). This channel is transmitted at fixed rate (SF = 256) and is always associated with a S-CCPCH, where the PCH is mapped.
- *High Speed – Shared Control Channel (HS-SCCH).* The HS-SCCH is a fixed rate (60 kb/s, SF=128) downlink physical channel used to carry downlink signalling related to HS-DSCH transmission. (See Section 2.4.4.)
- *High Speed – Physical Downlink Shared Channel (HS-PDSCH).* The HS- PDSCH carries the HS-DSCH. A HS-PDSCH corresponds to one channelisation code of fixed spreading factor SF=16 from the set of channelisation codes reserved for HS-DSCH transmission. Multi-code transmission is allowed, which translates to UE being assigned multiple channelisation codes in the same HS-PDSCH subframe, depending on its UE capability. An HS-PDSCH may use QPSK or 16QAM modulation symbols. All relevant Layer 1 information is transmitted in the associated HS-SCCH i.e. the HS-PDSCH does not carry any Layer 1 information.
- *High Speed – Dedicated Physical Control Channel (HS-DPCCH)*. The HS-DPCCH carries uplink feedback signalling related to downlink HS-DSCH transmission. The HS-DSCH-related feedback signalling consists of Hybrid-ARQ Acknowledgement (HARQ-ACK) and Channel-Quality Indication (CQI). There is at most one HS-DPCCH on each radio link. The HS-DPCCH can only exist together with an uplink DPCCH. The spreading factor of the HS-DPCCH is 256 i.e. there are 10 bits per uplink HS-DPCCH slot.

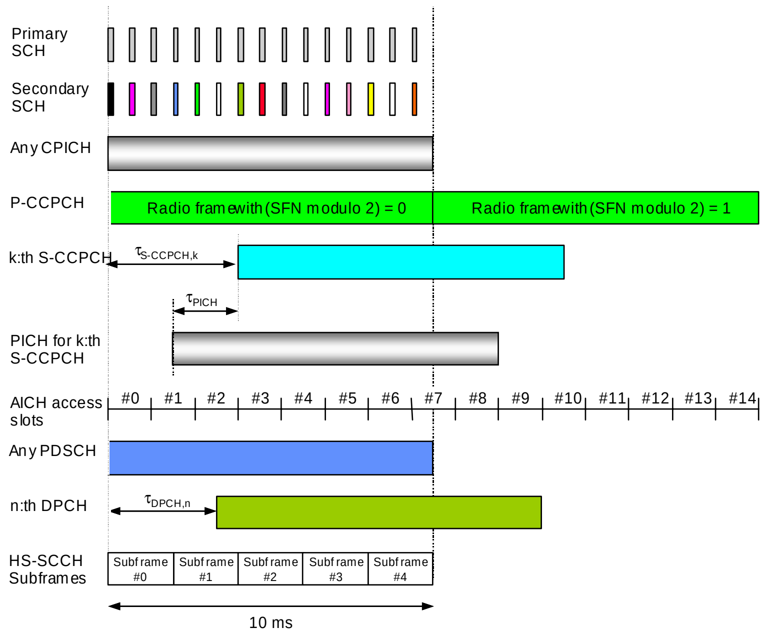


Figure 2.11. Radio frame timing and access slot timing of downlink physical channels.

2.4.3.1 Timing Relationship Between Physical Channels

The radio frame and access slot timing structure of the downlink physical channels are illustrated in Figure 2.11. As shown in the figure, the cell system frame number (SFN) is transmitted on the P-CCPCH, which is used as timing reference for all physical channels, since the transmission timing in the uplink is derived from the timing of the downlink physical channels [72].

The SCH (primary and secondary), CPICH (primary and secondary), P-CCPCH, and PDSCH have identical frame timing. The S-CCPCH timing may be different for different S-CCPCHs, but the offset from the P-CCPCH frame timing is a multiple of 256 chips. The PICH timing is 7680 chips prior to its corresponding S-CCPCH frame timing, i.e. the timing of the S-CCPCH carrying the PCH transport channel with the corresponding paging information. The AICH access slots #0 starts at the same time as P-CCPCH frames with (SFN modulo 2) = 0. The DPCH timing may be different for different DPCHs, but the offset from the P-CCPCH frame timing is always a multiple of 256 chips. The start of HS-SCCH subframe #0 is aligned with the start of the P-CCPCH frames. The HS-PDSCH starts 5120 chips (2 slots) after the start of the HS-SCCH.

2.4.4 Formats and Configurations

In order to describe how the mapping of TrCHs is performed and controlled by L1, some generic definitions and terms valid for all types of TrCH are introduced in this section. Further information can be found in [70].

- *Transport Block* (TB), is the basic unit exchanged between L1 and MAC for L1 processing; a TB typically corresponds to an RLC PDU or corresponding unit; layer 1 adds a CRC to each TB.
- *Transport Block Set* (TBS), is defined as a set of TBs, which are exchanged between L1 and MAC at the same time instance using the same transport channel.
- *Transport Block Size*, is defined as the number of bits in a TB; the Transport Block Size is always fixed within a given TBS, i.e. all TBs within a TBS are equally sized.
- *Transport Block Set Size*, is defined as the number of bits in a TBS.
- *Transmission Time Interval* (TTI), is defined as the inter-arrival time of TBSs, and is equal to the periodicity at which a TBS is transferred by the physical layer on the radio interface. It is always a multiple of the minimum interleaving period (i.e. 10ms, the length of one Radio Frame). MAC delivers one TBS to the physical layer every TTI.
- *Transport Format* (TF), is the format offered by L1 to MAC (and vice versa) for the delivery of a TBS during a TTI on a given TrCH. It consists of – one *dynamic part* (Transport Block Size, Transport Block Set Size); and one *semi-static part* (TTI, type of error protection [turbo code, convolutional code or no channel coding], coding rate, static RM parameter, size of CRC).
- *Transport Format Set* (TFS), is a set of TFs associated to a TrCH. The semi-static parts of all TFs are the same within a TFS. TB size, TBS size and TTI define the TrCH bit rate before L1 processing. As an example, for a DCH, assuming a TB size of 336 bits (= 320 bit payload + 16 RLC header), a TBS size of 2 TBs per TTI, and a TTI of 10 ms; the DCH bit rate is given by 336*2/10 = 67.2 kbit/s. Whereas the DCH user bit rate, which is defined as the DCH bit rate minus the RLC headers, is given by 320*2/10 = 64 kbit/s. Depending on the type of service carried by the TrCH, the variable bit rate may be achieved by changing between TTIs either the TBS Size only, or both the TBS and TBS Size.
- *Transport Format Combination* (TFC), an authorised combination of the currently valid TFs that can be simultaneously submitted to layer 1 on a CCTrCH of a UE, i.e. containing one TF from each TrCH being a part of the combination.
- *Transport Format Combination Set* (TFCS), is defined as a set of TFCs on a CCTrCH and a proprietary algorithm in the RNC produces it. The TFCS is what is given to MAC by L3 for control. When mapping data onto L1, MAC chooses between the different TFCs specified in the TFCS. MAC has any and only control over the dynamic-part of the TFC, since the semi-static part corresponds to the service attributes (quality, transfer delay) set by the admission control in the RNC. The selection of TFCs can be seen as the fast part of the radio resource control dedicated to MAC, close to L1. Thereby the bit rate can be changed very fast, without any need of L3 signalling. An example of data exchange between MAC and physical layer when two DCHs are multiplexed in the connection is illustrated in Figure 2.12. The TFCS may be produced as shown in Figure 2.13, i.e. as a Cartesian product between TFSs of the TrCHs that are multiplexed onto a CCTrCH, each one of those considered as a vector. In theory every TrCH can have any TF in the TFC, but in practice only a limited number of possible combinations are selected.
- *Transport Format Indicator* (TFI), as pointed out in the introduction, is a label for a specific TF within a TFS. It is used in the inter-layer communication between MAC and L1, each time a TBS is exchanged between the two layers on a transport channel.

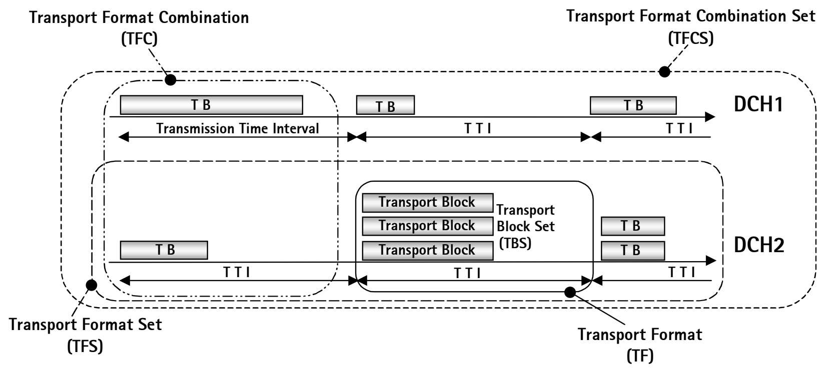


Figure 2.12. Example of data exchange between MAC and the physical layer when two DCHs are employed.

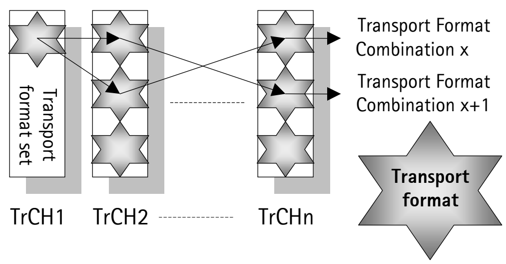


Figure 2.13. Relations of transport format, transport format set and transport format combination.

- *Transport Format Combination Indicator* (TFCI), as explained in the introduction, is used in order to inform the receiving side of the currently valid TFC, and hence to decode, de-multiplex and transfer the received data to MAC on the appropriate TrCHs. MAC indicates the TFI to L1 at each delivery of TBSs on each TrCH. L1 then builds the TFCI from the TFIs of all parallel TrCHs of the UE, processes the TBs and appropriately appends the TFCI to the physical control signalling (DPCCH). Through the detection of the TFCI the receiving side is able to identify the TFC.
- *Transport Format for HS-DSCH*, consists of three parts – one *dynamic part*, one *semi-static part* and *one static part*. The Transport Format for HS-DSCH is always explicitly signalled. There is no support of blind transport format detection. Attributes of the dynamic part are: Transport Block size (same as Transport Block Set size); Redundancy version/Constellation; and modulation scheme. No semi-static attributes are defined. Attributes of the static part are: Transmission time interval (fixed to 2ms in FDD); error protection scheme to apply (turbo coding; coding rate is 1/3); and size of CRC (24 bits).
- *Hybrid ARQ for HS-DSCH*. With the help of the HARQ information the UE is enabled to identify the process being used for the transport block that is received on the HS-DSCH. The HARQ information also includes information that indicates whether a new

data block is transmitted for the first time or a retransmission. Furthermore it is used to decode the received data correctly.

- *Transport Format and Resource Indication (TFRI).* The TFRI includes information about the dynamic part of the HS-DSCH transport format, including transport block set size and modulation scheme. The TFRI also includes information about the set of physical channels (channelisation codes) onto which the HS-DSCH is mapped in the corresponding HS-DSCH TTI.

2.4.5 Models and Functions of the Physical Layer

One UE can transmit only one CCTrCH at once, but multiple CCTrCHs can be simultaneously received in the forward direction. In the uplink one TFCI represents the current TFs of all DCHs of the CCTrCH. RACHs are always mapped one-to-one onto physical channels (PRACHs), i.e. there is no physical layer multiplexing of RACHs. Further, only a single CPCH of a CPCH set is mapped onto a PCPCH, which employs a subset of the TFCs derived by the TFS of the CPCH set. A CPCH set is characterised by a set-specific scrambling code for access preamble and collision detection, and it is assigned to the terminal when a service is configured for CPCH transmission. In case of a configured HS-DSCH one physical channel (HS-DPCCH) is configured for the reporting of HS-DSCH transport block acknowledgement/negative acknowledgement and channel quality indicator [70].

In the downlink the mapping between DCHs and physical channel data streams works in the same way as in the reverse direction. The current configuration of the coding and multiplexing unit is either signalled (TFCI) to the terminal, or optionally blindly (Blind TF Detection - BTFD) detected. Each CCTrCH has only zero or one corresponding TFCI mapped (each 10 ms radio frame) on the same DPCCH used in the connection. When the DSCH is employed in the communication, the DSCH TFI also indicates the channelisation code used for the shared channel. A PCH and one or several FACH can be encoded and multiplexed together forming a CCTrCH, one TFCI indicates the TFs used on each FACH and PCH carried by the same S-CCPCH. The PCH is always associated with the Paging Indicator Channel (PICH), which is used to trigger off the UE reception of S-CCPCH where the PCH is mapped. A FACH or a PCH can also be individually mapped onto a separate physical channel. The BCH is always mapped onto the P-CCPCH, without any multiplexing with other transport channels. For each HS-DSCH TTI, each HS-SCCH carries HS-DSCH-related downlink signalling for one UE. The following information is carried on the HS-SCCH: Transport Format and Resource Indicator (TFRI); Hybrid-ARQ-related Information (HARQ information); UE Identity via a UE specific CRC [70].

The main functions of the physical layer are – FEC encoding/decoding of transport channels, measurements and indication to higher layers (e.g. BER, SIR, interference power, transmission power, etc.), macro diversity distribution/combining and softer handover execution, error detection on transport channels (CRC), multiplexing of transport channels and de-multiplexing of CCTrCHs, rate matching, mapping of CCTrCHs on physical channels, modulation/demodulation and spreading/despreading of physical channels, frequency and time (chip, bit, slot, frame) synchronisation, closed-loop power control (inner loop PC), power weighting, combining of physical channels and RF processing.

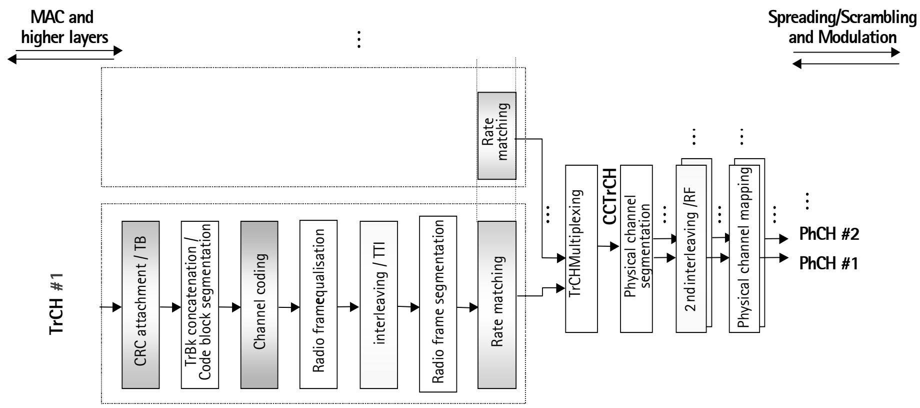


Figure 2.14. Uplink multiplexing and channel coding chain.

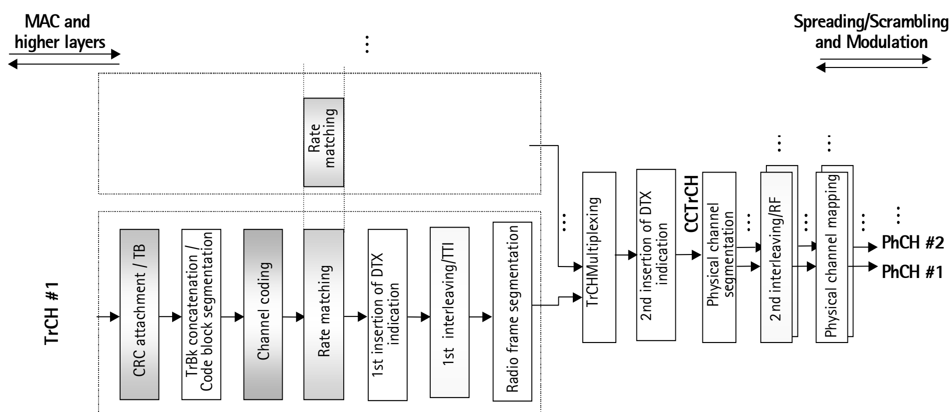


Figure 2.15. Downlink multiplexing and channel coding chain.

2.4.6 General Coding and Multiplexing of Transport Channels

This section only applies to DCH, RACH, CPCH, DSCH, BCH, FACH and PCH. For further details, other transport channels and HSDPA related descriptions see [71].

The multiplexing and channel coding chain is depicted in Figure 2.14 and Figure 2.15 for the uplink and downlink direction, respectively. As shown in these figures, data arrives to the coding/multiplexing unit in form of transport block sets once every Transmission Time Interval. The transmission time interval is transport-channel specific from the set {10 ms, 20 ms, 40 ms, 80 ms} [71].

Error detection is provided on transport blocks through a Cyclic Redundancy Check. The CRC length is determined by the admission control in the RNC and can be 24, 16, 12, 8 or 0 bits; the more bits the CRC contains, the lower is the probability of having undetected errors in the receiver. The CRC is carried out based on the CRC parity bits of each transport block; and the parity bits are generated using the cyclic generator polynomials specified in [71]. Regardless of the result of the CRC check, all TBs are delivered to L2 along with the associated error indications. This estimation is then used as quality information for uplink macro diversity selection/combining in the RNC. Also, this indication may be directly used as an error indication to L2 for each erroneous TB in TM, UM and AM RLC, provided that RLC PDUs are one-to-one mapped onto TBs.

Depending on whether the TB fits in the available code block size (channel coding method), the transport blocks in a TTI are either concatenated or segmented to coding blocks of suitable size.

Channel coding and radio frame equalisation is performed on the coding blocks after the concatenation or segmentation operation. Only the channel coding schemes reported in Table 2.2 can be applied to TrCHs, i.e. convolutional coding (CC), turbo coding or no coding (no limitation on the coding block size).

Convolutional coding is supposed to be used with relative low data rate, e.g. the BTFD using Viterbi decoder is much faster than turbo coding. Whereas turbo coding is applied for higher data rates and brings performance benefits when large enough block size are achieved for significant interleaving effect [17]. As an example, the AMR speech service (coordinated TrCHs, multiplexed in the FP) uses UEP (Unequal Error Protection): Class A bits, strong protection (1/3 cc and 12 bit CRC); Class B bits, less protected (1/3 cc); and Class C bits, the least protected (1/2 cc).

The function of the radio frame equalisation (padding) is to ensure that data arrive after channel coding can be divided into equalised blocks when transmitted over more than a single 10 ms radio frame. The radio frame size equalisation is only performed in the UL, because in the DL the rate matching output block length is already produced in blocks of equal size per frame.

The 1st interleaving (or radio-frame interleaving) is used when the delay budget allows more than 10 ms of interleaving period. The 1st interleaving period is related to the TTI and can be 20, 40 or 80 ms.

Rate matching is used to match the number of bits to be transmitted to the number of bits available on a single frame (DPCH). This is achieved either by puncturing or by repetition. The amount of repetition/puncturing for each service depends on the service combination and their QoS requirements. The rate matching procedure takes into account the number of bits of all TrCHs active in that frame. The admission control in the RNC provides a semi static parameter, the *rate-matching attribute* (RMA), to control the relative rate matching between different TrCHs. The rate-matching attribute is used to calculate the rate matching value when multiplexing several TrCHs for the same frame. With the aid of the rate matching attribute and TFCI the receiver can calculate backwards the rate matching parameters used and perform the inverse operation. By adjusting the rate-matching attribute, admission control of the RNC fine-tunes the quality of different services in order to reach an equal or near equal symbol power level requirement for all services.

The variable rate handling is performed after TrCH multiplexing for matching the total instantaneous rate of the multiplexed TrCHs to the channel bit rate of the DPDCH (when the transport block sets do not contain the maximum number of DPDCH bits). The number of bits on a TrCH can vary between different TTIs. In the downlink the transmission is interrupted if the number of bits is lower than maximum allowed by the DPDCH.

Table 2.2 TrCH coding schemes [71].

<table>
<tr><th>Type of TrCH</th><th>Coding scheme</th><th>Coding rate</th></tr>
<tr><td>BCH</td><td rowspan="4">Convolutional coding</td><td rowspan="3">1/2</td></tr>
<tr><td>PCH</td></tr>
<tr><td>RACH</td></tr>
<tr><td rowspan="3">CPCH, DCH, DSCH, FACH</td><td>1/3, 1/2</td></tr>
<tr><td>Turbo coding</td><td>1/3</td></tr>
<tr><td colspan="2">No coding</td></tr>
</table>

In the uplink bits are repeated or punctured to ensure that the total bit rate after TrCHs multiplexing is identical to the total channel bit rate of the allocated DPCHs. The rate matching is performed in a more dynamic way and the SF may vary on a frame-by-frame basis (see Appendix D).

The multicode transmission is employed when the total bit rate to be transmitted on a CCTrCH exceeds the maximum bit rate of the DPCH. The multicode transmission depends on the multicode capabilities of UE and Node B, and consists of several parallel DPDCHs transmitted for one CCTrCH using the same spreading factor.

- In the downlink, if several CCTrCHs are employed for one UE, each CCTrCH can have a different spreading factor, but only one DPCCH is used for them in the connection.
- In the uplink, the UE can use only one CCTrCH simultaneously. Multi-code operation is possible if the maximum allowed amount of puncturing has already been applied. For the different codes it is mandatory for the terminal to use the SF 4. Up to 6 parallel DPDCHs and only one DPCCH per connection can be transmitted.

The second interleaving is also called intra-frame interleaving (10 ms interleaving). It consists of block inter-column permutations, separately applied for each physical channel (if more than a single code channel is transmitted). The amount of bits output at this stage is exactly the number of bits the spreading factor of that frame can transmit.

2.5 QoS Concept and Architecture

In UMTS, the network QoS is provided by a Bearer Service, which is set up from the source to the destination of a service. The bearer service includes all aspects to enable the provision of a contracted QoS. The UMTS bearer service layered architecture is shown in Figure 2.16. Each bearer offers its individual services using services provided by the layers below [1]. A TE (Terminal Equipment) / MT (Mobile Terminal) Local Bearer Service, a UMTS Bearer Service, and an External Bearer Service realise an End-to-End-Service used by the TE. The UMTS bearer service provides the UMTS QoS, and it is what the operators offer. The UMTS bearer consists of two parts: The Radio Access Bearer (RAB) and the Core Network (CN) Bearer Service. The RAB Service provides confidential transport of signalling and user data between the MT and SGSN (Serving GPRS Support Node) with the QoS negotiated when the UMTS bearer is set up or with the default QoS for signalling. The CN Bearer Service connects the SGSN with the GGSN to the external packet data (PD) network. The role of this service is to efficiently control and utilise the backbone network in order to provide the contracted UMTS bearer service quality.

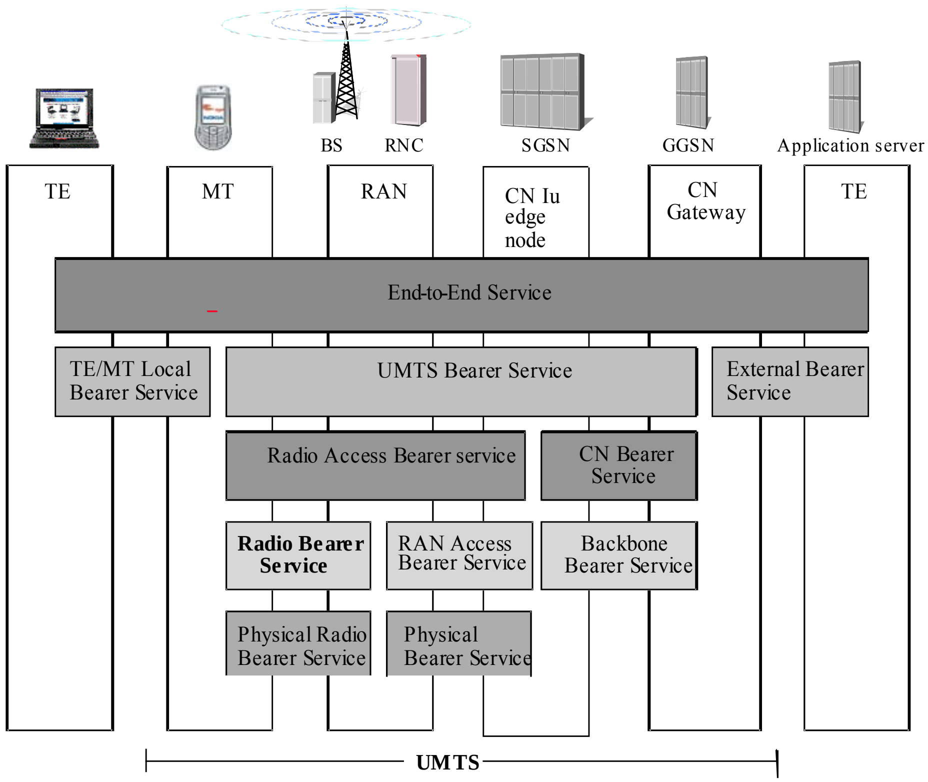


Figure 2.16. UMTS QoS Architecture.

A RAB consists of a Radio Bearer and a RAN Access Bearer Service. The Radio Bearer Service covers all the aspects of the radio interface transport. This bearer, provided by the UTRA FDD networks, is the only service within the scope of this work. The RAN Access Bearer Service together with the Physical Bearer Service provides the transport between RAN and CN. Ultimately, the Backbone Network Service covers the layer 1/layer2 functionality to fulfil the QoS requirements of the Core Network Bearer Service.

2.6 QoS Management Functions in the Network

This section gives an overview of functionality needed to establish, modify and maintain a UMTS Bearer Service with a specific QoS. *The relations between the functions internal to the nodes are implementation specific.* The allocation of these functions to the UMTS entities indicates the requirements for the specific entity to enforce the QoS commitments negotiated for the UMTS bearer service. *The specific realisation of these functions is implementation dependent and has only to maintain the specified QoS characteristics*. The QoS management functions of all UMTS entities together must ensure the provision of the negotiated service between the access points of the UMTS bearer service. The end-to-end service is provided by translation/mapping with UMTS external services [1].

2.6.1 QoS Functions for UMTS Bearer Service in the Control Plane

The QoS management functions for controlling the UMTS bearer service are shown in Figure 2.17.

- ***Service Manager*** co-ordinates the functions of the control plane for establishing, modifying and maintaining the service. It also provides all user plane QoS management functions with the relevant attributes. The service manager may perform an attribute translation to request lower layer services. Furthermore, it may interrogate other control functions to receive permission for service provision.
- ***Translation function*** converts between the internal service primitives for UMTS bearer service control and the various protocols for service control of interfacing external networks. The translation includes the converting between UMTS bearer service attributes and QoS attributes of the external networks service control protocol (e.g. between IETF TSPEC and UMTS service attributes). The service manager may include a translation function to convert between its service attributes and the attributes of a lower layer service it is using.
- ***Admission/Capability control*** maintains information about all available resources of a network entity and about all resources allocated to UMTS bearer services. It determines for each UMTS bearer service request or modification whether the required resources can be provided by this entity and it reserves these resources if allocated to the UMTS bearer service. The function checks also the capability of the network entity to provide the requested service, i.e. whether the specific service is implemented and not blocked for administrative reasons. The resource control performed by the admission control supports also the service retention.
- ***Subscription Control*** checks the administrative rights of the UMTS bearer service user to use the requested service with the specified QoS attributes.

2.6.2 Functions for UMTS Bearer Service in the User Plane

The QoS management functions of the UMTS BS for the user plane are shown in Figure 2.18. User plane QoS management functions maintain the signalling and user data traffic based on QoS attributes. These functions ensure the provision of the QoS negotiated for a UMTS bearer service.

- ***Mapping function*** provides each data unit with the specific marking required to receive the intended QoS at the transfer by a bearer service.
- ***Classification function*** assigns data units to the established services of a MT according to the related QoS attributes if the MT has multiple UMTS bearer services established. The appropriate UMTS bearer service is derived from the data unit header or from traffic characteristics of the data.

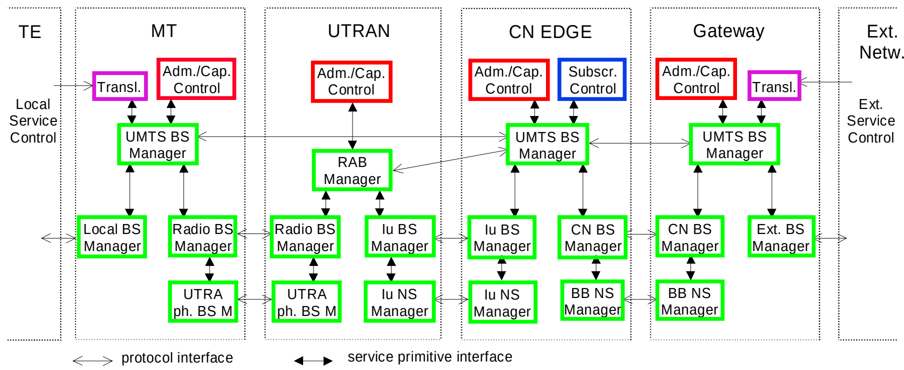


Figure 2.17. QoS management functions for UMTS bearer service in the control plane.

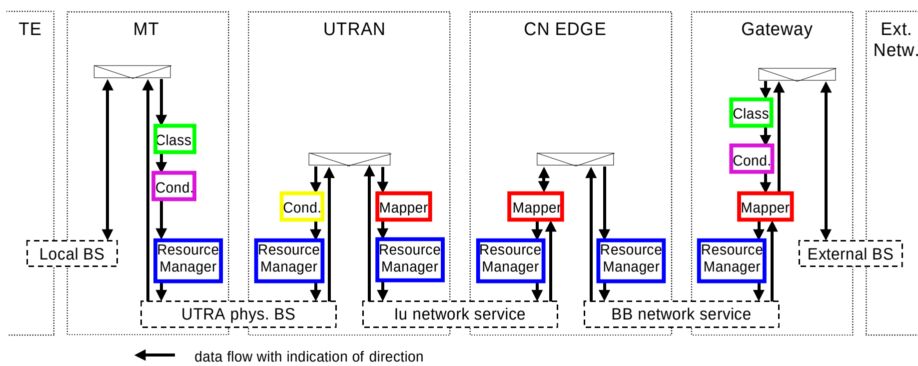


Figure 2.18. QoS management functions for UMTS bearer service in the user plane.

- ***Resource Manager*** distributes the available resources between all services sharing the same resource. The resource manager distributes the resources according to the required QoS. Example means for resource management are scheduling, bandwidth management and power control for the radio bearer.
- ***Traffic Conditioner*** provides conformance between the negotiated QoS for a service and the data unit traffic. Traffic conditioning is performed by policing or by traffic shaping. The policing function compares the data unit traffic with the related QoS attributes. Data units not matching the relevant attributes will be dropped or marked as not matching, for preferential dropping in case of congestion. The traffic shaper forms the data unit traffic according to the QoS of the service.

2.7 UMTS Bearer Service Attributes

UMTS bearer service attributes describe the service provided by the UMTS network to the user of the UMTS service. A set of QoS attributes (QoS profile) specifies this service. At

UMTS bearer service establishment or modification different QoS profiles are taken into account.

- The UE capabilities form a QoS profile, which may limit the UMTS bearer service that can be provided.
- The UE or the TE within the terminating network may request a QoS profile at UMTS bearer establishment or modification.
- A QoS profile in the UMTS subscription describes the upper limits for the provided service if the service user requests specific values.
- A Network specific QoS profile characterising for example the current resource availability or other network capabilities or limitations may limit the provided UMTS bearer service or initiate a modification of an established UMTS bearer service.

2.7.1 UMTS Bearer Service Attributes

The QoS profile consists of the following attributes:

- ***Traffic Class ('Conversational', 'Streaming', 'Interactive', 'Background')***. Application type for which the UMTS bearer service is optimised. The main distinguishing factor between these QoS classes is how delay sensitive the traffic is: Conversational class is meant for traffic which is very delay sensitive, while Background class is the most delay insensitive traffic class. Conversational and Streaming classes are mainly intended for carrying real-time traffic flows, such as video telephony and audio/video streams. Interactive class and Background are mainly meant for carrying traditional Internet applications like WWW, Email, Telnet, FTP and News. Due to looser delay requirements, compared to Conversational and Streaming classes, both provide better error rate by means of channel coding and retransmission. Traffic in the Interactive class has higher priority in scheduling than Background class traffic, so background applications use transmission resources only when interactive applications do not need them. However, these are only typical examples of usage of the traffic classes. There is in particular no strict one-to-one mapping between classes of service and the traffic classes defined in [1]. For instance, a streaming service by nature can use the Interactive traffic class to save radio resources, as shown in the following chapters of this work.
- ***Maximum Bit Rate (kb/s)***. Upper bit rate limit a user/application can accept or provide. All UMTS bearer service attributes may be fulfilled for traffic up to the Maximum bit rate depending on the network conditions. It may be used to make code reservations in the downlink of the radio interface for applications able to operate with different rates.
- ***Guaranteed Bit Rate (kb/s)***. Bit rate the UMTS bearer service must guarantee to the user or application. The Guaranteed bit rate may be used to facilitate admission control based on available resources, and for resource allocation within UMTS.
- ***Delivery Order (y/n)***. Derived from the user protocol (PDP type) and specifies if out-of-sequence SDUs are acceptable or not. Whether out-of-sequence SDUs are dropped or re-ordered depends on the specified reliability.
- ***Maximum SDU Size (octets)***. Maximum SDU size for which the network must satisfy the negotiated QoS. The maximum SDU size is used for admission control and policing and/or optimising transport (optimised transport in for example the RAN may be dependent on the size of the packets). (Note: The Maximum Transfer Unit (MTU) of the IP layer and the Maximum SDU Size have no relationship.)

- ***SDU Format Information (bits)***. List of possible exact sizes of SDUs. RAN needs SDU size information to be able to operate in transparent RLC protocol mode, which is beneficial to spectral efficiency and delay when RLC re-transmission is not used.
- ***SDU Error Ratio***. Fraction of SDUs lost or detected as erroneous. It is defined only for conforming traffic. It is used to configure the protocols, algorithms and error detection schemes, primarily within RAN.
- ***Residual Bit Error Ratio***. Indicates the undetected bit error ratio in the delivered SDUs. If no error detection is requested, Residual bit error ratio indicates the bit error ratio in the delivered SDUs. As above, it is used to configure radio interface protocols, algorithms and error detection coding.
- ***Delivery of Erroneous SDUs (y/n/-)***. Indicates whether SDUs detected as erroneous must be delivered or discarded.
- ***Transfer Delay (ms)***. Indicates maximum delay for 95th percentile of the distribution of delay for all delivered SDUs during the lifetime of a bearer service, where delay for an SDU is defined as the time from a request to transfer an SDU at one SAP to its delivery at the other SAP. This attribute allows RAN to set transport formats and ARQ parameters. Transfer delay of an arbitrary SDU is not meaningful for a bursty source, since the last SDUs of a burst may have long delay due to queuing, whereas the meaningful response delay perceived by the user is the delay of the first SDU of the burst.
- ***Traffic Handling Priority***. *S*pecifies the relative importance for handling of all SDUs belonging to the UMTS bearer compared to the SDUs of other bearers. Within the Interactive class, there is a definite need to differentiate between bearer qualities. This is handled by using the traffic handling priority attribute, to allow UMTS to schedule traffic accordingly. By definition, priority is an alternative to absolute guarantees, and thus these two attribute types cannot be used together for a single bearer.
- ***Allocation/Retention Priority***. It specifies the relative importance compared to other UMTS bearers for allocation/retention of the UMTS bearer. The Allocation/Retention Priority attribute is a subscription attribute that is not negotiated from the mobile terminal. In situations where resources are scarce, the relevant network elements can use the ARP to prioritise bearers with a high Allocation/Retention Priority over bearers with a low ARP when performing admission control.
- ***Source Statistics Descriptor ('speech'/'unknown')***. It specifies characteristics of the source of submitted SDUs.
- ***Signalling Indication (Yes/No)***. Indicates the signalling nature of the submitted SDUs. This attribute is additional to the other QoS attributes and does not over-ride them. This attribute is only defined for the Interactive traffic class. If signalling indication is set to 'Yes', the UE should set the traffic handling priority to '1'. This attribute permits enhancing the RAN operation accordingly. An example use of the Signalling Indication is for IMS signalling traffic. (Note: this indication is sent by the UE in the QoS IE.)

When establishing a UMTS bearer and the corresponding RAB for support of a service request, some attributes on UMTS level do not have typically the same value.

For example requested transfer delay for the UMTS bearer is typically larger that the requested transfer delay for the Radio Access Bearer, as the transport through the core network will use a part of the acceptable delay. The residual BER and SDU error ratios for the RAB service are reduced with the errors introduced in the CN. The relation between the attribute values for UMTS Bearer service and RAB service is implementation depended.

Further, the following attributes/settings only exist on the RAB level:

- ***SDU Format Information*** - exact format of SDU payload is retrieved from the codec integrated in the core network.
- ***Source Statistics Descriptor*** is set to speech if the RAB transports compressed speech generated by the codec integrated in the core network.

For the other attributes/settings the attribute value for the UMTS bearer is normally the same as the corresponding attribute value for the Radio Access Bearer.

Some considerations of the above attributes per traffic class and a discussion of 2G – 3G interworking issues can be found in [1], which also specifies the mapping between the release 99 QoS attributes and the QoS attributes for GPRS releases 97 and 98. The ranges of the RAB attributes are reported in Table 2.3.

2.8 Packet Data Transfer Across UMTS Networks

The section introduces the end-to-end packet data transmission and combined models for protocols used to control, support and carry user plane application data (see [73] for more information). Our target is to explain the mapping between bearer services and service access points (SAPs) of protocols, and the information available in the network elements in order to classify performance counters and indicators during measurements. As explained in Chapter 6, such identifiers will allow the network management system (NMS) to monitor the distinct offered services based on the corresponding PDP contexts.

Table 2.3. UMTS QoS profile (attributes): Ranges of values for Radio Access Bearer services [1].

Traffic Class	**Conversational**	**Streaming**	**Interactive**	**Background**
Maximum bit rate (kb/s)	≤ 16000	≤ 16000	≤ 16000 – overhead	≤ 16000 – overhead)
Delivery order	Yes/No	Yes/No	Yes/No	Yes/No
Maximum SDU size (octets)	≤ 1502	≤ 1502	≤ 1502	≤ 1502
Delivery of erroneous SDUs	Yes/No/-	Yes/No/-	Yes/No/-	Yes/No/-
Residual BER	$5*10^{-2} – 10^{-6}$	$5*10^{-2} – 10^{-6}$	$4*10^{-3} – 6*10^{-8}$	$4*10^{-3} – 6*10^{-8}$
SDU error ratio	$10^{-2} – 10^{-5}$	$10^{-2} – 10^{-5}$	$10^{-3} – 10^{-6}$	$10^{-3} – 10^{-6}$
Transfer delay (ms)	< 80	< 250	-	-
Guaranteed bit rate (kb/s)	≤ 16000	≤ 16000	-	-
Traffic handling priority	-	-	1,2,3	-
Allocation/Retention priority	1,2,3	1,2,3	1,2,3	1,2,3

2.8.1 Control Plane Protocol Stack

GMM supports mobility management functionality such as attach, detach, security, and routeing area update. SM supports PDP context activation, modification, deactivation, and preservation. SMS supports the mobile-originated and mobile-terminated short message service. The RANAP protocol encapsulates and carries higher-layer signalling, handles signalling between the 3G SGSN and I_u mode RAN, and manages the GTP connections on the I_u interface. RANAP is specified in 3GPP TS 25.413. The layers below RANAP are defined in 3GPP TS 25.412 and 3GPP TS 25.414. The RLC protocol offers logical link control over the radio interface for the transmission of higher layer-signalling messages and SMS. RLC is defined in 3GPP TS 25.322. GPRS Tunnelling Protocol for the control plane (GTP-C) is used for signalling messages between SGSNs and GGSNs (G_n), and between SGSNs in the backbone network (G_p). The User Datagram Protocol (UDP) is the transport protocol for signalling messages between GSNs. UDP is defined in RFC 768.

2.8.2 User Plane Protocol Stack

The UMTS user plane protocol stack is depicted in Figure 2.20. The numbers in the figure define the SAPs between protocol layers where the performance of the related bearer and thus the corresponding offered QoS may be assessed. The mapping of bearer services onto protocol service access points is reported in Table 2.4. The metrics defined in Chapter 6 characterise only the performance of the Radio Bearer service, which operates between SAP 4 and 6, as illustrated in Figure 2.20.

The Packet Data Convergence Protocol (PDCP) provides transparency for upper-layer protocols, e.g. IPv4, PPP and IPv6, and protocol control information compression. The GPRS Tunnelling Protocol (GTP) encapsulates all PDP PDUs, i.e. tunnels user data, between the RNC and SGSN, and between GSNs in the backbone network. UDP/IP is the backbone network protocol used for routing user data and control signalling. The Radio Link Control (RLC) protocol provides logical link control over the radio interface. There may be several simultaneous RLC links per UE and each link is identified with a Bearer Id. The Medium Access Control (MAC) controls the access (request and grant) procedures for the radio channel. The MAC-hs handles the HS-DSCH specific functions.

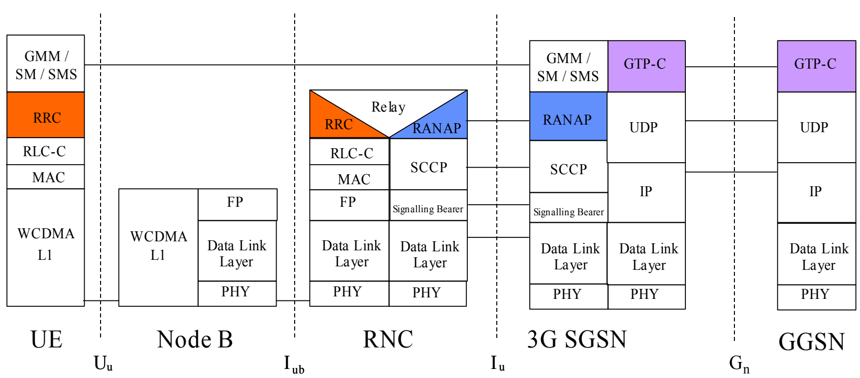


Figure 2.19. PS domain - Control plane protocol stack.

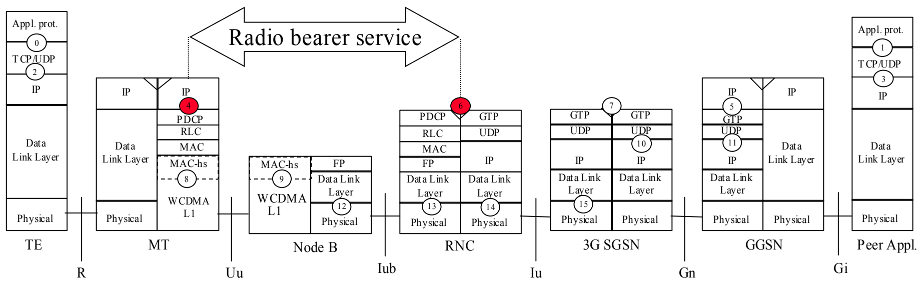


Figure 2.20. PS domain - User plane protocol stack.

Table 2.4. Mapping of bearer services onto protocol service access points.

Bearer service	**Service Access Point (SAP)**	
Service applications	0	1
Network services	2	3
UMTS bearer service	4	5
Radio Access Bearer service	4	7
Core network bearer service	6	5
Radio Bearer service	4	6
Iu bearer service	6	7
Backbone network service	10	11
Physical bearer service	12 (14)	13 (15)
UTRA FDD	8	9

A PDP (Packed Data Protocol) context is a virtual communication pipe established between the UE and the GGSN (SAPs 4 and 5 in Figure 2.20) for delivering the data traffic stream. The PDP context is defined in the UE, SGSN and GGSN by:

- A PDP Context Identifier (index of the PDP context)
- A PDP Type (e.g. PPP or IP)
- A PDP Address (e.g. an IP address)
- An Access Point Name (label describing the access point to the PD network)
- A QoS Profile (bearer service attributes).

There is a one to one correspondence between PDP context, UMTS bearer and RAB, as well as between RAB and RB Service, which, however, can be carried by more transport channels of the same type at the radio interface. A QoS profile is associated with each PDP context. The QoS profile is considered to be a single parameter with multiple data transfer attributes, as illustrated in Table 2.3.

2.8.3 Procedure Example

This section describes the PDP context activation procedure and mapping of bearer service characteristics onto radio interface protocols. More information on UMTS functionalities and system procedures can be found in [64] and [73].

The PDP context activation procedure is depicted in Figure 2.21. The UE sends an Activate PDP Context Request message to the SGSN. The message includes: A Network Sublayer Access Point Id (NSAPI), Transaction Identifier (TI), PDP Type, PDP Address, Access Point Name, QoS Requested, PDP Configuration Options. The UE leaves the PDP Address empty to request a dynamic PDP address and may use Access Point Name to select a reference point to a certain external network and/or to select a service (PDP Configuration Options may be used to request optional PDP parameters from the GGSN). The SGSN validates the Activate PDP Context Request using the PDP Type (optional), PDP Address (optional), and Access Point Name (optional) provided by the UE and the PDP context subscription records. If a GGSN address can be derived, the SGSN creates a Tunnel Endpoint Id (TEID) for the requested PDP context. If the UE requests a dynamic address, the SGSN lets a GGSN allocate the dynamic address. The SGSN may restrict the requested QoS attributes given its capabilities and the current load, and it must restrict the requested QoS attributes according to the subscribed QoS profile. The SGSN sends a Create PDP Context Request message to the affected GGSN. The message contains the PDP Type, PDP Address (empty if a dynamic address is requested), Access Point Name, QoS Negotiated, TEID, NSAPI, MSISDN, Charging Characteristics, and PDP Configuration Options. The GGSN may use the APN to find an external network and optionally to activate a service for this APN.

The GGSN creates a new entry in its PDP context table and generates a Charging Id (the new entry allows the GGSN to route PDP PDUs between the SGSN and the external PDP network, and to start charging). The GGSN then returns a Create PDP Context Response (TEID, PDP Address, PDP Configuration Options, QoS Negotiated, Charging Id, Cause) message to the SGSN. PDP Address is included if the GGSN allocated a PDP address; if the GGSN has been configured by the operator to use External PDN Address Allocation for the requested APN, PDP Address must be set to 0.0.0.0, indicating that the PDP address is negotiated by the UE with the external PDN after completion of the PDP Context Activation procedure. If QoS Negotiated received from the SGSN is incompatible with the PDP context being activated, the GGSN rejects the Create PDP Context Request message. In I_u mode, RAB setup is done by the RAB Assignment procedure.

The GGSN confirms the new QoS attributes by sending an Update PDP Context Response to the SGSN. The SGSN inserts the NSAPI along with the GGSN address in its PDP context. If the UE has requested a dynamic address, the PDP address received from the GGSN is inserted in the PDP context. The SGSN selects Radio Priority and Packet Flow Id based on QoS Negotiated, and returns an Activate PDP Context Accept message to the UE. The message includes the PDP Type, PDP Address, TI, QoS Negotiated, Radio Priority, Packet Flow Id, and PDP Configuration Options. The SGSN is now able to route PDP PDUs between the GGSN and the UE, and to start charging. For each PDP Address a different QoS profile may be requested. If a QoS requirement is beyond the capabilities of a PLMN, the PLMN negotiates the QoS profile as close as possible to the requested QoS profile. The UE either accepts the negotiated QoS profile, or deactivates the PDP context. If the PDP Context Activation Procedure fails or if the SGSN returns an Activate PDP

Context Reject (Cause, PDP Configuration Options) message, the UE may attempt another activation to the same APN up to a maximum number of attempts.

During the RAB assignment procedure, the AC in the RNC, based on the QoS profile it receives from the CN, produces the appropriate layer 2 and layer 1 parameters to establish and maintain the radio bearer characteristics.

The RB configuration encompasses, among other things, the definition of the RLC transmission mode, i.e. Transparent (TM), Acknowledged (AM) or Unacknowledged Mode (UM), quality targets (BLER, Block Error Rate and SIR, Signal to Noise Interference Ratio), the selection of the transport channels, and the derivation of the corresponding codes of the physical channels employed at the radio interface.

Figure 2.22 depicts a possible mapping of the radio bearer characteristics for circuit switched (CS) and packed switched (PS) services onto RLC modes and types of transport channels between layer 1 and 2. In CS domain, guaranteed bit rate (GB) services are carried on Conversational or Streaming Class using TM RLC and dedicated transport channels (DCHs). Packet switched services with guaranteed bit rate run either on Conversational Class using UM RLC, or on Streaming Class using UM or AM RLC, depending on the Transfer Delay attribute value (see Table 2.3); for GB traffic DCHs are always used. Non-guaranteed bit rate (NGB) services are only PS, and are mapped onto Interactive or Background Class using AM RLC. In this case, the user data transmission is possible using the Random Access Channel (RACH) and Forward Access Channel (FACH) in Cell_FACH state, or employing the combination of DCHs with High Speed Downlink Shared Channels (HS-DSCH) or DCHs in Cell_DCH state. (Note: All packet switched services presented in the following chapters, namely: audio and video streaming, PoC, SWIS, WAP, MMS and Dialup, are offered using AM RLC.)

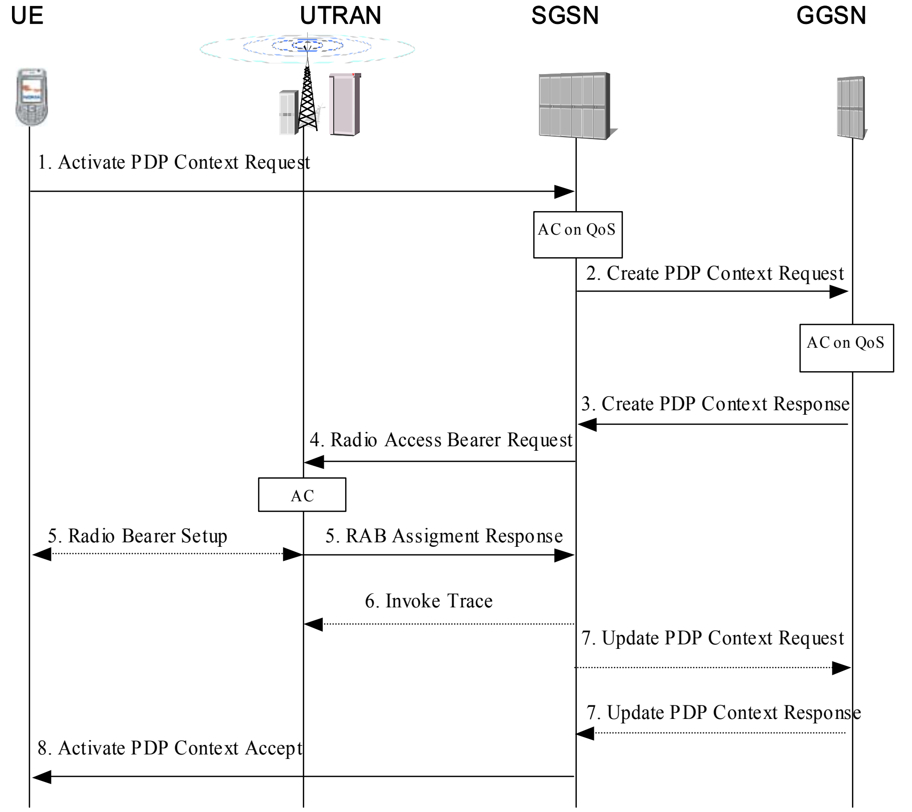


Figure 2.21. PDP context activation procedure.

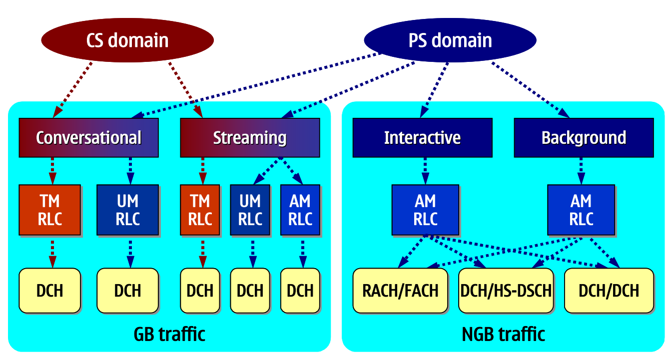


Figure 2.22. Mapping of bearer service characteristics onto RLC transmission modes and transport channels between layer 2 and layer 1.

3

Models and Assumptions

This chapter describes traffic, system, and environment models adopted in this work. The main assumptions and adopted methodology (description of the proposed simulators) are also reported in the following sections. As a part of this framework, a virtual time simulator (denoted by *Ares*, in this work) for studying the provisioning of QoS in UTRAN is also proposed. Besides this, to investigate the impact and entailed limitations of the deployment of new services on the performance of the offered ones, this chapter describes plain methods and a simplified version of the above tool (denoted as *Max*) for WCDMA radio interface dimensioning. The simulator (based on throughput estimates and snapshots of the system status) supports essential RRM functions, such as admission control and packet (bit rate) scheduler with QoS differentiation. Measures used for assessing system performance in this thesis are also presented at the end of the chapter. For each of the offered services, the corresponding user satisfaction criterion is defined, and performance results, at a given QoE, are expressed in terms of spectral efficiency. Ultimately, the performances obtained with Ares are compared to related results attained with a system level dynamic simulator (denoted by *Wallu* in this work). The comparison was performed using the same simulation scenario, traffic models, user satisfaction criteria and network parameter settings. Some differences resulted in the magnitudes of powers, throughputs, and satisfaction ratios, especially, in high loaded cells. In general, due to the simplified RRM algorithms, Ares indicated better performance than Wallu. However, the trends of results were similar in that sorting the cells in the order of almost any tested measure would produce similar lists for Ares and Wallu. The obtained comparison results validated that the virtual time simulator proposed in this work does quite well the task it was designed for, i.e. planning networks from the QoS point of view. The study did not reveal such aspects that would question the use of Ares, for instance, in preliminary screening of good methods and parameter values prior to selected trials with Wallu.

3.1 Introduction

3GPP specifies only a layered bearer service architecture and QoS attributes and leaves the implementation, dimensioning, as well as planning aspects of the actual QoS management functions needed to handle bearer services with a specific QoS to vendors' and operators' choice [1]. Since the design and parameter settings of radio resource management (RRM) functions have a direct influence on the overall service performance and infrastructure costs, the deployed algorithms in the UTRAN and configurations thereof will definitely be an issue in the mature networks. Hence, due to the complexity of the system, any practical realisation and deployment of the above QoS management functions and offered services needs to be validated a priori by means of static or dynamic simulations, depending on the desired level of time resolution and accuracy. As a matter of fact, several analytical methods and tools are reported in the literature, for example (see Section 1.2.2 for more information): Capacity and coverage estimations in a multi-service context based on an analytical expression for the total power of the base station was covered in [8]. New dimensioning tables, equivalent to Erlang laws for circuit switched, stemming from simulations to convert the volume of packet data into a required pipe (radio spectrum bandwidth) respecting the QoS were proposed in [9]. A static simulator for studying WCDMA radio network planning issues was presented in [10] and attached in [11]. A comprehensive description of an advanced WCDMA dynamic simulator for an effective analysis of service performance was described in [16]. Evaluation of the feasibility and accuracy of such a tool, as well as a presentation thereof, can be found in [17]. However, none of the published analytical methods and tools showed enough flexibility for an efficient and effective WCDMA radio interface dimensioning and QoS provisioning. The theoretical approaches, mainly based on circuit switched type of communications, lose the particulars, and radio network planning tools do not take into consideration the possibility of handling radio resources according to offered traffic mix characteristics, such as priorities and quality of experience (QoE) of the users of the services. On the other hand, dynamic simulators typically run with far too high time resolution, and thus need lengthy simulation times to design (plan) radio access networks and/or to analyse thoroughly the deployment of service applications.

This main theme of this chapter is to introduce a virtual time simulator that gives more insight into the challenge of QoS provisioning and overcomes limitations and complexity of static tools and dynamic system level simulators, respectively. The described tool is an enhanced version of the simulator the author designed and presented in [52], which now supports a multi-cell scenario with realistic power control and traffic models for relevant packet switched services, such as Push to Talk over Cellular (PoC), See What I See (SWIS, or real time video sharing, RTVS), Multimedia Messaging (MMS), WAP browsing and Dialup connections. Also, for each of these services customised user satisfaction criteria are defined [53]. Besides this, a simplified version (single cell) of the above simulator based on throughput estimates is presented. This software, in its simplicity, provides a fast response upon the effects caused by the deployment of a new (novel) service on existing applications in a WCDMA cell, and makes it possible for the operator to know when would be the case to enhance the capacity of the access network by e.g. deploying a new carrier [51].

The remainder of this chapter is organised as follows: Section 3.2, 3.3, and 3.4 describe, traffic, system (network) and environment models, respectively. Section 3.5 presents the methodology (simulators) used for studying the system performance in this thesis. Section 3.6 reports the performance measures employed in this work. The verification of the virtual time prediction method is discussed in Section 3.7. The overall conclusions on the adopted methods and assumptions are drawn in Section 3.8.

3.2 Traffic Models

All call and session arrivals are generated following a Poisson process [74]. Then according to the given traffic mix each connection is associated with a certain service, which in turn is mapped onto a corresponding QoS profile. Speech and video calls are circuit switched (CS) type of communications, hence held for an exponentially distributed service time [74]; their inter-arrival period follows exactly the same type of distribution, as a consequence of the Poisson distributed number of arrivals per time unit [74]. All other supported services are packet switched. The sessions are implemented as an ON/OFF process with the object size generated using limited range (truncated) distributions [75]. The model for a packet service session is depicted in Figure 3.1. The duration of the ON period depends mainly on the allocated bit rate and object size, which is modelled differently depending on the carried application. (Packet calls are blocks of continuous user data; distributions of single packets are not modelled.) Different distributions are also used to model the corresponding OFF time behaviours and session lengths (expressed in number of objects). The utilised traffic models are listed in Table 3.1. More details on a part of the adopted models can be found in [74]-[78], the distributions and pseudo code for the Poisson arrivals process are reported in Appendix A. PoC and SWIS are novel services for mobile networks [79]. PoC is similar to the walkie-talkie, where one person is talking while one or more people are listening. SWIS enables, in parallel (concomitantly) to an ongoing 2-way person-to-person communication, the sharing of video and/or audio data. In this work, SWIS is a peer-to-peer, unidirectional streaming service between terminals, i.e. there is a sending and a receiving terminal without an associated backward media or communication channel. Traffic mixes (*share of calls*, in percentage) and traffic volumes (*mean arrival rate*, in seconds, for call/session arrivals, which means that on average a user makes a connection every mean arrival rate seconds) are set differently depending on the case study.

3.3 System Models

In this thesis, performance results were analysed using a simulated macro cellular network located in the downtown of Finland's capital Helsinki. The network consists of one RNC connected to nine sites with up to three sectors each, totalling in 19 cells with an average antenna height of 18 m (see Figure 3.2). All sectors used 65° directional antennas. The parameter settings for each of the simulated cells are reported in Appendix B, Table B.1. The most important system based parameters (common to all case studies presented in this work) are listed in Table 3.2.

3.4 Environment Models

The propagation loss was calculated using an Okumura-Hata model with an average area correction factor of –6.3 dB. Figure 3.3 and Figure 3.4 show the cell dominance areas and the path losses, respectively, for the simulation scenario adopted in this work depicted in Figure 3.2.

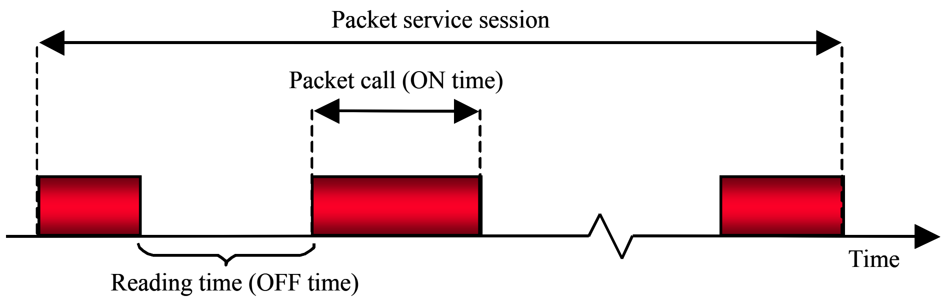


Figure 3.1. Model for packet service sessions.

Table 3.1. Adopted traffic models (see Appendix A).

Service	Data rate (kb/s)	Buffer size (s)	Object size (k octets)	Off time (s)	Session length (Objects)
PoC	8	1	Exponential 6 mean, 0.5 min, 40 max	Exponential 60 mean, 1 min, 1200 max	Geometric 8 mean, 1 min, 30 max
Streaming (Audio and video)	64	8	Uniform 160 min, 3200 max	-	1 (1 packet call)
MMS	Best Effort	-	Exponential 20 mean, 3 min, 200 max	-	1 (1 packet call)
Dialup	Best Effort	-	Log-normal (μ=5, σ=1.8) 0.1 min, 20000 max	Pareto (k=2, α=1) 2 min, 3600 max	Inverse Gaussian (μ=3.86, λ=6.08) 1 min, 50 max
SWIS (RTVS)	64	1	Exponential 80 mean, 32 min, 2400 max	-	1 (1 packet call)
WAP	Best Effort	-	Log-normal (μ=2, σ=1) 0.1 min, 50 max	Exponential 20 mean, 1 min, 600 max	Geometric 3 mean, 1 min, 50 max
CS Speech	12.3	-	-	-	Exponential 90 s
CS Video	64	-	-	-	Exponential 120 s

Table 3.2. Most important system based parameters.

Parameter	Value
Orthogonality (α), see (4.3)	0.5
E_b/N_0 values[a]:	
Speech	7 dB
SWIS	6 dB
Streaming	6 dB
PoC	7 dB
MMS/WAP	5/5.5 dB
Dialup	5.5 dB
Maximum BTS Tx power	43 dBm
P-CPICH Tx power and all other CCH Tx powers	33 dBm and 30 dBm

a. Required received values for a satisfactory connection quality in terms of BER.

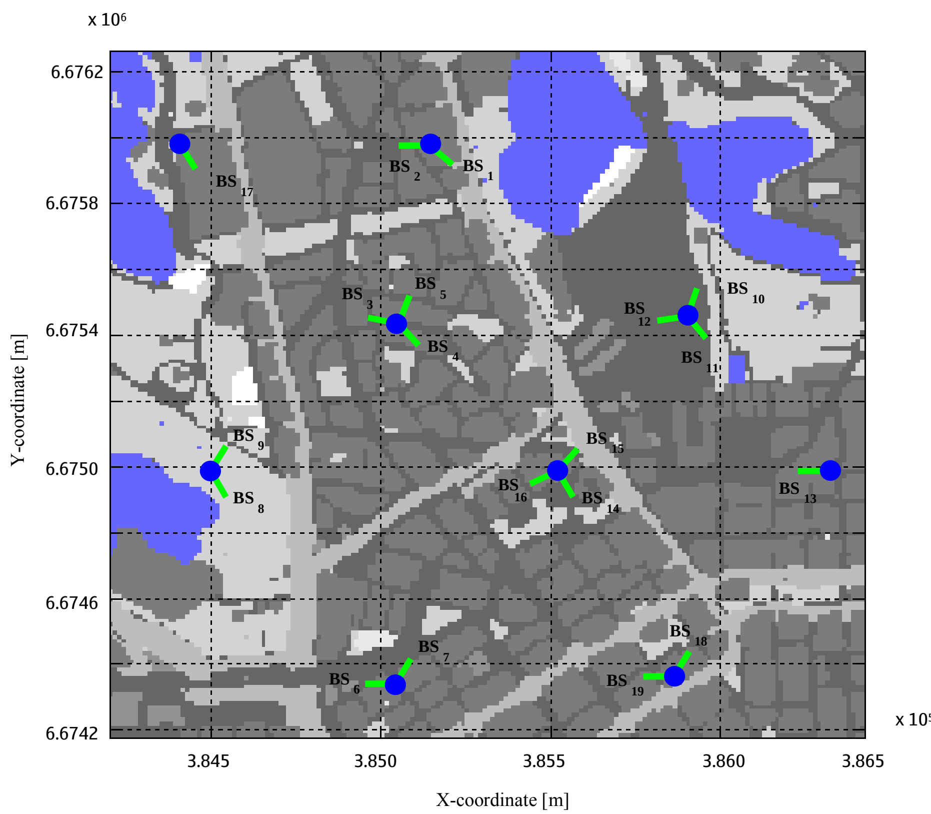


Figure 3.2. Simulation scenario used in the case studies discussed in this thesis.

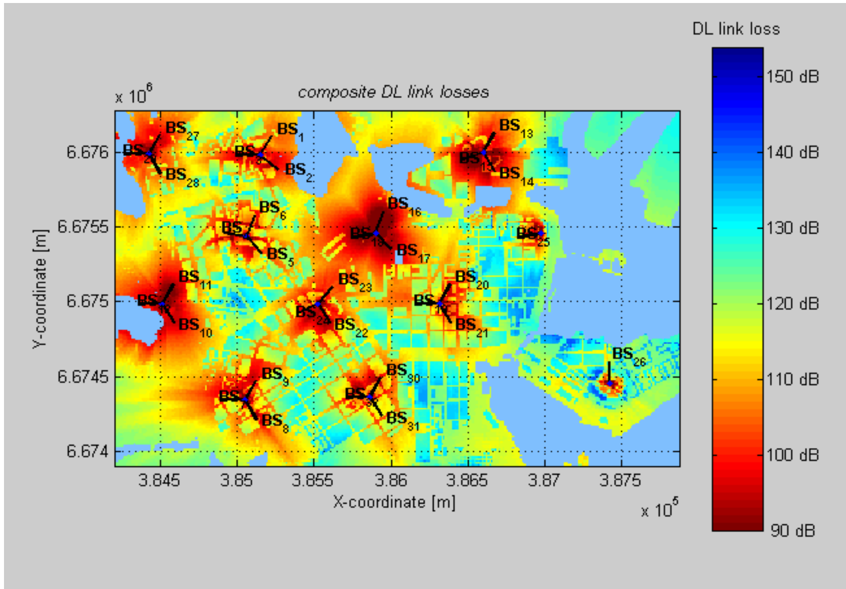


Figure 3.3. Link losses of the simulation scenario illustrated in Figure 3.2.

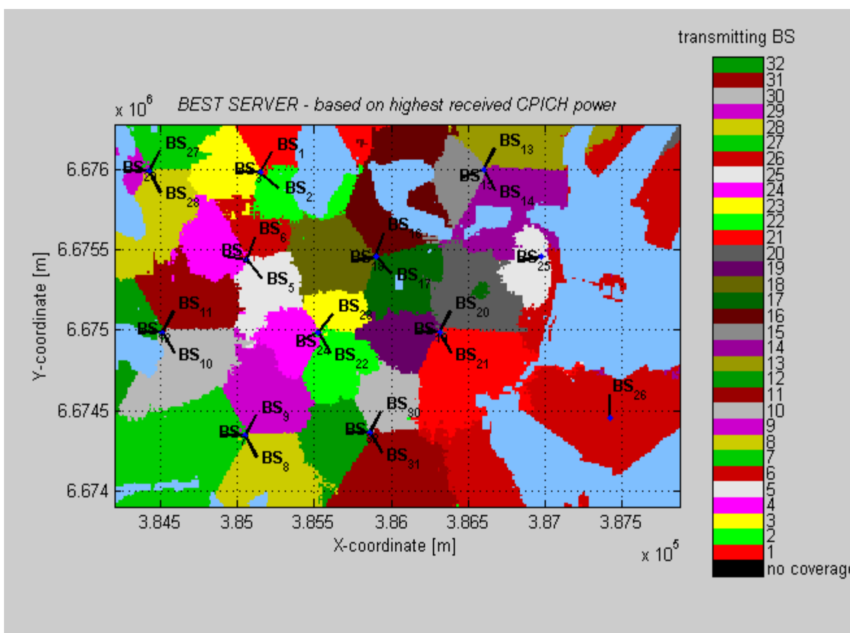


Figure 3.4. Cell dominance areas of the simulation scenario illustrated in Figure 3.2.

3.5 Methodology

In this section, the methodology for studying the system performance is presented. This includes an explanation of the virtual time simulator (Ares) used in this work for:

- Analysing the QoS management functions proposed in Chapter 4.
- Service driven radio network planning, as discussed in Chapter 5.
- QoS "optimisation", as described in Chapter 7.

In this section, a description of a simplified version of the above tool (Max) is also included. This tool is used for addressing the radio interface dimensioning issues that may arise from the introduction of new services in WCDMA networks, as presented in Chapter 5.

The validation of the virtual time prediction method is reported in Section 3.7. The verification methodology is based on the comparison of performance results with a short-term extensively tested and validated dynamic simulator. Measurement data at the time of writing was not available. Only the channel propagation model introduced in Section 3.4 was tuned using experimental data (drive tests).

3.5.1 Virtual Time Simulator (Ares)

The simulator structure consists of seven modules, namely: A generator for traffic mix and path losses, an Admission Control (AC) component, a Load Control (LC) component, a Packet Scheduler (PS) component, a Power Control (PC) component, a Process Calls (PrC) component and a Performance Management (PM) function. AC, LC and PS are cell-based functions, whereas PC, PrC and PM are system-based functions. The statistically large enough amount of terminals in the system does not make it necessary to have them really moving: The mobility effects may be taken into account by e.g. speed dependent E_b/N_0 requirements. (This approach leads to good results if the mobile speed in the analysed traffic scenarios is supposed to be low with respect to the data traffic dynamic, e.g. sessions terminate before mobile terminals can travel for a significant distance. The mobility issue is further discussed in Section 3.7, where the proposed methodology is verified through a Wallu-Ares performance results comparison.) Soft handover (SHO) affects mainly AC and PS. In the former, diversity (DHO) branches are processed first, followed by the main branches. In the latter, the bit rate assigned to the radio link set (UE) is the minimum of the bit rates allocated separately (for each cell) to all radio links of the active set. SHO gains may be taken into account in the E_b/N_0 requirements based on SHO condition.

The simulation flowchart for studying the downlink user data traffic in the radio access network is illustrated in Figure 3.8, where the aforementioned modules, explained in detail in the following sections, can be easily identified. Since the system in high traffic situations is downlink capacity limited (see, for example, [11] and/or [17]), the simulator supports only this direction. The structure to simulate the functions in the uplink would be exactly the same except that the transmission powers at base station (BS) would need to be replaced by the received ones. Terminals could be moved after the system power control. In this case, path losses and SHO conditions would be recomputed. The maximum resolution is one radio resource indication period (*RRI*), which is the time needed to receive the power measurements from the base stations.

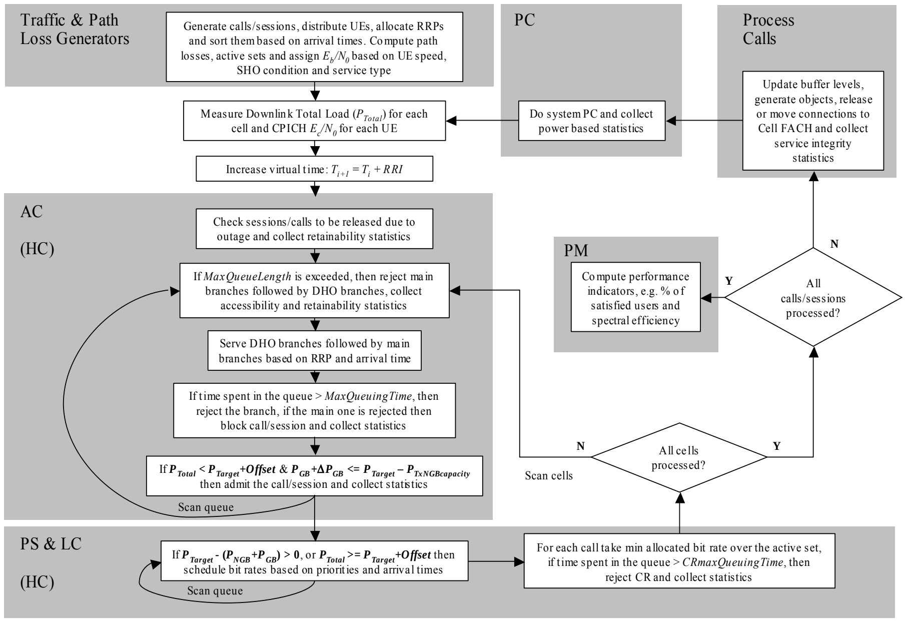


Figure 3.5. Simulation flow chart.

3.5.1.1 Traffic Generator

The traffic generator supports the models presented in Section 3.2. All call/session arrivals are generated at the beginning of a simulation. If an access to radio resources is denied, the corresponding call (or session) set-up is considered blocked with no possibility of retrying. Then according to the given traffic mix each connection is associated with a certain service, which in turn is mapped onto a corresponding QoS profile. The *share of calls* parameter defines the percentage of call arrivals of each of the offered services during the simulation period. All calls/sessions, generated at the beginning of the simulation, are subsequently processed (played back) taking into account the corresponding arrival times, service activities and priorities, hence the name *virtual time simulator*.

3.5.1.2 Path Loss Generator

After all calls have been generated, each of them gets assigned a random location across the simulation area. For the terminal positions, other distributions are possible as well, e.g. hot spots could be generated around strategic locations such as office parks, shopping malls or highways. For each mobile location, the received power levels from all cells are calculated first and then the cells satisfying the SHO conditions are assigned as active. Each base station can be configured separately in terms of location, sectorisation, antenna configuration (height, main lobe direction, tilt, beam width and gain), total transmit powers; transmit powers of P-CPICH and other common channels. For the path loss calculations, the following propagation models are supported: Okumura-Hata model as described in [80] and the models defined in [81]. By implementing an appropriate interface it is even supported to import the propagation calculated by another radio network planning tool. The latter option was used in the simulations described in this thesis. Correlated slow fading can be overlaid as described in [74]. In power computations, the effect of multipath fading is taken into account only in the required E_b/N_0 and orthogonality factor values. This assumption does not entail any limitation on the accuracy and precision of the performance results within the scope of this work.

3.5.1.3 QoS Management Functions in UTRAN

The supported RRM functions, i.e. AC, PS, and LC with QoS differentiation (priorities and differentiated parameters for different QoS profiles, or bearer service attributes), are described in Chapter 4. The algorithms are cell based, which means that the bit rates of the offered services are allocated and modified for each of the simulated cell independently, as illustrated in Figure 3.5. All other supported functions, i.e. Power Control, Performance Management and Process Calls, are system based as described in the following sections. A graphical representation of the AC, PS and LC function is illustrated in Figure 3.6, where the total downlink transmission power is reported as a function of the served guaranteed (GB) and non-guaranteed (NGB) traffic volume in the cell. From the figure, we can see that the GB traffic is admitted if the actual (measured) power, P_{GB}, exploited by GB services plus the power estimated for the bearer in question, ΔP_{GB}, does not exceed the maximum planned, $P_{TxTarget}$, minus the power dedicated to NGB traffic, $P_{TxNGBcapacity}$. New radio link requests are arranged into a queue and served following the strict priority principle (branch additions have top priority) and, at a given priority, based on their arrival times (FIFO). Diversity branches are not set up if the following condition is satisfied:

$$P_{GB} + \Delta P_{GB} > P_{TxTarget} + Offset\ , \tag{3.1}$$

where P_{GB} is the actual non-controllable power in the target cell, ΔP_{GB} is the estimated power increase in the same cell due to the radio link in question, and $P_{TxTarget} + Offset$ is the corresponding overload threshold. The NGB traffic is always admitted if the overload threshold, defined by $P_{TxTarget}$ + *Offset*, is not exceeded, and it is accommodated in the cell on the best effort (BE). In this case, PS relies upon the power left out by GB traffic and common channels (CCH). In other words, the power budget available for NGB traffic is given by the power target $P_{TxTarget}$ minus the power already in use (dissipated) in the cell. The tool supports two algorithms for bit rate scheduling, namely: Fair Throughput and Fair Resources scheduling. The former is more fair in terms of allocated bit rates among peers, regardless the distance of the terminal from the base station, whereas the latter, which takes into account the path loss at a given transmission power, is more spectral efficient (for more information see Chapter 4). In the case of SHO, the allocated bit rate is the minimum of the bit rates scheduled for each of the links of the radio link set. Overload control actions, which affect the bit rate of NGB services, are taken only if the overload threshold in that particular cell is trespassed. As already pointed out, the maximum resolution of the RRM functions is one radio resource indication period (RRI), which defines the time step between two power measurements at the base station (BS).

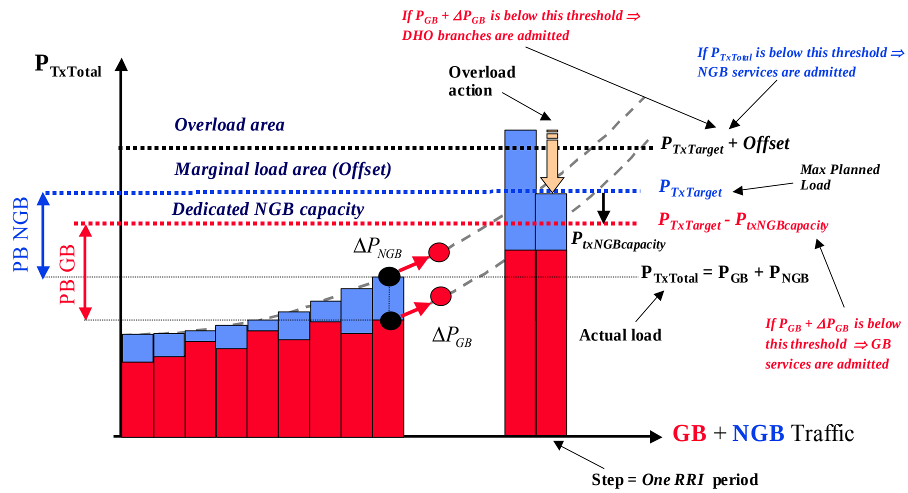


Figure 3.6. Graphical representation of QoS management functions supported by the simulator.

3.5.1.4 Process Calls (PrC) Function

All active calls in the system are processed together each radio resource indication period, as illustrated in Figure 3.8. The call processing consists of: Find the calls that are admitted or scheduled, check whether the connections are packet switched (PS) or circuit switched (CS) and process them accordingly.

If the ongoing connection is CS, the simulator collects its throughput, increases the active connections counter, and releases the communication in the case it has lasted longer then the corresponding call duration (see Table 3.1).

For each PS connection, we check firstly whether either the RNC buffer or the source buffer is not empty. If there is data to transmit, the active session throughput is collected and the active connections counter accordingly increased. Besides this, the status of the corresponding buffer in the terminal is updated and monitored. If during the ON period the buffer gets empty, the user of that service is considered unsatisfied and the re-buffering procedure is activated. When the user is reading (or the connection is in idle mode in the case of PoC), the transfer delay of the delivered object is calculated, the connection is marked as inactive and the inactivity period monitored. If the dwelling time of the DCH in question lasted longer than the corresponding inactivity timer, the terminal is moved to Cell_FACH state and the corresponding allocated resources released. When the reading time is over, either a new object to be downloaded and the corresponding reading time are regenerated, or the ongoing packet communication is released, depending on the session length (see Table 3.1). In the former case, if the time needed to fill up the buffer in the terminal is more than the corresponding buffering delay, the user of the service in question is considered unsatisfied and the related counter increased by one unit.

3.5.1.5 Power Control (PC) Function

The purpose of the standardised power control functionality in WCDMA is to keep the received signal to interference ratio in such a level that the link quality (block error rate, BLER) is kept at the target level. Ideally this means that the SIR and thus the received E_b/N_0 is constant during the communication using a particular application. With the fast power control scheme implemented in WCDMA this is approximated fairly well. In fact, with an update-rate of 1500 Hz, i.e. once per slot, a step size of 1dB can effectively track a typical Rayleigh fading channel up to a Doppler frequency of about 55Hz, which corresponds to a mobile speed of about 30km/h [11].

The simulator supports an ideal power control function that, within limits (PC dynamic range), compensates effectively for the slow and fast fading of the propagation channel. Yet, for the sake of simplicity, without losing generality, the models for the path losses adopted in this work, and presented in Section 3.4 and 3.5.1.2, do not include fast fading. The multi-path fading and SHO effects are taken into account in the E_b/N_0 requirement for each of the supported services, which yield sufficiently accurate and precise results for the scope and purpose of this work (see Section 3.7). The power control function also includes a realistic model for interference: At any simulation time step, the fully received power from all cells except from the best server is counted as interference, and hence the corresponding coupling effect is fully taken into account. From the best server only the fraction determined by the non-orthogonality is considered as interference.

The power control system of equations can be written as:

$$\frac{Wp_{i_m} / L_{m,i_m}}{R_{i_m} P_m / L_{m,i_m}\left(1-\alpha_{i_m}\right) + \sum_{n,n\neq m} P_n / L_{n,i_m} + N_{i_m}} = \rho_{i_m}, \quad , \qquad i_m \in I(m), \quad m = 1,\ldots,M \tag{3.2}$$

where the symbols in (3.2) are explained in the following table.

Figure 3.7 depicts the situation where the $MS_{i,m}$ is served by BS_m and the BS_n adds a component to the interference experienced by the mobile station. Equation (3.2) simply equates the received E_b/N_0 with given transmission powers to the required E_b/N_0 needed for sufficient quality of the connection.

Table 3.3. Symbols in the power control system of equations.

Symbol	Explanation
i_m	Index of a MS served by BS m
m, n	Indices of BSs
$I(m)$	Set of MS indices served by BS m
M	Number of cells
p_{i_m}	BS transmitted power for MS i_m
P_m, P_n	Total transmit power of BS m and BS n

Symbol	Explanation
L_{m,i_m}	Pathloss from BS m to MS i_m served by BS m
L_{n,i_m}	Pathloss from BS n to MS i_m served by BS m
R_{i_m}	Bit rate used by MS i_m
α_{i_m}	Orthogonality factor for MS i_m
N_{i_m}	Noise power (thermal plus equipment) of MS i_m
ρ_{i_m}	Required E_b/N_0 for MS i_m

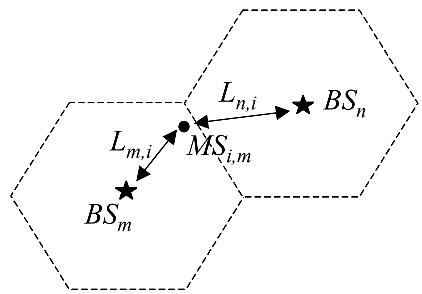


Figure 3.7. Mobile station $MS_{i,m}$ served by BS_m and interfered by BS_n.

Taking into account that

$$P_m = \sum_{i_m \in I(m)} p_{i_m} + p_{c,m} , \tag{3.3}$$

where $p_{c,m}$ is the sum of common channel powers from BS_m, (3.2) can be rewritten in the compact formant as a M times M linear system of equations of the type $Ax = b$, where the unknowns are the total BS powers of each BS. During the simulations, we firstly solve the linear system and then from (3.2) we derive the individual link powers. The solutions are used to estimate the transmission power of GB and NGB services in (4.1), (4.2), (4.3), and (4.5). AC and PC functions ensure that the BS total transmission power is kept below the maximum and the WCDMA pole capacity is not exceeded. The existence of solutions to the type (3.2) of equations was studied e.g. in [82]. The power of common channels is a cell based management parameter (see Section 3.3).

3.5.1.6 Performance Monitoring (PM) Function

Several performance indicators can be collected during the *measurement period* (simulated time), e.g. *call block ratio* (CBR) caused by queuing and/or buffer overflow, *call drop ratio* (CDR) due to power outage, *active session throughput* (AST) as defined in [74], *capacity request rejection ratio* (CRRR) for NGB traffic, *object transfer delay* for Browsing, MMS and dialup connections, and *re-buffering* for streaming, PoC and SWIS applications.

Link and *cell based powers*, as well as E_c/N_0 (energy per bit per noise spectral density) measurements on common pilot channel (CPICH) are also computed during the simulated time. From such measurements the *geometry factor* (*G*) is derived. *G*, defined as the ratio between the received power from the serving cell and the received power from the surrounding cells plus noise, is used to represent the distance from the base station, as explained in [11].

All above performance metrics may be displayed on the digital map to show where the most critical areas, such as hot spots or low coverage zones, are.

3.5.2 Simulator for WCDMA Radio Interface Dimensioning (Max)

The simulation flow chart is illustrated in Figure 3.8. As can be noticed in the figure, the simulator consists of four modules and two loops, denoted by inner and outer loop in the following. The first module is the call generator, which is described in Section 3.5.2.1, followed by an admission control (AC) and packet scheduler (PS) function, which are presented in Section 3.5.2.2. The last block has been designed to collect performance statistics, as explained in Section 3.5.2.3. The inner loop constitutes a computational procedure in which a cycle of operations (new configuration of active users) is repeated to approximate the desired result more closely. In other words, the more iterations are allowed, the more accurate the performance results are. In fact, the tool, being event based (from the active connections viewpoint), during each computation (iteration) analyses a snapshot of the cell status, where more users are simultaneously active depending on the corresponding probability to generate a call (use a certain service). The outer loop may be enabled to increase the offered traffic (load) till at least one of the thresholds defined by the percentage of dissatisfied users of that particular service is trespassed. When such a condition is satisfied, the simulation ends and the offered and served traffic, together with the average cell throughput, are collected, as described in the following sections.

3.5.2.1 Call Generator

For each of the services presented in Table 3.1 defines a different *QoS profile*, *mean service time*, *mean arrival rate* and *share of subscription*. In this case, a user is supposed to make a call every mean arrival rate seconds, and the corresponding connection may last on average mean service time seconds. The ratio between the two corresponding parameters yields the erlangs per user for that particular service type, whereas the share of subscriptions defines the percentage of users that have contracted to receive and pay such a service.

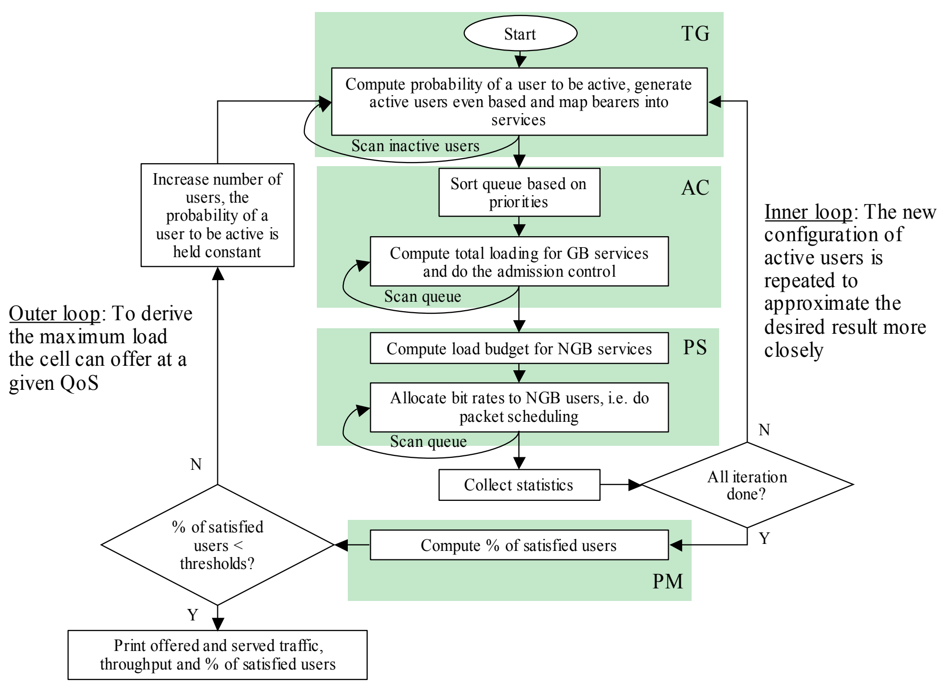


Figure 3.8. Cell based simulation flow chart.

Hence, during a snapshot of the cell status, the probability P_i of a user to make use of a certain service type i, when the subscribers are scanned for the first time, is given by

$$P_i = \frac{T_i}{A_i}\frac{S_i}{100}, \tag{3.4}$$

where, T_i is the mean service time, A_i the mean arrival rate (reciprocal of the mean arrival intensity, denoted by λ in the literature) and S_i the share of subscription to that particular service type. For PoC, the possibility of a user to communicate simultaneously in the downlink to more peers is taken into account by introducing the group factor in (3.4). In this work, the group factor follows a geometric distribution with a cut-off: The average group size is supposed to be composed by 4 people, and the minimum and maximum crowd size is assumed to be 1 and 25 users, respectively. Furthermore, during the inner iterations, the probability of a user to make a new call is conditioned by the fact that such a subscriber may have already established one or more connections and (3.4) is accordingly corrected. Ultimately, when the simulation is over, e.g. the condition to exit the outer loop of Figure 3.8 is satisfied, for each of the services deployed in the cell, the maximum offered load is estimated as follows:

$$N_i = U_i \cdot \frac{A_i}{T_i}, \tag{3.5}$$

where U_i is the average number of the active bearers carrying the service type i, A_i and T_i are as defined in (3.4), and N_i denotes the offered traffic in number of subscribers per cell.

3.5.2.2 Radio Resource Management Functions

The simulator supports only an admission control (AC) and packet scheduler (PS) function, as illustrated in Figure 3.8. The connection requests are arranged into a queue and served following the strict priority principle by the same standard as presented in Section 3.5.1.3. The lower the associated priority value, the higher the priority of the service in question when processed by the radio resource management functions. Guaranteed bit rate (GB) services are not admitted if either the total load in the cell, L_{Total}, exceeds the overload threshold, L_{Target} + *Offset*, i.e.

$$L_{Total} = L_{NGB} + L_{GB} > L_{Target} + Offset, \tag{3.6}$$

or the load increase ΔL_{GB} due to the bearer in question (if admitted), plus the actual non-controllable (GB) load L_{GB}, goes beyond the target load L_{Target} planned for the cell, i.e.

$$L_{GB} + \Delta L_{GB} > L_{Target}, \tag{3.7}$$

The NGB traffic is admitted if (3.6) is not satisfied, and bit rates are allocated using the Fair Throughput (FT) scheduling algorithm presented in Section 4.2.2.1. The load budget LB_{NGB}, which is shared fairly in terms of allocated bit rates between peers, is given by:

$$LB_{NGB} = L_{Target} - (L_{NGB} + L_{GB}), \tag{3.8}$$

The load increase and decrease computations are based on the downlink fractional load equation presented in [11], which, for one radio bearer service (k), reduces to:

$$\eta_{DL}^{k} = \frac{1+SHO}{W} \cdot \rho_k \cdot R_k \cdot \nu_k \cdot \left(\left(1-\alpha_k\right) + i_{k,DL} \right), \tag{3.9}$$

where SHO is the soft handover overhead, W is the chip rate, ρ_k is the required E_b/N_0 of the service in question, R_k the transport channel user bit rate, ν_k the service Activity Factor (AF), $i_{k,DL}$ the other to own cell interference ratio, and α_k is the orthogonality factor, which ranges from 0 to 1, depending on multipath conditions ($\alpha = 1$ means perfect orthogonality). The downlink load factor (L_{DL}) is given by [11]:

$$L_{DL} = \sum_k \eta_{DL}^{k} = P_{TxTotal} / P_{TxMax} \,, \tag{3.10}$$

where $P_{TxTotal}$ and P_{TxMax} denote the actual and maximum transmission power of the cell.

3.5.2.3 Performance Monitoring and Exit Condition

The performance indicators collected by the simulator are call block ratios and throughputs, based on which the following user satisfaction criteria are defined. Circuit Switched (CS) speech and video calls are satisfied if they do not get blocked. Their guaranteed bit rates (GB) are 12.2 and 64 kb/s, respectively. The same criterion is applied to SWIS, whose target bit rate (GB) is 64 kb/s. For NGB services, in addition to the above condition, the user satisfaction criteria take into account the average allocated bit rates. In particular: PoC, streaming, MMS, Dialup and WAP users are satisfied if the average bit rate during the iterations (inner loop) is higher or equal to 8 (or 16), 64, 32, 64 and 32 kb/s, respectively.

If the outer loop is enabled, the simulation is over when at least one of the following conditions results true (see Figure 3.8):

- Less than 90% of speech users are satisfied.
- Less than 90% of video users are satisfied.
- Less than 90% of streaming users are satisfied.
- Less than 70% of MMS users are satisfied.
- Less than 90% of PoC users are satisfied.
- Less than 90% of SWIS (RTVS) users are satisfied.
- Less than 50% of dialup users are satisfied.
- Less than 70% of WAP users are satisfied.

(Note: The above values are used in the case studies presented in Section 5.3, in general are parameters for the radio network planner to set.) This means that the maximum load that can be served in the cell, at a given QoE, and share of subscriptions, is achieved and the offered traffic from (3.5) can be accordingly derived. The main output of the simulator is thus the maximum number of subscribers a WCDMA cell can accommodate satisfactorily.

Besides this, the simulator may be used for studying the impact of a new service on the existing subscribers satisfaction. This can be done increasing gradually the subscription level for the new service while keeping the input load of the other services constant.

3.6 Performance Measures

In this work, system performance is assessed as recommended in [74], and for this purpose tailored *user satisfaction* criteria are defined.

A speech user is satisfied if the call neither gets blocked nor dropped during the call duration period. A call can be blocked in the case of AC queue overflow, or if either one of the two conditions reported in Section 4.2.1 is satisfied. A connection can be released (dropped) for coverage or quality reasons. In the former case, the received CPICH E_c/N_0 needs to be below – 18 dB. In the latter case, the radio link power needs to exceed its maximum physical value (40 dBm, i.e. 3 dB below the maximum base station transmission power, see [11]) for more than 5 s. This means that the power control headroom is not enough to maintain the quality of the communication as close as possible to the required level, which is represented by the E_b/N_0 values reported in Table 3.2. In addition to this criterion, for PoC, SWIS and streaming users no re-buffering in the terminal is allowed during the communications and the time to fill it up (denoted by *buffering delay*) needs be reasonably short; for Dialup (http, emails, ftp, etc…), WAP browsing, and MMS, the active session throughput has to be higher than 64 kb/s, 32 kb/s and 8 kb/s, respectively. Furthermore, any of the capacity requests (CRs) of NGB services must not be rejected in the packet scheduler queues.

The *spectral efficiency* is computed as the system load (average cell throughput normalised with respect to the chip rate, 3.84 Mchip/s) at which a certain percentage of users of the worst performing service is satisfied. Different thresholds can be set for the distinct bearers, though 90% is the default value in this work for all applications. Figure 3.9 presents an example of spectral efficiency gain computation for three distinct bearer services using the same quality limit (90% of satisfied users). The continuous and dashed lines are the performance results collected before and after the enforcement of a change in the parameter settings or supported QoS management functions in UTRAN. The spectral efficiency gain (or loss) is the difference between the two spectral efficiency values before and after the change in the configuration or algorithm of the analysed network. The more traffic that can be handled, at a given QoE (= % of satisfied users), for a given frequency band, the more efficiently the spectrum is utilised by the operator.

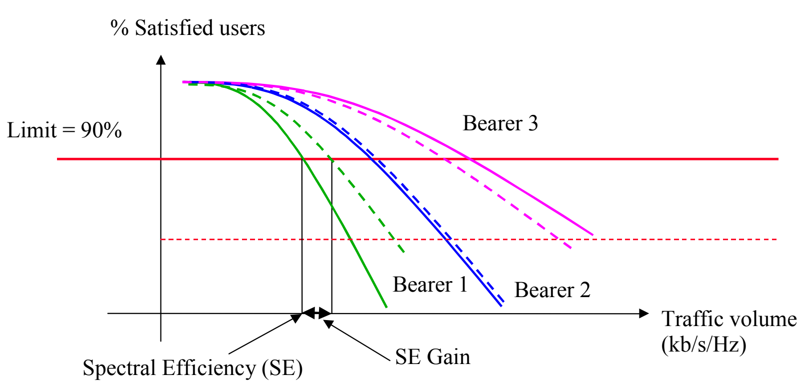


Figure 3.9. Example of gain computation in terms of spectral efficiency at a given QoE (≥ 90% of users satisfied).

3.7 Verification of the Virtual Time Prediction Method

In this section, the virtual time prediction method (Ares) described in Section 3.5.1 is compared to dynamic analysis results using the simulator (Wallu) presented in [16]. Based on this comparison it is demonstrated that the results attained with Ares are accurate enough for assessing UTRAN system performance. The same methodology was adopted in [83] to verify the static tool for WCDMA radio network planning presented in [10].

3.7.1 Introduction

Ares is a virtual time WCDMA network simulator designed for studying and validating QoS management functions before their deployment in real networks (see Chapter 4). Ares is also used for service driven WCDMA radio network planning and for thus studying how different traffic scenarios and/or network configurations can influence the QoE of offered services (See Chapter 5 and 7). Wallu is a dynamic WCDMA network simulator designed for research and development of all diverse WCDMA system features, excluding possibly only higher layers and the radio link level. Wallu is a complex but extensively tested tool whose results are trusted and used in standardisation, network dimensioning, and feature marketing. Ares is a much-simplified simulator implementing only basic models and crucial features needed in the QoS management analysis, aiming to provide a good trade-off between simulator simplicity (for improved speed and maintenance) and generalisability of results to real networks.

The present study aims to illuminate the extent that Ares results can be generalised to real networks. The assumption is that Wallu provides reliable results. The task is carried out running the same network scenario with Ares and Wallu and comparing the obtained results for their similarity. The comparison is performed in the downlink performance, since Ares implemented only this direction when the tool was validated.

3.7.2 Feature Comparison Between Ares and Wallu

In the following, a high level description of the functions and models supported by the simulators is given.

3.7.2.1 Traffic Generator, Models, and Mix

Call and session arrival processes, mapping to QoS profiles, and distributions of traffic elements such as object sizes, reading times, and session lengths can be parameterised in the similar way in Ares and Wallu. Wallu does not readily model the inverse Gaussian distribution, which Ares does. Thus, in Wallu a geometric model was used instead.

3.7.2.2 Path Loss

Ares and Wallu can both import propagation models calculated with other tools. Ares and Wallu also support correlated slow fading.

Wallu supports multi-path profiles with different path gains for radio propagation in downlink. Signals from the same base station cell propagating along the same path are totally orthogonal; that is, they do not interfere with each other. Signals propagating along different paths are totally unorthogonal. The signals from different paths are combined in the mobile. The effect of multi-path propagation is incorporated in Ares with a single parameter referred to as the orthogonality factor (α), ranging from zero to one, and in the E_b/N_0 requirements, which take into account the UE speed and SHO conditions.

Wallu allows addition of fast fading process and shadow fading process with specific mean and deviation to the path loss, which are not modelled in Ares.

3.7.2.3 Admission Control

Wallu and Ares apply similar methods in the admission control and in the estimation of downlink power due to the admitted bearer. However, Ares admits all NGB bearers if the overload threshold is not exceeded, whereas Wallu requires that a spreading code is also available in the code tree.

3.7.2.4 Handover Control

Wallu is a dynamic simulator with moving users making hard, soft, and softer handovers. Ares does not support mobility of users. Ares includes a parameter for defining the soft-handover offset, which is used to scale up the E_b/N_0 requirements in admission control and packet scheduling formulas.

3.7.2.5 Packet Scheduler

Wallu and Ares apply similar methods in the computation of allowed NGB power that can be allocated at each packet-scheduling period. However, there are some differences. Wallu limits the change in the total downlink transmission power. In one scheduling period, the power increase is limited to 1.6 dB by default. Ares does not limit the power change.

Ares allocates immediately the highest possible bit rate, whereas Wallu allocates the *minimum allowed bit rate* first (see Chapter 4), unless the data amount in the RLC exceeds a specific limit (1024 octets by default). Higher bit rates are requested and possibly upgraded once the buffer limit is exceeded.

Ares estimates the downlink power increase due to a particular bearer with an initial power formula that applies the received CPICH E_c/N_0, orthogonality factor, and the total downlink transmission power. In Wallu, the allocation of the initial minimum bit rate is based on the same formula. However, the following bit rate upgrades are based on scaling the average link power with the ratios of new bit rates and E_b/N_0 requirements to old ones.

3.7.2.6 Power Control

Ares simulation cycle equals to the radio *resource indication period* (200 ms by default). Wallu simulation cycle is one slot (0.7 ms). The power controls in Ares and Wallu are thus quite different. Wallu uses fast closed-loop power control to keep signal strengths at the required levels. Moreover, the required levels are adjusted with an outer-loop power control to meet the requirements of block error rates. Ares supports an ideal power control function that includes the effects of a large-scale propagation channel, but not fast fading. Multi-path fading and SHO effects are taken into account in the service E_b/N_0 requirements. In this simplified tool, however, the interference is realistically modelled: At any simulation time step, the received power from all cells except from the best server is counted as interference, and hence the corresponding coupling effect is fully taken into account. (Note: Form the best server only the fraction of the power determined by the non-orthogonality is considered as interference.) During each radio resource indication period, Ares resolves the system of equation that equates the received E_b/N_0 with given transmission powers to the required E_b/N_0 for sufficient quality of the connection from which derives the individual radio link and cell powers. The solutions are then used in all formulas to estimate the transmission powers of GB and NGB services (See Section 3.5.1 and Chapter 4 for more information).

3.7.2.7 Common Channel Powers

The pilot power is the only common channel power modelled with Wallu. Ares allows power setting for other common channels as well.

3.7.3 Simulated Scenario

In the following sections the traffic profile, simulated scenario and network parameters are presented.

3.7.3.1 Traffic Profile

The traffic scenario included services from all traffic classes with different QoS profiles. As shown in Table 3.4, NGB services were run on Interactive and Background classes. GB applications were offered on Conversational and Streaming classes with fixed bit rates. The bit rates of NGB services could vary as shown in the table. The range is however typically limited between the *minimum allowed bit rate* and *maximum allowed bit rate* (see Section 4.2.2, for parameters description), which in this case were set equal to 64 kb/sand 384 kb/s, respectively. Services of Interactive traffic class were divided into three different THP priorities. THP1 has the highest priority in resource allocation by PS. The background service Dialup is also referred to as THP4 service in this work. In the packet scheduler queue, PoC is served first, followed by the streaming application, and WAP & MMS. Dialup connections, holding the lowest priority, are scheduled last.

Table 3.1 and Table 3.4 show the specific traffic characteristics of the services. (Note: In Wallu, the Dialup session length, in number of objects, was modelled with a Geometric distribution, Mean = 3.86, Min = 1, and Max = 50.) Data stream rate refers to the constant application generated bit rate of services such as speech, video or streaming applications. The streaming-type services apply a buffer in the terminal that stores data to compensate for varying packet delays. Different services require buffers of different sizes. If the buffer gets empty during transmission, the effect is shown as degraded application quality. The unit over which the QoS was monitored is referred to as session, which consisted of one or several objects (voice or data) separated by inactivity periods (reading time).

Table 3.4. Mapping of services onto QoS profile, bit rates, and traffic mix.

QoS Profile	Service	Bit rates (kb/s)	Share of Calls (%)
CS Conversational	Speech Video	12.2 64	20 7
PS Streaming	SWIS (RTVS)	64	10
PS Interactive THP1 THP2 THP3	 PoC Streaming application WAP/MMS	 0, 64 0, 64, 128 0, 64, 128, 144, 256, 384	 18 12 13/5
PS Background	Dialup	0, 64, 128, 144, 256, 384	15

Table 3.1 shows the distributions of the number of objects or time duration in the session (session length), the object size, and the reading time for each service. The distributions were based on studies conducted on service quality published in international forums or on educated guess (see Section 3.2). In Table 3.4, the share of calls gives the probability of a specific service once a new call/session was generated. The traffic mix was selected to generate a traffic volume considerably large for each of the deployed services.

Once a session was completed, the satisfaction data of the user was collected. Different services had different satisfaction criteria (see Section 3.6). Firstly, a user was considered unsatisfied if the call was blocked or dropped. The blocking also covered those packet-switched services whose capacity requests were rejected by the PS. Secondly, users running streaming-type applications controlled by the PS were considered unsatisfied if the terminal streaming buffer run out during the session. Finally, users of other PS controlled services were unsatisfied if the active session throughput did not meet specific levels shown in the table.

3.7.3.2 Simulation Setup

The simulators modelled a 19-sector Helsinki macro scenario, as shown in Figure 3.2. A set of mobile terminals was deployed uniformly in the area without any restrictions with streets, buildings, or water. The mobiles were static with Ares but moved walking pace with Wallu.

The main parameters used in the simulations are presented in Table 3.5. The channel multi-path profile was that of ITU Vehicular A with 5-paths propagation. The path gains with Wallu are shown in the table, and they produced 60% orthogonality factor on the average. However, the planned orthogonality factor applied by the downlink initial power estimation was set to 50%. Ares just assumed that 50% of the base station power was interference. The propagation loss model was imported from data generated with the tool presented in [10], named *NPSW*, and attached in [11]. The slow fading from different sites was not correlated. Zero-mean shadow fading with 8 dB deviation and Jakes fast fading were added to the process with Wallu.

The E_b/N_0 values applied to power estimation were the same for all bit rates with one service (see Table 3.2). In practice, higher bit rates require lower E_b/N_0 values than lower bit rates.

3.7.4 Comparison Results

In this section the corresponding Ares-Wallu performance results are discussed.

3.7.4.1 Offered Load in Call Arrivals

The upper part of Figure 3.10 shows the numbers of generated calls/sessions on a per-cell basis during the simulations with Ares and Wallu. The differences between the call numbers were small, which indicates that the user distributions and cell sizes were similar. The lower part of the figure shows how the calls are distributed into the services. The figure verifies that the call generators were configured in the same way with the simulators. The small differences in the call counts can be regarded as random.

Table 3.5. Selected parameters used in the simulations.

Parameter	Value	
Call/session mean arrival rate	0.5 s (0.72 calls/h/mobile)	
Radio resource indication period (RRI)	0.2 s	
Simulation time (s)	3 600 s	
Number of mobile stations	10 000	
BTS total power	20 W	
Power target for DL AC	3 dB below BTS total power	
Overload offset for DL AC	1 dB above power target	
Orthogonality in RRM functions	0.5	
Period for load control actions	0.4 s (2 RRI)	
Period for Packet Scheduling	0.2 s (1 RRI)	
Inactivity timer with PS	5 s	
Maximum queuing time with PS	4 s	
Minimum allocation time with PS	15 s	
Maximum BTS Tx power	43 dBm	
P-CPICH Tx power	33 dBm	
Parameter	**Ares**	**Wallu**
Sum of all other CCH Tx powers	30 dBm	0 dBm
Multi-path propagation gains	-	51, 30, 11, 6, 3%
Mobile station speed	0	3 km/h
Correlated slow fading	Not used	

3.7.4.2 Served Load

Figure 3.11 shows the average number of (active) users (connections) per service and cell. In Wallu, the average numbers were about two users higher than in Ares, which were mostly due to the PoC, the streaming application, and WAP users that were not served as quickly in Wallu as in Ares.

3.7.4.3 Cell Throughput

Figure 3.12 shows the average throughput in each cell. The throughputs in Wallu were lower than those in Ares. The difference was mainly in the Dialup service, most likely due to different traffic models, whose average throughput in Ares was up to 200 kb/s (Cell 10) higher than in Wallu.

3.7.4.4 Load Status in Cell 11

As illustrated in Figure 3.10-Figure 3.12, Cell 11 was one of the most loaded cells. Figure 3.13 shows the Cell 11 total transmission power, $P_{TxTotal}$, and the power for NGB users, P_{TxNGB}, during the last 1000 seconds of the simulation with Ares (upper) and Wallu (lower). The mean P_{TxNGB} was 53% of $P_{TxTotal}$ with Wallu. With Ares, $P_{TxTotal}$ was closer to the target, which was likely due to the 1 W higher common channel power. The ratio of Ares P_{TxNGB} to $P_{TxTotal}$ was similar to that with Wallu. Figure 3.14 shows the histograms of P_{TxNGB} and

$P_{TxTotal}$ during the entire simulated period. The modes of powers with Ares were about 1 W higher than those with Wallu, which the 1 W higher CCH power explains.

3.7.4.5 Satisfied Users

Figure 3.15 shows the ratios of satisfied users for each service and cell. The average number of satisfied users per cell for each of the analysed services is illustrated in Figure 3.16. Performance results were consistent and similar trends were shown for each cell and deployed service. In particular, Streaming and Dialup were the poorest performing services, followed by PoC, WAP and MMS. GB services performed best. However, in high loaded cells, especially for long packet calls, Ares produced better performance then Wallu. This is due to the differences in propagation channel modelling, power control, power estimates and bit rate allocation in packet scheduling, and the fact that terminals with Ares are not moving. The single performance indicators are discussed in the following sections.

3.7.4.6 Call Block Ratio

The call block ratio (CBR) for each service and simulated cell is depicted in Figure 3.17. In high loaded cells, the CBR experienced with Wallu was higher than in Ares. Performance and trends were however similar. GB services, such as speech, video and SWIS calls, performed worst. Among the NGB call attempts, Dialup and WAP were the most blocked.

3.7.4.7 Active Session Throughput

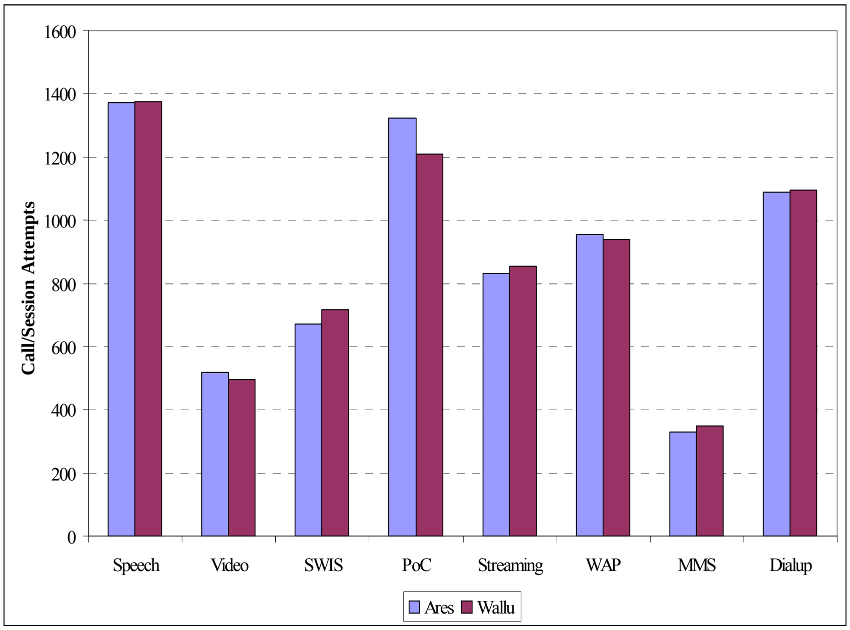
Figure 3.18 shows the 10th percentile of the active session throughput per service and cell indicating the performance of the worst decile at best. With Wallu, many cells showed very low throughput for all packet-switched connections. This was mostly due to capacity rejections, delays in transport channel and bit rate allocations, and packet retransmissions in bad channel quality conditions. Overall, the throughputs were lower than those with Ares. The Ares averages over the cells were over 50 kb/s higher for WAP and MMS, 30 kb/s higher for Dialup, 2 kb/s higher for streaming and PoC applications, than those with Wallu.

3.7.4.8 Capacity Request Rejection Ratio

Figure 3.19 shows the capacity request rejection ratios (CRRRs) per service and cell with Ares (upper) and Wallu (lower). The ratios were similar from service to service. For PoC, WAP&MMS the rejection ratios with Ares were some points higher than with Wallu. This is most likely due to the higher bit rates allocated with Ares that for small amount of data to transmit increased the number of transitions from Cell_FACH to Cell_DCH, and thus the probability of being blocked in the packet scheduling queues. The CRRR for Dialup and streaming was fairly better with Ares than with Wallu.

3.7.4.9 Re-Buffering Ratio

The re-buffering ratio (RBR) for SWIS, PoC and streaming services for each of the simulated cells is illustrated in Figure 3.20. The trends and performance numbers of the corresponding curves attained with Ares (upper) and Wallu (lower) are similar. The RBR for SWIS, being offered with guaranteed bit rate, was zero with Ares, whereas with Wallu some of the users in the most loaded cells experienced bad quality. Except for Cell 3, where the RBR for streaming with Ares was much higher then in Wallu, the performance of the two simulators matched quite well. For PoC, the RBR was higher with Wallu than with Ares. This could be explained considering the lower active session throughput with Wallu and the small buffer size in the terminals (one second for PoC).



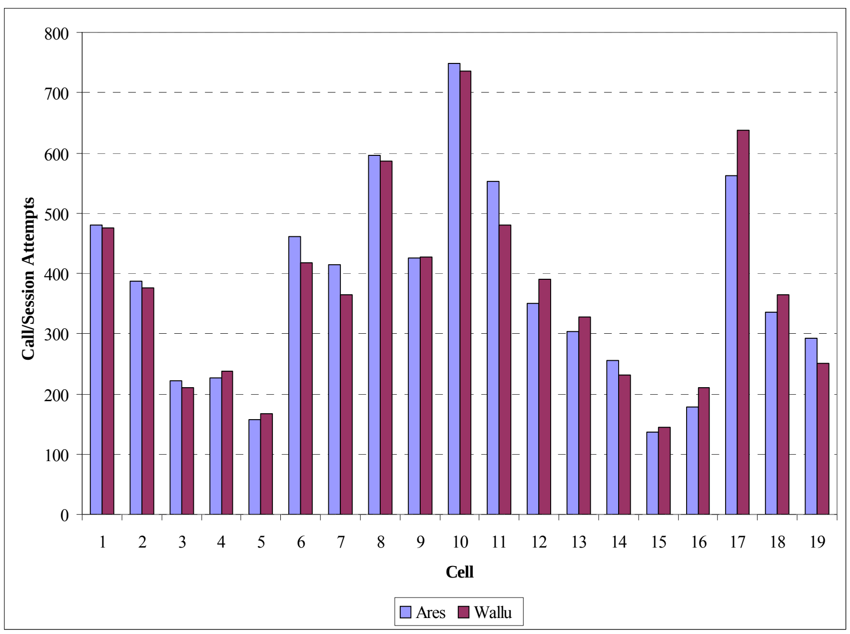
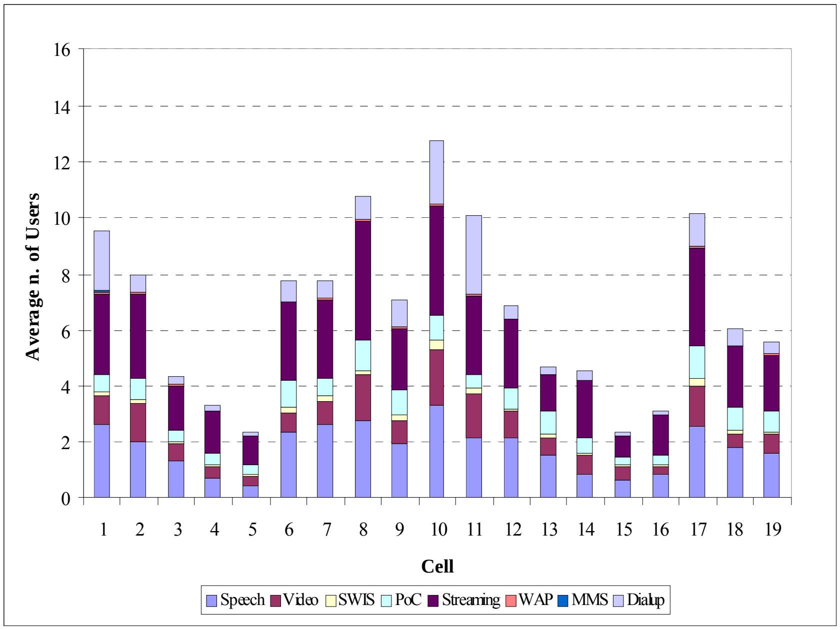


Figure 3.10. Number of call/session attempts per service (upper) and per cell (lower).



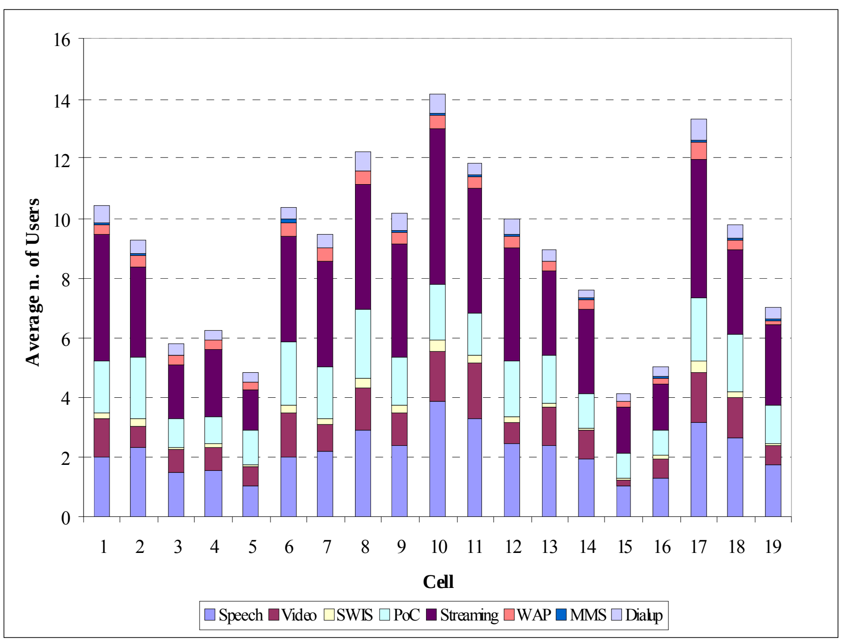
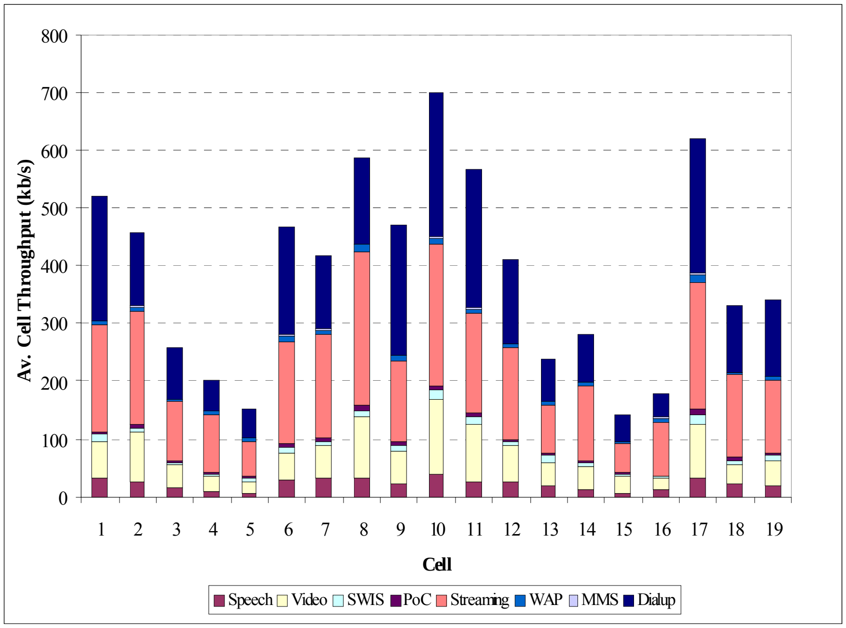


Figure 3.11. Average number of users per service and cell in Ares (upper) and Wallu (lower).



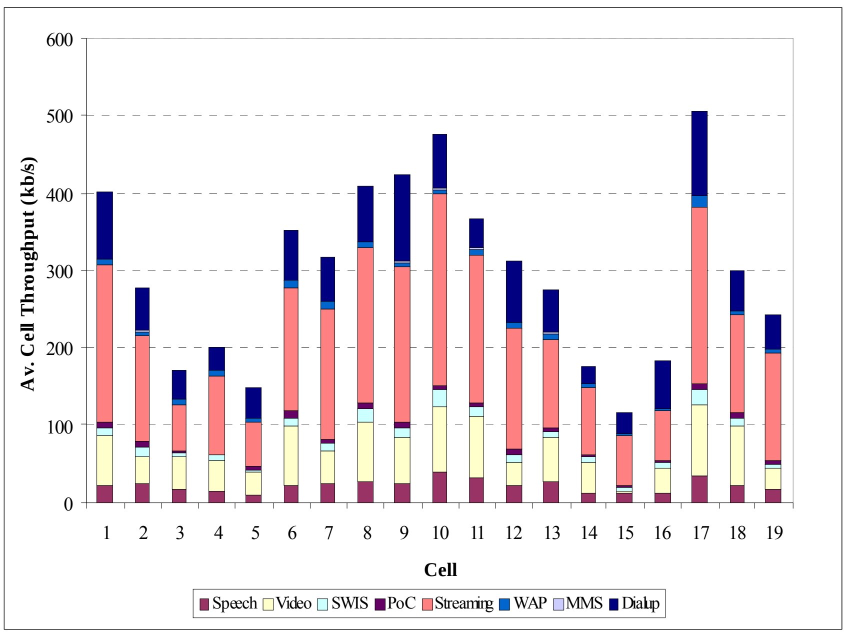


Figure 3.12. Average cell throughput per service in Ares (upper) and in Wallu (lower).

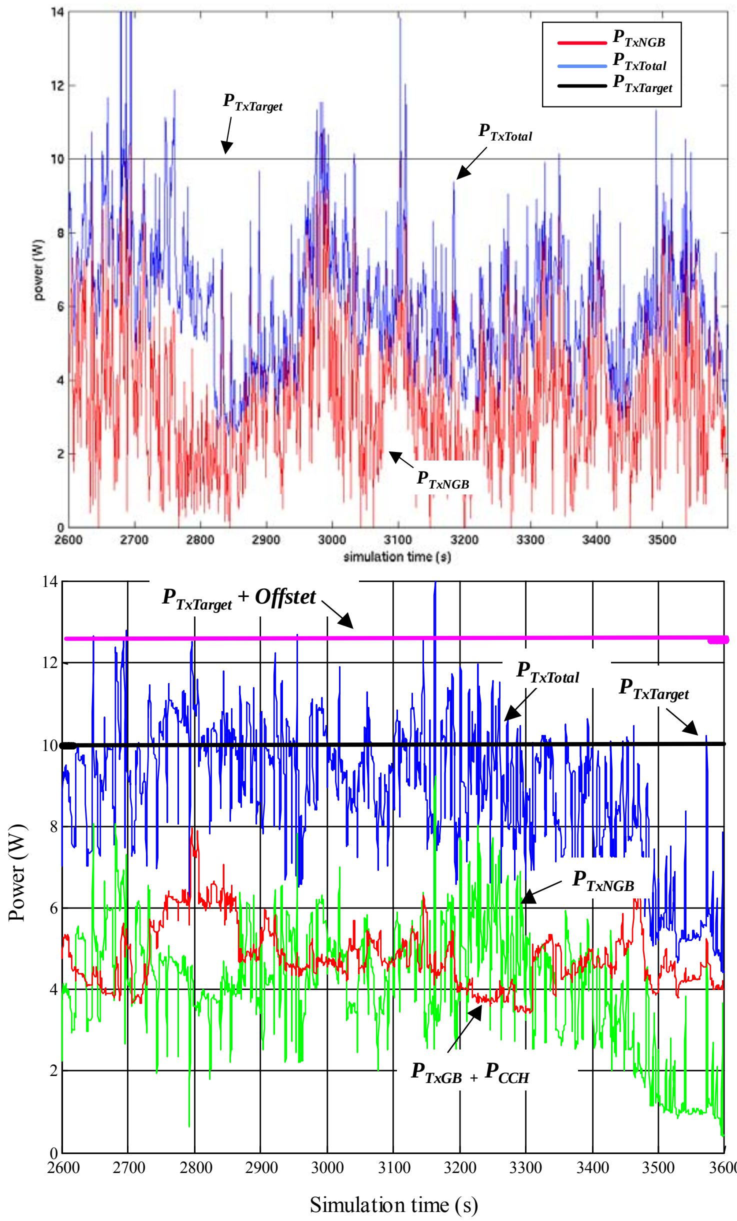


Figure 3.13. Load status in Cell 11 during the last 1000 seconds of the simulation with Ares (upper) and with Wallu (lower).

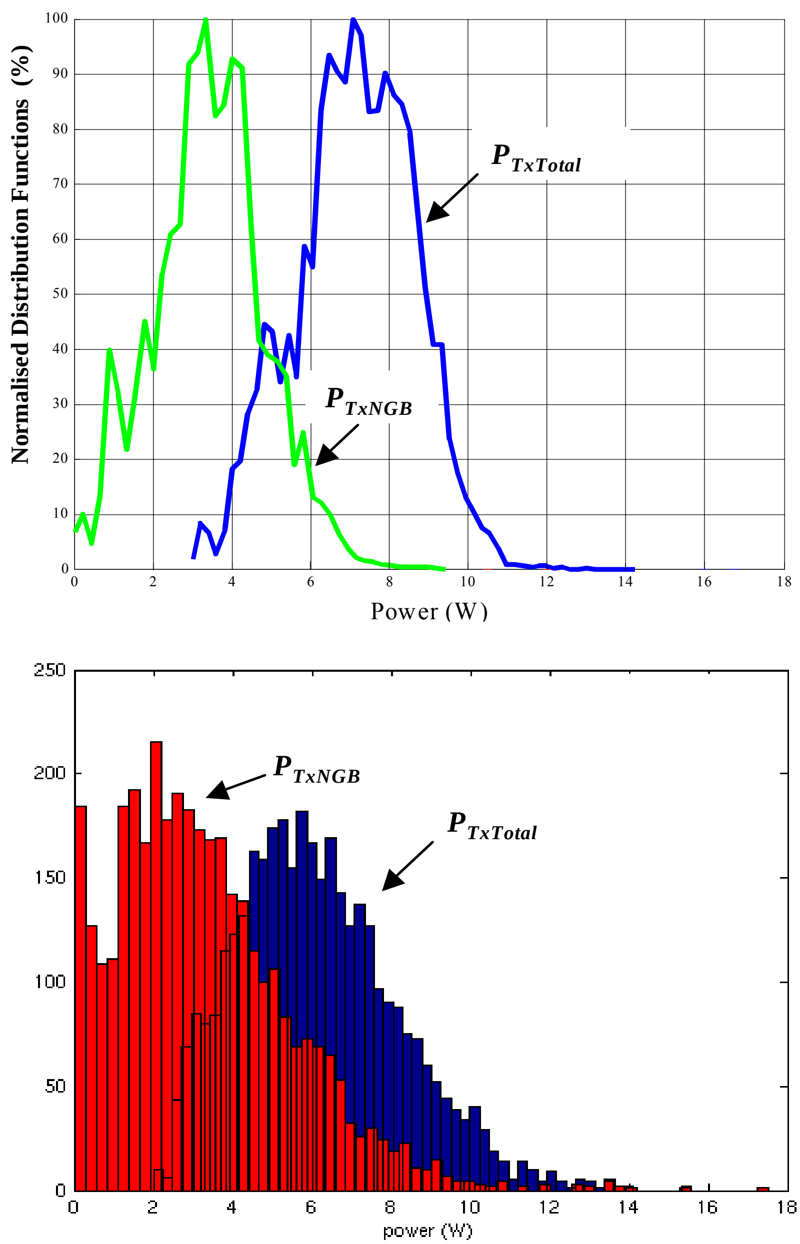
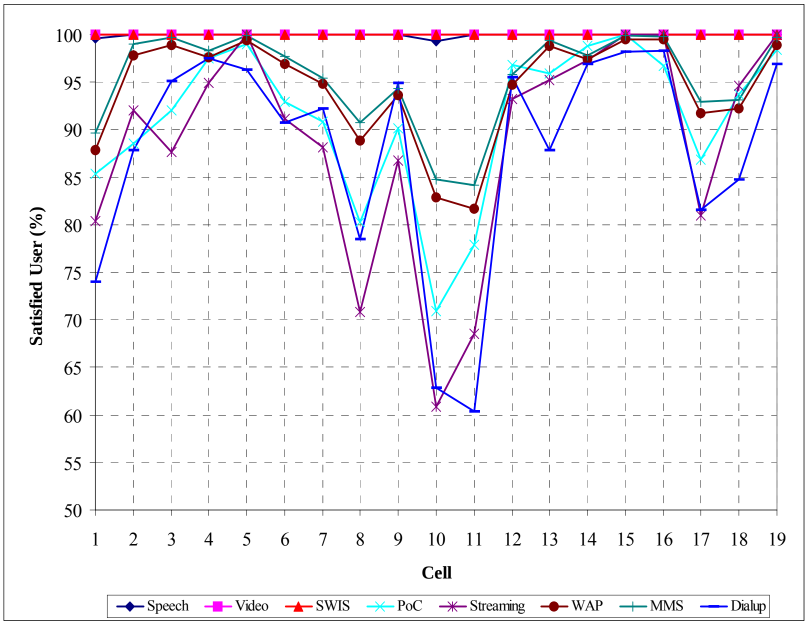


Figure 3.14. Histograms of P_{TxNGB} and $P_{TxTotal}$ during the last 1000 seconds of the simulation with Ares (lower) and with Wallu (upper).



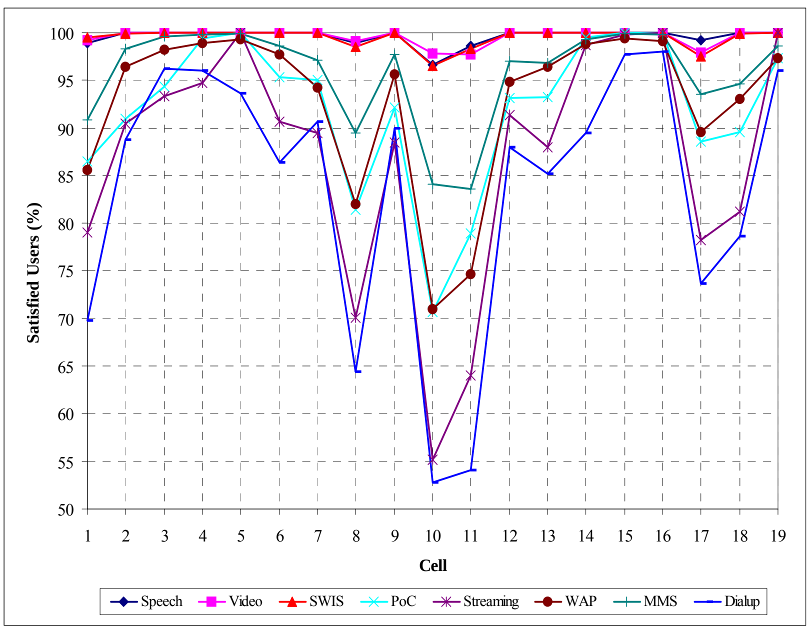


Figure 3.15. Ratio of satisfied users per service and cell with Ares (upper) and Wallu (lower).

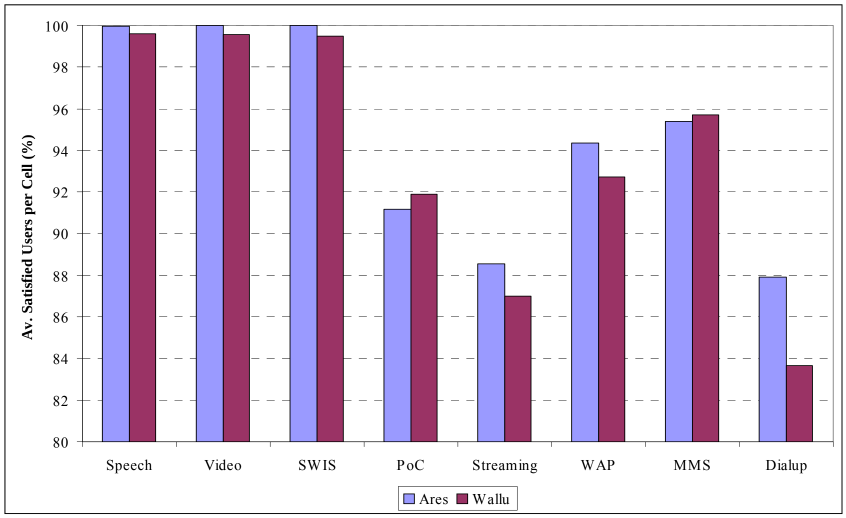


Figure 3.16. Average ratio of satisfied users per service.

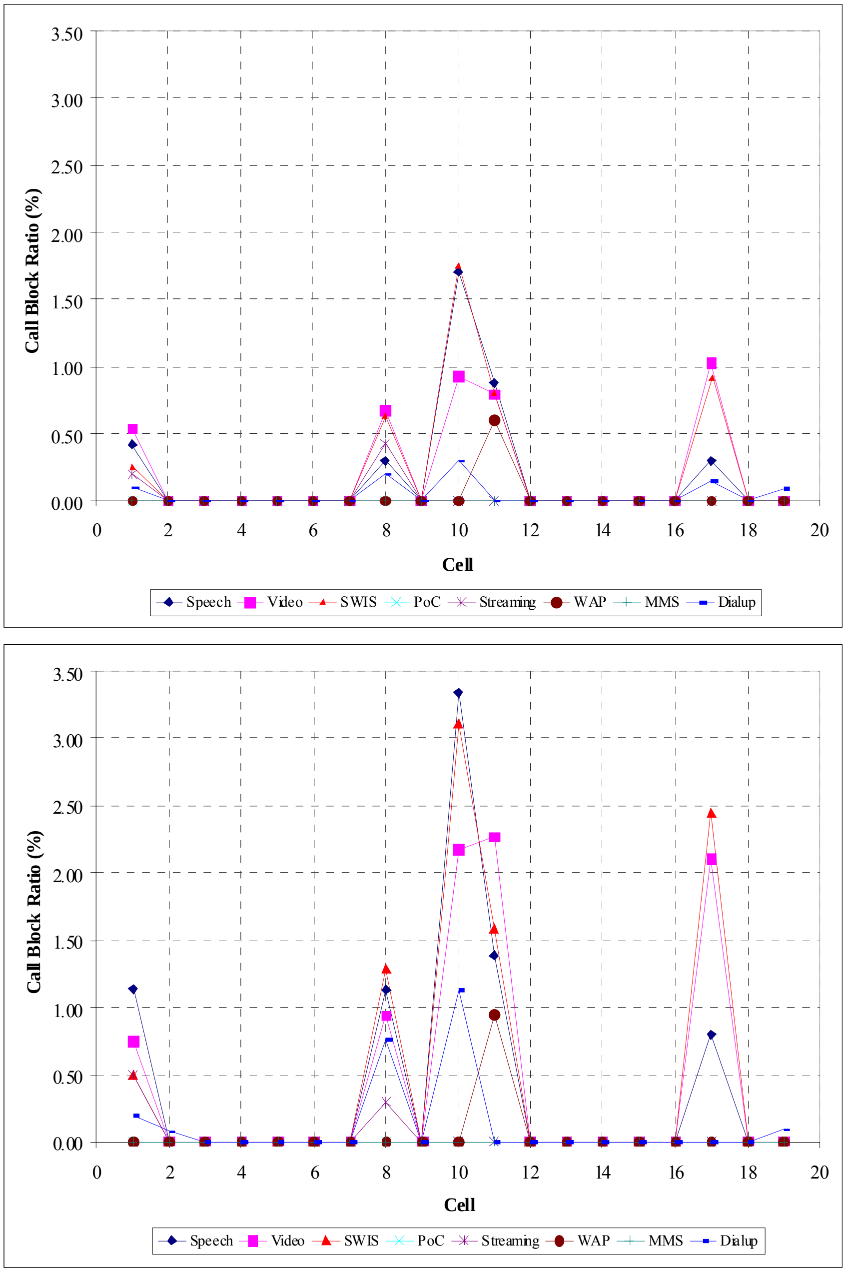
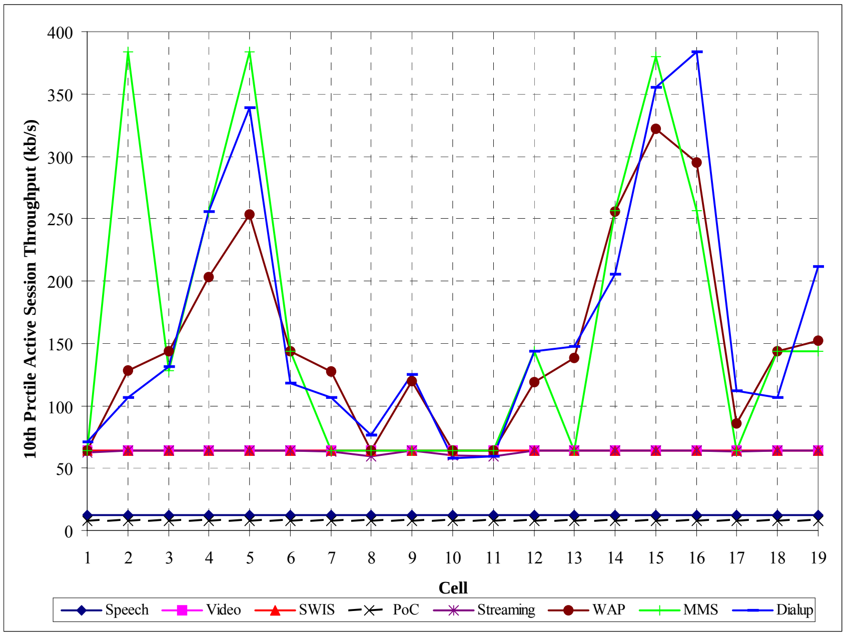


Figure 3.17. Call block ratio per service and cell with Ares (upper) and Wallu (lower).



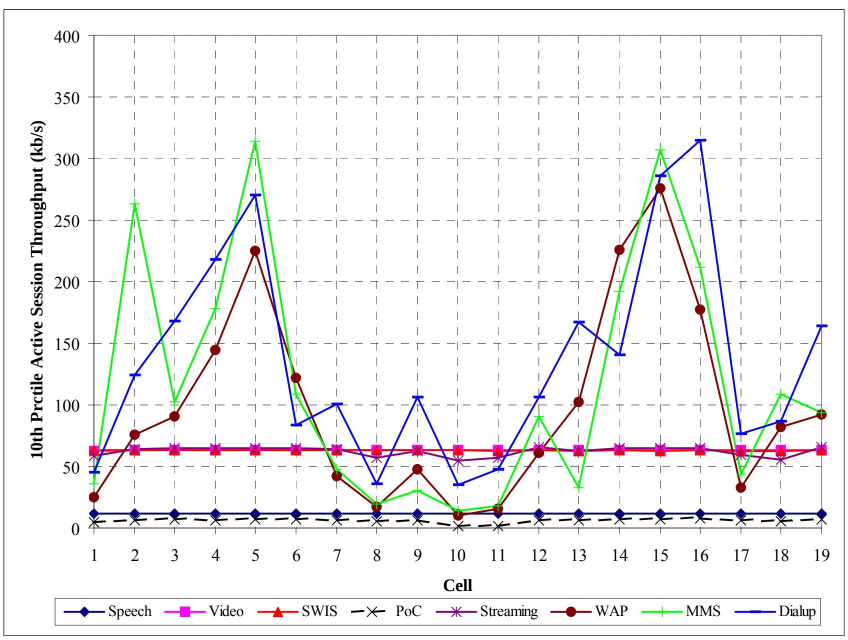


Figure 3.18. 10th percentile of active session throughput per service and cell with Ares (upper) and Wallu (lower).

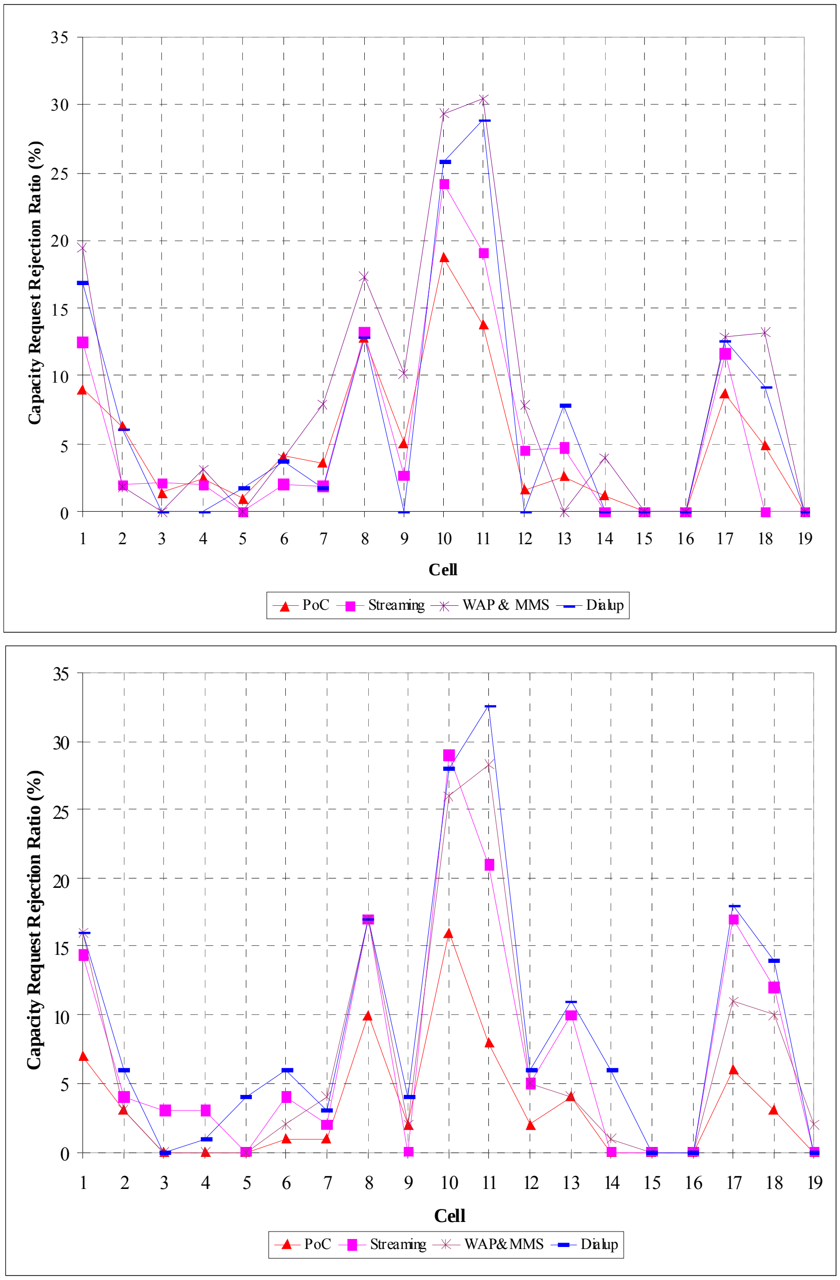


Figure 3.19. Capacity request rejection ratios per service and cell with Ares (upper) and Wallu (lower).

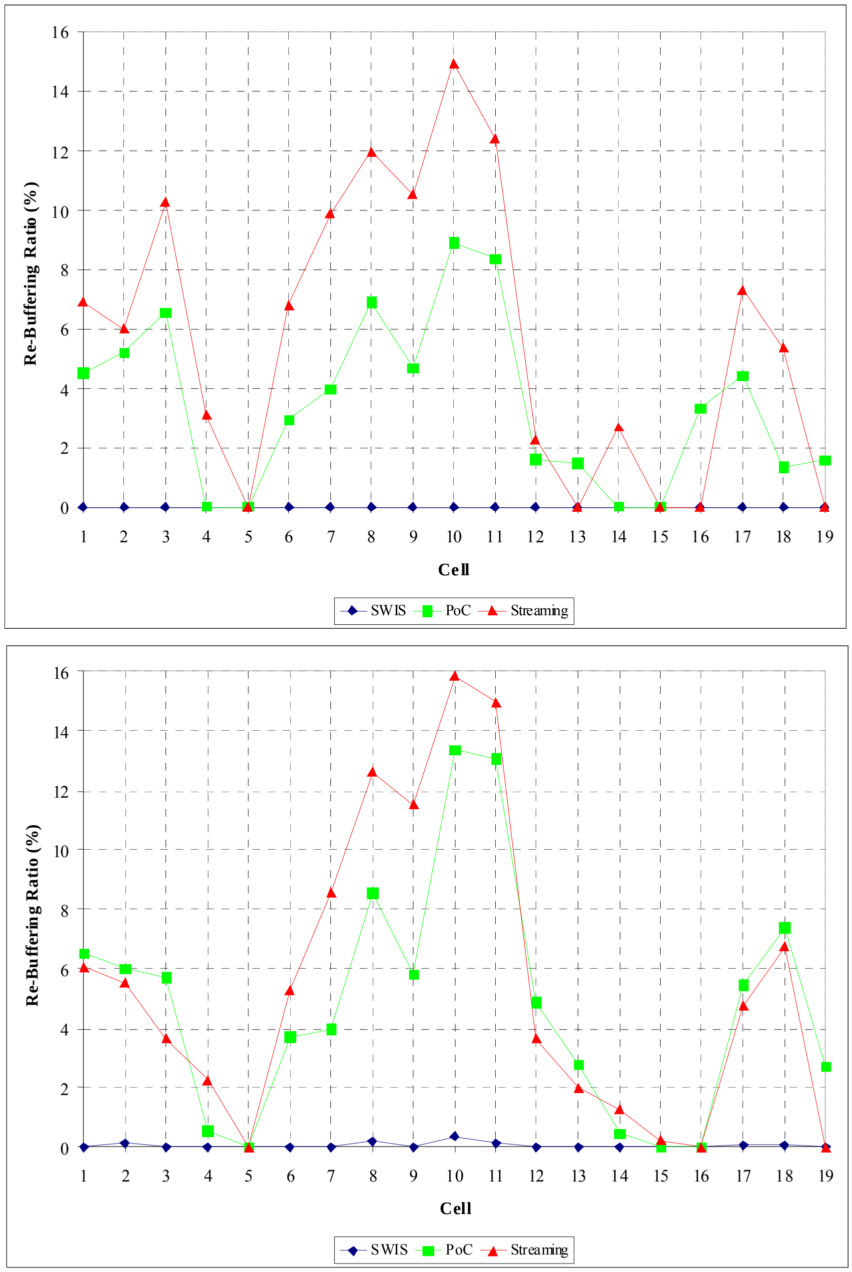


Figure 3.20. Re-buffering ratios per service and cell with Ares (upper) and Wallu (lower).

3.8 Conclusions

In this chapter, a comprehensive description of the assumptions, models and methodologies adopted in this work was given. As a part of this framework, an efficient and effective virtual time simulator, denoted by Ares in the thesis, for studying the provisioning of QoS in UTRAN was proposed, and a simplified version of tool, denoted as Max, for WCDMA radio interface dimensioning was described. The employed performance measures, environment, traffic and system models were also presented. Furthermore, in order to validate the precision and accuracy of the proposed tool, the performances of the virtual time prediction method were compared with the corresponding data attained using a trusted short-term dynamic model (simulator), denoted by Wallu in this thesis. The modelled network comprised 19 macro cells in Helsinki downtown area and a variety of real-time and non-real-time service users. Performances were monitored in terms of loads, throughputs, and satisfaction ratios. Many significant differences between the two simulator implementations were pointed out. For instance, the length of simulation step with Ares is hundreds times longer than that with Wallu. All short-time processes and their effects are thus ignored in Ares. Ares has no fast power control: Powers, compensating for the only slow fading, produce the target error rates exactly. There are no random signals fading processes in Ares, and terminals in this tool are not moving. The packet scheduler of Ares does not operate exactly the same way as that of Wallu. Due to the differences, it would be unfair to expect exactly overlapping performance results with the two prediction methods. Indeed, differences were observed in the magnitudes of powers, throughputs, and satisfaction ratios, especially, in high loaded cells. In general, Ares indicated better performance than Wallu. However, the trends of the curves and performance results were consistent, and satisfactorily to conclude that Ares does quite well the task it was designed for, as presented in the following chapters, where the motivation why this new tool was developed instead of using Wallu is further clarified.

4

QoS Provisioning

This chapter describes effective QoS management functions with service differentiation for WCDMA networks. The proposed radio resource management (RRM) algorithms, i.e. Admission Control, Packet Scheduler, Load Control, Power Control and Handover Control with QoS differentiation, enable network administrators to offer new services in a cost-effective way. As a part of this framework, in order to illustrate the benefits provided by the proposed differentiated service treatment in UTRAN respect to non-prioritised scenarios and/or undifferentiated parameter settings, six case studies are presented in the following sections. Several network configurations and mappings of RT services onto different QoS profiles are analysed by varying offered traffic loads in a realistic propagation environment. This includes the possibility of offering audio/video streaming with and without bit rate guarantees using two packet-scheduling algorithms, i.e. Fair Throughput (FT) and Fair Resources (FR) scheduling. Simulations were run using the prediction method presented in Section 3.5.1. Performance results are compared in terms of user satisfaction and, at a given QoE, capacity gains are expressed in terms of spectral efficiency, as introduced in Section 3.6. Discussions regarding the research methodology, parameter settings for a set of traffic mixes, impact of additional services on the offered ones and parameterisation thereof, as well as whether the results would hold in other deployment scenarios are within the scope of Chapter 7.

4.1 Introduction

An overview of QoS management functions needed to set up, modify and maintain bearer services with specific QoS according to 3GPP standards was given in Section 2.6. RRM strategies supported by UTRAN specifications and typical algorithms are described in [84]. However, as already pointed out in Section 1.1 and 2.6, features details and the relationship between the functions internal to the network elements are implementation specific, which the standardisation body leaves to the operators' and vendors' design.

On the matter, several studies are available in the literature. A review of the previous work (the most relevant radio resources allocation schemes) was reported Section 1.2.1.

In [54]-[56], the author presented some aspects of the performance of differentiated services offered to distinct users, and derived quantitative answers to accessibility gains provided by prioritisation of radio resource requests in UTRAN. The performance results were attained using the basic version of Ares, the author created and implemented on purpose, as described in [52]. In particular, in [54], the author presented an overview of RRM functions with user and bearer service differentiation, and analysed the impact of different parameter settings on prioritised connections in high traffic situations. Besides this, the author demonstrated the capability of the proposed algorithms to discriminate between users/services with different QoS profiles in terms of waiting time, CBR, CRRR, active session throughput and perceived data transfer delay. Further, the author showed that, while providing differentiation between NGB services, yet maintaining a high degree of fairness between peers (users of the same QoS class), is an easy task for the proposed resources allocation scheme; the differentiation between GB services turns out to be more difficult for the network administrator to handle. For that purpose, the author proposed some modifications to the AC queuing system. In [55], using the RRM functions described in [54], and in the following sections, the author studied the performance of Web browsing, e-mail, FTP and UDP data as perceived by the end users of WCDMA networks. The experienced quality was characterised by the object transfer delays and active session throughput. Simulation results showed that the flexibility of the proposed RRM functions allows for improved perceived quality and utilisation of radio resources in UTRAN. In the analysed traffic mix scenarios, the number of Interactive users accommodated into the system, at a given QoE, turned out to be 80% more compared to the non-prioritised case. Ultimately, in [55], the author introduced the concept of policy based QoS management and examined the feasibility of policies to configure QoS functions (such as admission control and packet scheduler) in UTRAN through two case studies. Simulation results showed that, at a given quality of experience, 40% more prioritised users of NGB services could be accommodated in the system by adapting policy parameters to the observed network and user behaviour. The investigated parameters were, in particular, the minimum allowed bit rate and dedicated power for NGB traffic, which are described in the following sections.

In the above papers, the author demonstrated the possibility for the operator to divide users into distinct categories (gold, silver, bronze) and accordingly enforce different charging policies. However, the benefit of this discrimination in terms of revenues and Background users satisfaction turns out to be noteworthy only at high traffic loads, where the advantages of such a classification can be really appreciated. Eventually, this model could be too complex from the operator's marketing point of view and its usability may be limited to corporate, Internet or machine users. Business wise, a more adequate solution appears to be the one that encompasses the prioritisation of service applications and supports guaranteed bit rates for the most delay sensitive ones. Such an example, which may provide the same level of end-user service quality improvements and the same gains in WCDMA network dimensioning, was explored in [57] and in the following sections.

This chapter describes thoroughly the proposed QoS management functions in UTRAN and enhancements thereof to improve the quality of prioritised bearer services. Besides this, to study the feasibility of the proposed resource allocation schemes, several network configurations, and mappings of real time services onto different QoS profiles are analysed. The analysis is performed by varying the offered load using the models and measures described in Chapter 3.

The reminder of this chapter is organised as follows: Section 4.2 describes the proposed RRM functions. Section 4.3 presents the specific models and simulation assumptions. Section 4.4 examines the performance of the proposed RRM algorithms and differentiated parameter settings. Section 4.5 concludes the chapter with a summary of the key issues. An evaluation of the statistical confidence on the simulation results can be found in Appendix C.

4.2 QoS management Functions in UTRAN

This section presents the simulated RRM algorithms, i.e. Admission Control (AC), Packet Scheduling (PS) and Load Control (LC) with QoS differentiation. Differentiated handoffs were not within the scope of this simulation campaign, since handover queues, enduring the same discrimination principle as in the AC ones, would lead to similar conclusions upon the added values provided by the differentiated treatment. Hence, only the AC queues were simulated, as explained in Section 4.2.1. The Power Control (PC) function supported by the simulator was presented in Chapter 3, where more insights into the Handover Control (HC) functionality were also given. Furthermore, since the system is downlink capacity limited in high traffic situations [11], only the downlink (DL) direction was simulated. Moreover, during the transport channel type selection, only dedicated channels (DCHs) were used for user data transmission.

4.2.1 Admission Control

For each cell, when a connection is set up or modified, AC assigns a *resource request priority* (*RRP*) value to the bearer service in question based on the QoS profile it receives from the core network (CN). The RRP value is a differentiated parameter (table) for the operator to set. The lower the RRP value the higher the priority assigned to the RAB in question. New radio link requests are arranged into priority queues and served following the strict priority principle (during handoffs, branch additions have top priority) and, at a given priority, based on their arrival times (FIFO). Resource requests (RR) are rejected if either the *maximum queuing time* or the corresponding *maximum allowed queue length* (in number of RABs) is exceeded. These parameters may be set differently depending on the bearer service characteristics, such as QoS class, ARP and THP, in the case of Interactive class. Except for the overload situation, defined by

$$P_{TxTotal} = P_{NGB} + P_{GB} > P_{TxTarget} + Offset\ , \tag{4.1}$$

where $P_{TxTotal}$ is the total current transmission power in the cell, exploited by the served traffic with guaranteed (GB) and non guaranteed bit rate (NGB), and $P_{TxTarget} + Offset$ is the overload threshold, NGB bearer services are always admitted, whereas GB traffic is not admitted if either (4.1) or the following inequality is satisfied:

$$P_{GB} + \Delta P_{GB} > P_{TxTarget} - P_{TxNGBcapacity}\ , \tag{4.2}$$

where $P_{TxNGBcapacity}$ is the power (capacity) dedicated to NGB traffic, and ΔP_{GB} is the estimated power increase in the cell if the bearer in question is admitted. ΔP_{GB} is calculated using the following equation, which is a modified version of the formula used to calculate the initial downlink radio link power in [11]:

$$\Delta P_{Tx} = \frac{\rho R}{W}\left(\frac{p_{tx,CPICH}}{\rho_c} + (1-\alpha)P_{TxTarget} - P_{TxTotal}\right), \tag{4.3}$$

where ρ and R are the required E_b/N_0 and maximum bit rate of the bearer in question, $p_{tx,CPICH}$ is the power of the common pilot channel in the cell, ρ_c is the energy per chip per noise spectral density (CPICH E_c/N_0) received by the mobile, W is the chip rate (3.84 Mchip/s), and α is the factor of DL code orthogonality ($\alpha = 1$ means perfect orthogonality). During soft HO, diversity branches are not set up if (3.1) is satisfied.

4.2.2 Packet (Bit Rate) Scheduler

Bit rates of admitted NGB bearer services are scheduled based on the actual radio resource priority values and, at a given priority, based on the arrival times of the corresponding capacity requests (CRs), following the FIFO principle. The bit rate allocation method is based on the *minimum* and *maximum allowed bit rates*, which define the lowest and highest limits of the TFS that can be allocated to the requesting bearer service. These parameters are differentiated depending on the bearer service characteristics.

In order to increase the degree of fairness (reduce the variance of the allocated bit rates) among peers, the radio resource priority value of a persisting CR in the PS queue may be modified taking into account the actual active session throughput of that particular connection. The new radio resource priority value of the bearer in question is calculated as follows:

$$RRP_i(t) = RRP_i - x\frac{R^i_{Target} - R_i(t)}{\max\{TFS_i\}}, \tag{4.4}$$

where RRP_i is the initial priority assigned by AC (see Section 4.2.1) to the bearer in question i, x is the *priority step* parameter, which can be set differently for different cells, R^i_{Target} is the target bit rate of that particular connection, $R_i(t)$ is the measured average active session throughput at the time t during the packet data transmission, and $\max\{TFS_i\}$ is the maximum bit rate of the TFS of the dedicated channel used to transport the corresponding user data. (Note: In this work, the effects of the priority step settings are not investigated, since, at this stage, the fairness between peers does not appear to be such an issue to justify the many degree of freedom introduced by this parameter.)

During the bit rate allocation, PS follows the best effort model and relies upon the power budget left by the GB and NGB active (*a*) and inactive (*i*) connections, i.e.

$$P^{NGB}_{Allowed} = P_{Target} - (P_a + P_{i,GB} + kP_{i,DTX}) \, . \tag{4.5}$$

The power of inactive GB traffic $P_{i,GB}$ takes into account the required power of the bearer services just admitted, but not yet on air, whereas $P_{i,DTX}$ is the power that needs to be reserved for the bearer services in discontinuous transmission (DTX) during their idle or reading periods. Since it is unlikely that all inactive connections get active at the same time, we introduce k as a management parameter to restrict the reserved power to a lower value. k ranges from 0 to 1, where 1 is the most conservative number, which yields the lowest effectiveness, since in this case all required power would be reserved. In the case of SHO, the allocated bit rate is the minimum of the bit rates scheduled for each of the links of the radio link set.

Capacity requests are rejected if they stay longer in the PS queue than the value specified by *CR maximum queuing time*. Allocated bit rates may be rescheduled if the ongoing communication has lasted longer than the *corresponding granted DCH minimum allocation time* (see Figure 4.1, where an example of bit rate modification is depicted). Both parameters may be set differently for distinct bearer service characteristics.

During the transport channel type selection, only dedicated channels (DCHs) are available for high rate packet data transmission. Bearer services that have been longer in DTX than the corresponding *inactivity timer* are moved to Cell_FACH state (see Section 2.3.2.6) and the corresponding allocated resources are released. However, there is no transmission using FACH: When the new data arrives in the buffer of the radio network controller (RNC), a new CR is sent to PS and subsequently another DCH is allocated to the bearer in question based on the selected scheduling algorithm, as explained in the following sections. The inactivity timer may be a differentiated parameter.

4.2.2.1 Fair Throughput (FT) Bit Rate Scheduling

In the Fair Throughput (FT) scheduling algorithm, the unexploited power in (4.5) is shared fairly in terms of throughput among the different bearer services every scheduling period. This is achieved by matching the bit rates and thus the estimated transmit powers of the bearer services to sum up to (4.5). Power estimates in the bit rate increase algorithm are based on (4.3).

The algorithm may be implemented as follows. Consider a matrix of N_R x N_C entries, where the number of rows N_R corresponds to the max TFS size of the employed DCH, and the number of columns N_C is the amount of CRs in queue to be scheduled. Rows (bit rates) are sorted in ascending order and columns (CRs) are arranged in descending order based on priorities and arrival times (FIFO). The TFs are then scanned (from left to right) using a sliding window (pointer), whose size equals the number of CRs, starting from the highest bit rates of the most demanding bearer services down to the lowest ones across the different TFSs. The corresponding exploited capacity is estimated at each step. The selected (allocated) bit rates are the ones that fit into the available capacity. In other words, the window stops sliding when the estimated power increase, for the selected bit rates, is less or equal to the available power budget for NGB traffic. In the example of Figure 4.2, the assigned bit rates to the five CRs are marked in red. Using this scheme bit rates are fairly allocated accordingly to the priority values set by the operator. This bit rate allocation scheme may run in MAC layer, thus speeding up the process of handling bit rates without involving any RRC procedure at layer 3 [69].

An example of bit rate allocation for up to five capacity requests in PS queue using the FR algorithm is illustrated in Figure 4.3. The chronological order of the capacity requests (CRs) is written in brackets and the priorities of the bearers in question represented by different colours. In the figure, the possible bit rates in the TFS are: 128, 144, 256 and 384 kb/s, where 128 kb/s and 384 kb/s are the dedicated channel (DCH) *minimum* and *maximum allowed bit rate*, respectively. When five capacity requests are in the queue, only four of those are allocated 128 kb/s.

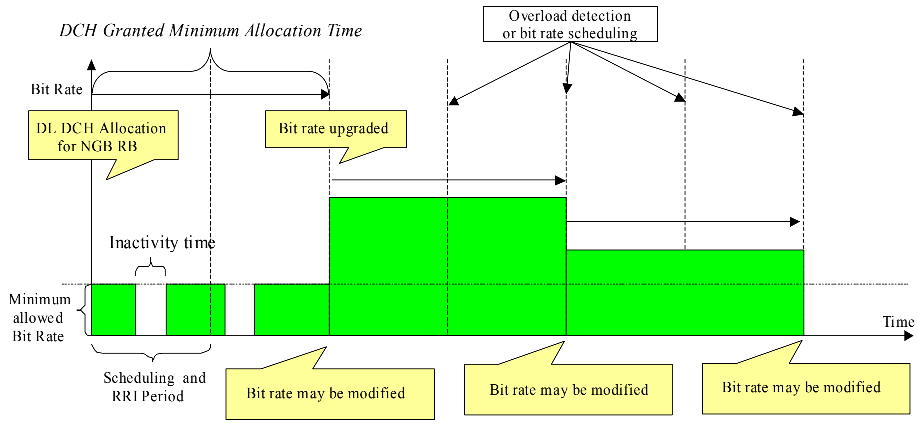


Figure 4.1. Example of bit rate modification by packet scheduler. The time resolution is one *scheduling period*, which in this case corresponds to one *radio resource indication* (RRI) period (interval between consecutive power measurements received by the radio network controller from the base station).

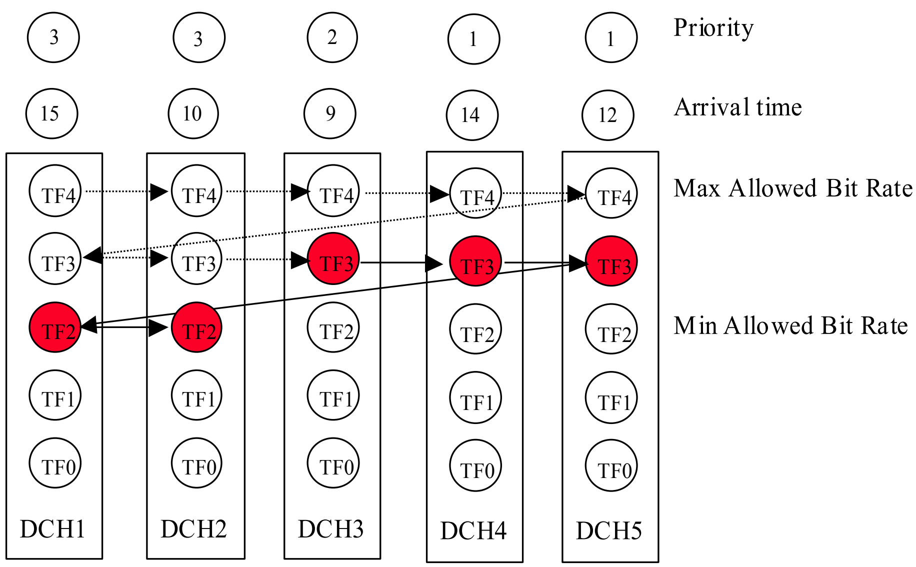


Figure 4.2. Example of bit rate allocation for 5 capacity requests in packet scheduler queue.

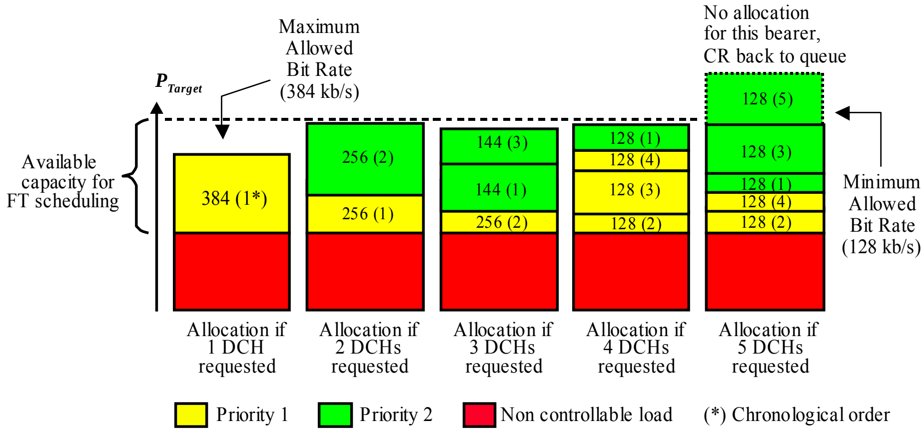


Figure 4.3. Example of bit rate allocation for 1-5 capacity requests in packet scheduler queue when FT scheduling is used.

4.2.2.2 Fair Resources (FR) Bit Rate Scheduling

In case of Fair Resource (FR) scheduling, all users get the same output power and the actual bit rates depend on the received SIR (signal to interference ratio). This means that users close to the base station will be allocated higher bit rates than the users at cell borders. In the case of QoS differentiation, different factors (*weights*) could be used for the output power, depending on the used service (TC) and subscription profile. The weights can have values between 0 and 1. (Note: In this work, the effects of the weights in the FR algorithms are not investigated, since distinct priority settings appear to be sufficient to provide discrimination between different peers.) The proposed algorithm is an enhancement of the FT scheduling presented in the previous section, and consists of the following steps:

1. $P_{NGBAllowed}$ is calculated as shown in (4.5).
2. $P_{NGBAllowed}$ is evenly divided (in linear scale) among all radio links prior to scheduling (or rescheduling), i.e. $p_{tx,link} = P_{NGBAllowed}/N$, where N is the number of calls under scheduling.
3. $p_{tx,link}$ is used in the below formula to calculate the "virtual" bit rate (R_i) of that particular session (i):

$$R_i = \frac{W}{\rho} \frac{\dfrac{w_i}{\sum_j w_j} p_{tx,link}}{P_{tx,target}(1-\alpha) + \dfrac{p_{tx,cpich}}{\rho_c} - P_{tx,tot}}, \tag{4.6}$$

where w_i are the differentiated power weights introduced above, and all other terms are as defined in (4.3), from which (4.6) was derived. This will now result in different bit rates depending on the corresponding path loss and interference situation.

4. For each session, the virtual bit rate is rounded upwards to the next bit rate in TFS. For those cases where R_i exceeds the maximum bit rate in TFS, it is rounded down to the maximum bit rate in TFS (see Figure 4.4).
5. A temporary TFS for each session is constructed. In this TFS the highest bit rate is the one found in step 4 for each session separately and the lowest ones are replaced by zero.
6. FT scheduling is performed as presented in Section 4.2.2.1, yet using the temporary TFS generated in the previous step. The sessions with zero bit rates are not scheduled.

Figure 4.5 shows an example of bit rate allocation using the algorithm presented above for the five cases presented in Section 4.3. The capacity requests are the ones illustrated in Figure 4.3 for FR scheduling. In this case, the possible bit rates in the TFS are: 64, 128, 144, 256 and 384 kb/s, where 64 kb/s and 384 kb/s are the dedicated channel (DCH) *minimum* and *maximum allowed bit rate*, respectively. Although the minimum allowed bit rate for the DCH has been reduced to 64 kb/s, when five capacity requests are in the queue, only four of those get a possibility to transmit.

In order to compensate for the slow fading fluctuations, the propagation path losses and interference conditions, E_c/N_0 and $P_{tx,tot}$ values in (4.6), should be continuously monitored for each mobile terminal in the cell and the allocated bit rate rescheduled whenever needed, regardless the value of the *granted DCH minimum allocation time* parameter.

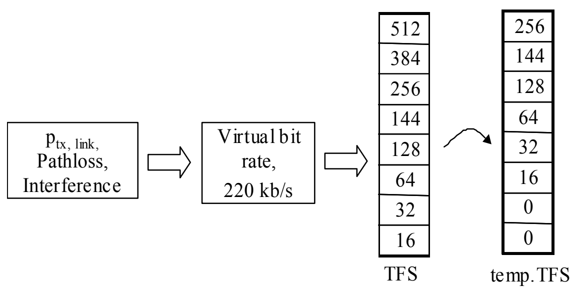


Figure 4.4. Example of transport format set (TFS) construction when the FR scheduling algorithm is used.

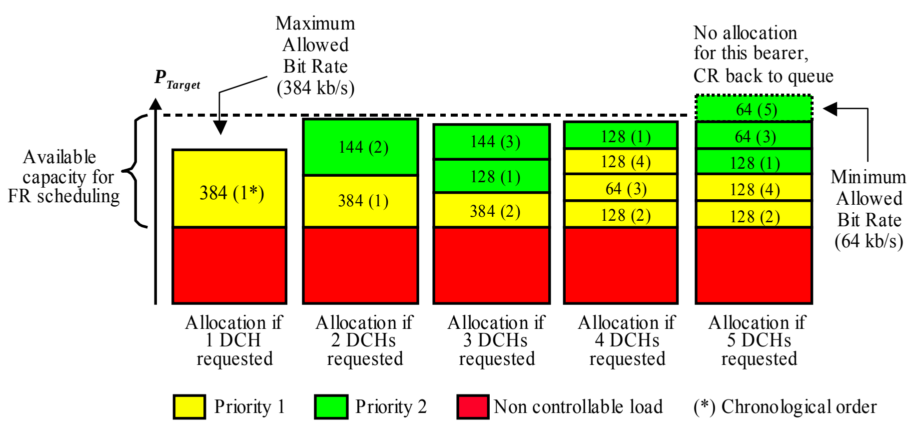


Figure 4.5. Example of bit rate allocation for 1-5 capacity requests in packet scheduler queue when FR scheduling is used.

4.2.3 Load Control

The only load control action considered in this work is the reduction of bit rates of NGB bearer services when (4.1) is satisfied. The DCH bit rate may be downgraded only when the allocation time of the carried service lasted longer than the corresponding *DCH granted* (*overload) minimum allocation time* (see Figure 4.1). Bit rates are reduced starting from bearers with the lowest priority, and at given priority, based on their arrival times (FIFO), but none of the sessions is released. Power estimates in the bit rate decrease algorithm are based on (4.3).

4.3 Specific Models and Simulation Assumptions

Several case studies were simulated to investigate the feasibility of the proposed algorithms and to calculate the spectral efficiency gains provided by the service differentiation. Six of those, focusing on the deployment of the services reported in Table 4.1, using the fair throughput and fair resources scheduling algorithms, are presented in this and the following section.

The provisioning of QoS is provided by service differentiation in admission control, load control, scheduling of bit rates and using service dependent values for the minimum allowed bit rates, maximum queuing times, inactivity timers, and granted DCH allocation times in normal situation and in the case of overload. Besides this, two mappings of services onto different QoS profiles are analysed in this work. In the first mapping, denoted by Mapping1 (M1), PoC and streaming are offered on best effort, whereas in the second configuration, denoted by Mapping 2 (M2), streaming is carried with guaranteed bit rate.

Simulations were performed over a period of 2 hours using a time step of 200 ms (*radio resource indication* period (RRI), which is the time needed to receive in the RNC power measurements from the base stations), and all statistics were collected over the entire system and simulation period. The impact of the transient phase (warm up time) is studied in Appendix C.

Simulated traffic mix (in share of calls, see Table 4.1) and traffic intensity were varied according to the case studies presented in the next sections. Users were randomly uniformly distributed over the area depicted in Figure 3.2. Table 4.2 summarises the most important network parameters. More information on the adopted traffic, network and environment models can be found in Sections 3.2-3.3.

Table 4.1. Adopted traffic mixes.

Service	PoC	Streaming (Audio and video)	MMS	Dialup	SWIS	WAP	CS Speech
Share of calls (%)	14 (20[a])	2 (20[a])	8 (20[a])	3 (20[a])	3 (0[a])	38 (20[a])	32 (0[a])

[a]) Used only in the Fair Throughput (FT) – Fair Resources (FR) bit rate scheduling comparison (Case 2).

Table 4.2. Most important system based parameters.

Parameter	Value
Call or session mean arrival rate [a]	3, 2, 1, 1.5, 1, 0.5, 0.25 s
Radio resource indication period (*RRI*)	0.2 s
Simulation time (s)	7200 s
Power target for DL AC, see (4.1) and (4.2)	3 dB below BTS total power
Overload offset for DL AC, see (4.1)	1 dB above power target
Orthogonality (α), see (4.3)	0.5
Period for packet scheduling and load control actions	0.2 s (1 RRI)
Length of AC queue	10 RABs
Dedicated NGB capacity [b], see (4.2)	0 W
Priority step (x) [c], see (4.4)	0
Power weights for inactive NGB traffic (k), see (4.5)	0.5
Weights for FR scheduling algorithm (w_i) [c], see (4.6)	1
Transport Format Set (TFS) for NGB services	0, 16, 32, 64, 128, 144, 256, 384 kb/s

a. On average a user makes a connection every mean arrival rate seconds. Values depend on the case study.

b. The author investigated the influence on QoS and setting of this parameter in [56].

c. The impact of this parameter setting was not investigated in this work.

4.3.1 Case 1 – QoS Differentiation in UTRAN with FT Scheduling

This section introduces four case studies to draw conclusion about the gains provided by the prioritisation, and to quantify the influence of differentiated parameters and mappings on the performance of the offered services. The adopted traffic mix was illustrated in Table 4.1 and the scheduling algorithm (FT) was described in Section 4.2.2.1.

4.3.1.1 Case 1.1: Non-Prioritised case, cell based parameters, service Mapping 1 (NPM1)

In this case, all resource requests have the same priority and none of the relevant parameter values are differentiated, as shown in Table 4.3. The DCH granted minimum allocation time was 5s shorter for overload control actions. Target bit rates for the user satisfaction criteria are set according to the service throughputs defined in Section 3.6. The minimum allowed bit rates correspond to applications source bit rates or upper rounded throughputs taken from the user satisfaction criteria. The inactivity timer is short to release unutilised radio resources as fast as possible. The CR maximum queuing time in the PS queues is long enough (10s) to reduce the probability of being rejected while scheduling the bit rates. Buffering delays reflect the users expectations, i.e. longer delays would not be tolerated. The AC maximum queuing time is stretched up to the limit a user bears in practice, after that a subscriber usually hangs up. The maximum allowed bit rate corresponds to the maximum bit rate supported in UTRAN for user data transport on dedicated channels. The services were mapped onto traffic classes according to Mapping 1, i.e. Speech was served as CS-Conversational and SWIS was carried on PS-Streaming class. These were the only services offered with guaranteed bit rate; all other applications were treated on the best effort (BE or non-guaranteed bit rate). Furthermore, PoC, video and audio streaming, and

WAP/MMS were mapped onto PS-Interactive. WAP/MMS were supposed to be placed by the operator behind the same Access Point Name (APN, see [73]). Dialup connections, which comprised, for example, ftp, emails and http traffic, were carried on PS-Background class.

4.3.1.2 Case 1.2: Prioritised case with cell based parameters, service Mapping 1 (PM1)

In this case, services are mapped as in Case 1.1, and using the same parameter values, but now they are offered with different priorities. Speech calls have top priority, followed by SWIS, PoC, and all other service applications. Within the Interactive class, using different traffic handling priorities (THP), PoC is handled first, followed by audio and video streaming, and WAP/MMS. Dialup, on Background class, is the service with the lowest priority, and hence served last, as illustrated in Table 4.3.

4.3.1.3 Case 1.3: Prioritised case with Differentiated parameters, service Mapping 1 (PDM1)

Also in this case, services are mapped as in Case 1.1, but now they are offered with different priorities and parameter values. To further improve PoC and streaming performance, the minimum allowed bit rate for WAP/MMS and Dialup connections was reduced, and the granted minimum allocation time for DCHs carrying PoC and streaming increased, whereas for Dialup was reduced, as illustrated in Table 4.3.

4.3.1.4 Case 1.4: Prioritised case with Differentiated parameters, service Mapping 2 (PDM2)

This case was introduced to study the benefit for the operator to offer audio and video streaming on Streaming QoS class with guaranteed bit rate. Besides this, we wanted to analyse the performance deterioration at high traffic volumes, in terms of dissatisfied users, when SWIS and Streaming services start to compete for radio resources. In this context, SWIS, being mapped on PS-Conversational class, should experience less blocking than Streaming due to the higher priority in the AC queue. The mapping for this case together with the other differentiated parameters, which are the same as in Case 1.3, is displayed in Table 4.3.

4.3.2 Case 2 – Comparison Between FR and FT Scheduling

The comparison between FR and FT bit rate scheduling was done using the parameter values of Case 1.2 (see Table 4.3). In other words, FR and FT scheduling were compared using the prioritised case with cell-based parameters. Two traffic mixes were tested, i.e. the one adopted in Case 1, and 20% for all NGB services (see Table 4.1).

Table 4.3. Mapping of services onto QoS classes and parameter values for the 4 case studies: Non Prioritised Mapping 1 (NPM1); Prioritised Mapping 1 (PM1); Prioritised with Differentiated parameters Mapping 1 (PDM1); Prioritised with Differentiated parameters Mapping 2 (PDM2).

Case	Application service	RRP	DCH Granted Overload Min Alloc Time (s)	Target Bit Rate (kb/s)	DCH Granted Min Alloc Time (s)	Minimum Allowed Bit Rate (kb/s)	Inactivity Timer (s)	CR Max Queuing Time (s)	Buffering Delay (s)	AC Maximum Queuing Time (s)	Maximum Allowed Bit Rate (kb/s)	Mapping onto Traffic Class
1.1 (NPM1)	Speech	1	-	12.2	-	GB	-	-	-	5	GB	CS Conversational
	Streaming	1	10	64	15	64	5	10	16	15	64	Interactive THP2
	SWIS	1	-	64	-	GB	-	-	4	10	GB	PS Streaming
	PoC	1	10	8	15	8	5	10	4	15	8	Interactive THP1
	WAP	1	10	32	15	64	5	10	-	15	384	Interactive THP3
	Dialup	1	10	64	15	64	5	10	-	15	384	Background
	MMS	1	10	8	15	64	5	10	-	15	384	Interactive THP3
1.2 (PM1)	Speech	1	-	12.2	-	**GB**	-	-	-	5	**GB**	CS Conversational
	Streaming	4	10	64	15	64	5	10	16	15	64	Interactive THP2
	SWIS	2	-	64	-	**GB**	-	-	4	10	**GB**	PS Streaming
	PoC	3	10	8	15	8	5	10	4	15	8	Interactive THP1
	WAP	5	10	32	15	64	5	10	-	15	384	Interactive THP3
	Dialup	6	10	64	15	64	5	10	-	15	384	Background
	MMS	5	10	8	15	64	5	10	-	15	384	Interactive THP3
1.3 (PDM1)	Speech	1	-	12.2	-	**GB**	-	-	-	5	**GB**	CS Conversational
	Streaming	4	5	64	10	64	5	10	16	15	64	Interactive THP2
	SWIS	2	-	64	-	**GB**	-	-	4	10	**GB**	PS Streaming
	PoC	3	10	8	15	8	60	4	4	15	8	Interactive THP1
	WAP	5	0.2	32	5	32	10	10	-	15	384	Interactive THP3
	Dialup	6	0.2	64	1	16	5	5	-	15	384	Background
	MMS	5	0.2	8	5	32	10	10	-	15	384	Interactive THP3
1.4 (PDM2)	Speech	1	-	12.2	-	**GB**	-	-	-	5	**GB**	CS Conversational
	Streaming	**3**	-	64	-	**GB**	-	-	16	15	**GB**	PS Streaming
	SWIS	**2**	-	64	-	**GB**	-	-	4	10	**GB**	PS Conversational
	PoC	4	10	8	15	8	60	4	4	15	8	Interactive THP1
	WAP	5	0.2	32	5	32	10	10	-	15	384	Interactive THP2
	Dialup	6	0.2	64	1	16	5	5	-	15	384	Background
	MMS	5	0.2	8	5	32	10	10	-	15	384	Interactive THP2

4.4 Simulation Results and Discussion

This section presents and discusses the performance results of the case studies described in the previous sections. Conclusions are drawn in Section 4.5.

4.4.1 Case 1 – QoS Gains with FT Scheduling

Performance results in terms of percentage of satisfied users and cell throughput normalised with respect to the chip rate are shown in Figure 4.6-Figure 4.9. Table 4.5 summarises the relative capacity gains due to different network configurations (QoS provisioning) when at least 90% of users of the worst performing service turn out to be satisfied using the quality criteria defined in Section 3.6. None of the speech and SWIS calls were blocked or dropped; hence all speech and SWIS users were satisfied in all analysed traffic scenarios (see Table 4.4, where the offered traffic is expressed in number of calls or sessions in the system and per cell during the simulated time as a function of the input arrival intensity).

The simulation results of Case 1.1 (non-prioritised, Mapping 1, NPM1), and Case 1.2 (prioritised scenario with cell based parameters, Mapping 1, PM1) are depicted in Figure 4.6, where the percentage of satisfied users, defined in Section 3.6, is shown separately for each service type. The percentages of unsatisfied users of PoC and streaming services increase more rapidly than the ones for the users of other applications when the offered traffic increases. This is due to the fact that neither PoC nor streaming has a guaranteed bit rate when carried on Interactive traffic class. Yet, the corresponding performance in Case 1.2 turns out to be less sensitive to changes in traffic volumes. In other words, deploying different priorities (according to the sensitivity of a particular application to delays) makes it possible to achieve a spectral efficiency gain of 30% with respect to the non-differentiated case. The gain comes at the expense of the Dialup connections, whose deterioration, however, remained acceptable in all analysed traffic scenarios.

Performance results of Case 1.1 (non-prioritised with cell based parameters, NPM1), and Case 1.3 (prioritised with differentiated parameters, Mapping 1, PDM1), are compared in Figure 4.7. From the figure a spectral efficiency gain of 50% can be noticed (derived). This means that 20% more gain would be provided by the deployment of differentiated parameters. Such a gain is still at the expense of Dialup connections, whose bit rate, however, resulted high enough for a tolerable quality in all analysed traffic scenarios.

Table 4.4. Analysed Traffic Scenarios.

Mean Arival Rate (s)	**1.5**	**1**	**0.75**	**0.5**	**0.25**
N. of Calls/Sessions (system)	4808	7212	9614	14340	28826
N. of Calls/Sessions per cell	**253**	**380**	**506**	**755**	**1517**

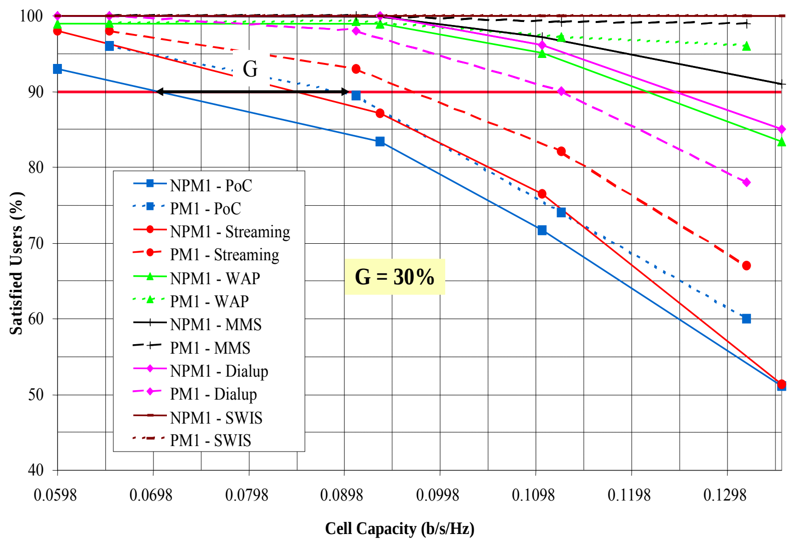


Figure 4.6. Case 1.1 vs. 1.2. Non-prioritised (NPM1) vs. prioritised scenario with cell-based parameters (PM1), using Mapping 1. The definition of spectral efficiency is given in Section 3.6.

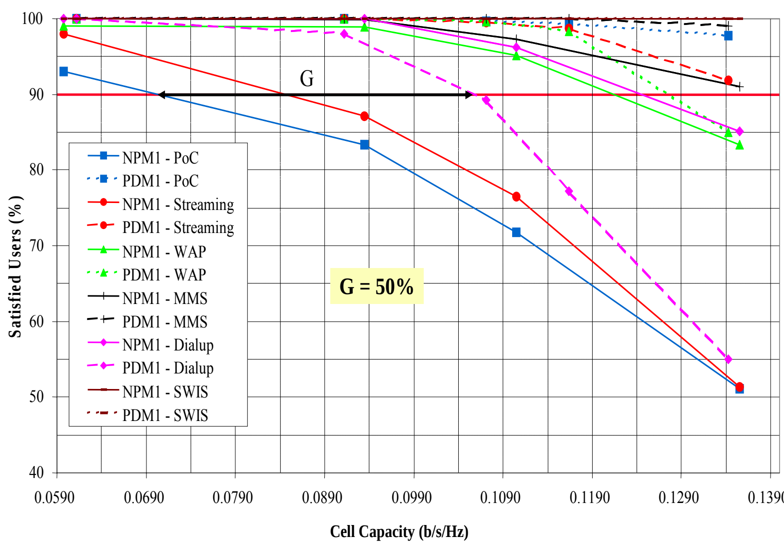


Figure 4.7. Case 1.1 vs. 1.3. Non-prioritised scenario (NPM1) vs. prioritised case with differentiated parameters (PDM1), using Mapping 1. The definition of spectral efficiency is given in Section 3.6.

Case 1.2 (prioritised, Mapping 1, PM1) was compared to Case 1.3 (prioritised with differentiated parameters, Mapping 1, PDM1). Performance results are shown in Figure 4.8. The spectral efficiency gain in this case is about 16%, if we consider Dialup the most demanding service in Case 1.3. However, the most important results stand on the benefit the RT communications would get from the differentiated parameters, which are essential to offer PoC and streaming on BE. As illustrated in Figure 4.8, at highest traffic volume (about 1517 calls-sessions per cell, see Table 4.4), the percentage of satisfied users of PoC and streaming services is still above the threshold (90%), whereas more than 30% of such users would be dissatisfied if the only priorities were set differently.

Ultimately, performance results of Case 1.3, and Case 1.4 (prioritised with differentiated parameters, Mapping 2, PDM2), are compared in Figure 4.9. In Case 1.4, streaming, as speech and SWIS, but with a lower priority, is offered with guaranteed bit rate. Dialup and WAP performances remain almost unaltered: The percentage of satisfied users of Dialup service (offered on Background Class), which determines the capacity gain between the two cases, is about the same, whereas 5% would be the spectral efficiency gain if WAP was considered the reference application (most demanding service). This is due to the fact that the available power in (4.5) is better estimated, and thus better exploited, when fewer NGB connections request capacity. In other words, the more NGB traffic is in DTX, the less accurate the estimate of the inactive NGB power in (4.5) is. From (4.2), we can see that streaming is admitted and served regardless the transmission power level of the best effort connections, since none of the base station power was reserved for NGB traffic ($P_{TxNGBcapacity}$ = 0, see Table 4.2). Nevertheless, Figure 4.9 reveals that setting different priorities and provisioning differentiated parameters is sufficient to sustain the quality of real time services offered on best effort at any of the investigated traffic volumes. In fact, PoC performance remains the same in the two cases and the percentage of streaming satisfied users in Case 1.3 (PDM1) never goes beyond the quality threshold.

As a summary, Table 4.5 collects the relative capacity gains between the explored cases, when at least 90% of users of the worst performing service were satisfied. Based on these values, from the cost-effectiveness perspective, we conclude that the best performing network configuration is the one with prioritised services and differentiated parameters. The proposed RRM algorithms appear to have enough flexibility for the operator to offer PoC and Streaming services on Interactive class, i.e. without any bit rate guarantee. This makes it possible to reserve resources during the sessions only when needed, and charge the packet communication simply when there is data to transmit.

Table 4.5. Performance results at a given user satisfaction.

Case	Spectral Efficiency Gain (%)
1.2 – PM1	30
1.1 – NPM1	
1.3 – PDM1	50
1.1 – NPM1	
1.3 – PDM1	16
1.2 – PM1	
1.4 – PDM2	0
1.3 – PDM1	

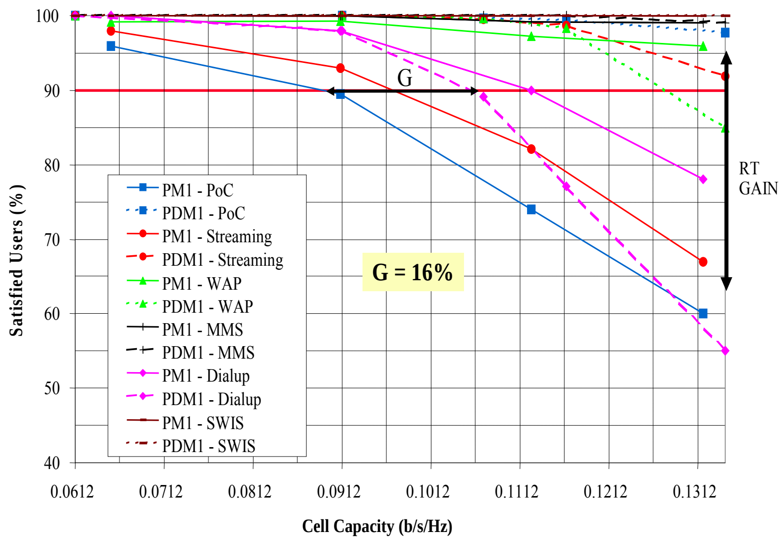


Figure 4.8. Case 1.2 vs. 1.3. Prioritised scenario without (PM1) and with differentiated parameters (PDM1), using Mapping 1. The definition of spectral efficiency is given in Section 3.6.

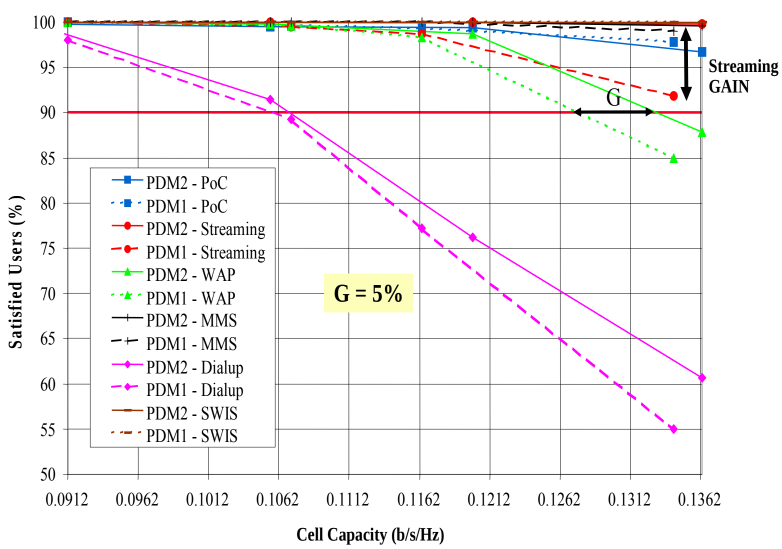


Figure 4.9. Case 1.3 vs. 1.4. Prioritised with differentiated parameters, Mapping 1 (PDM1) and Mapping 2 (PDM2). The definition of spectral efficiency is given in Section 3.6.

4.4.2 *Case 2 – Comparison Between FT and FR Scheduling*

Simulations results of this case study are shown in Figure 4.10 and Figure 4.11. The performance in terms of spectral efficiency using the traffic mixes of Table 4.1 is depicted in Figure 4.10 and Figure 4.11. From the figures we conclude that there would not be any benefit in terms of spectral efficiency if the FR algorithm was implemented in UTRAN to carry RT services as streaming and PoC on best effort. This is due to the fact that PoC and streaming are carried with a limited TFS, i.e. zero and the source bit rate, which is 8 and 64 kb/s, respectively. Hence, the link adaptation algorithm presented in Section 4.2.2.2, benefits the cell capacity only at high traffic volumes when already too many RT users are dissatisfied. Conversely, the spectral efficiency gain of the FR scheduling with respect to the FT one, at a given QoE, would be considerable if the NGB services were offered with a wide range of bit rates, as WAP, Dialup and MMS. In our case, the *additional* spectral efficiency gain provided by the FR algorithm would range from 10 to 30% depending on the input traffic mix, if WAP was considered the reference service (see Figure 4.10 and Figure 4.11).

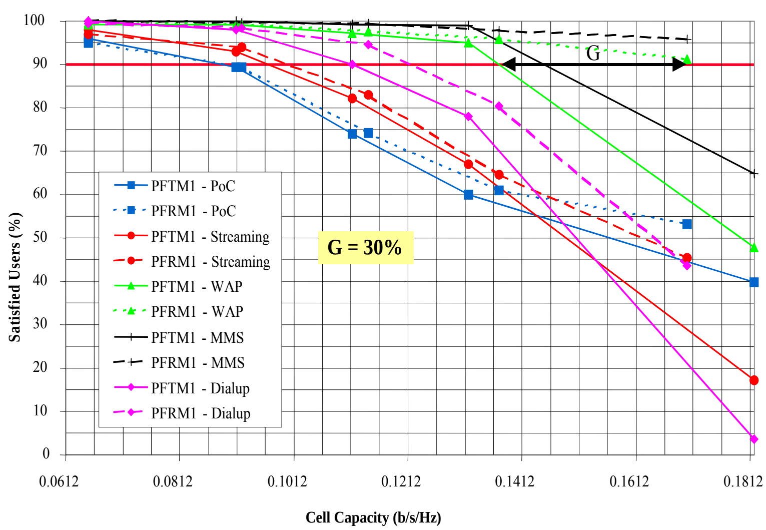


Figure 4.10. Case 2. Prioritised scenario with Fair Resources scheduling (PFRM1) vs. prioritised scenario with Fair Throughput scheduling (PFTM1). The traffic mix is the same as in Case 1. The definition of spectral efficiency is given in Section 3.6.

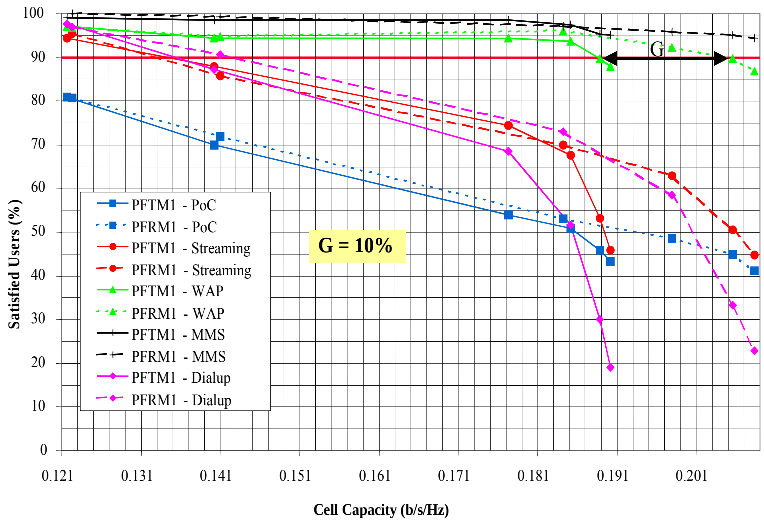


Figure 4.11. Case 2. Prioritised scenario with Fair Resources scheduling (PFRM1) vs. prioritised scenario with Fair Throughput scheduling (PFTM1). The simulated traffic mix is shown in Table 4.1 note (a). The definition of spectral efficiency is given in Section 3.6.

4.5 Conclusions

An effective framework to handle differentiated services in UTRAN has been presented. This encompasses an admission control, packet scheduling and load control functions with QoS differentiation. The feasibility and added values of the proposed QoS management functions were analysed by means of dynamic simulations using realistic traffic models for packet communications and traffic mix scenarios. Performance results showed enough flexibility for the operator to handle distinct services in a cost-effective manner and to sustain the required QoS of real time services, e.g. PoC, audio and video streaming, without any admission control or a priori physical radio channels reservation. Customised user satisfaction criteria were defined for the offered services, and the spectral efficiency was computed as the system load (mean cell throughput normalised with respect to the chip rate, 3.84 Mchip/s) at which 90% of users of the worst performing service were satisfied. Within this operation constraint, the spectral efficiency gain provided by the only prioritisation was 30% and the additional capacity gain when the non-guaranteed bit rate services were also provisioned with differentiated parameters was about 20%. Ultimately, the extra spectral efficiency gain provided by the proposed Fair Resources (FR) scheduling algorithm, with respect to Fair Throughput (FT) scheduling, depends heavily on the offered bearer services and traffic mix characteristics: The more traffic is offered with a wide range of bit rates, the more spectral efficient the FR scheduling is. Simulation results did not take into account any limitation entailed by I_{ub} (interface between the base station and radio network controller) bottlenecks, hardware capabilities, and/or availability of orthogonal codes. Yet, from the average cell throughput and the probability of a user (connection) to be inactive, we conclude that the code tree limitation was not an issue in this study.

5

Service Driven Radio Network Planning

This chapter addresses some of the radio dimensioning and detailed planning issues that may arise from the introduction of new services in WCDMA networks. To investigate the impact and entailed limitations of the deployment of new services on the performance of the offered ones, five case studies are presented in this chapter using the simulators and methods described in Chapter 3. Firstly, using Ares, the effectiveness of the supported QoS mechanisms for service differentiation is investigated in terms of cell throughput and required transmission power. The QoE is assessed separately for each of the simulated services through a customised combination of performance metrics, which determines the degree of satisfaction of the user of the service. Performance results show the proposed simulator to be an appropriate solution for studying several aspects of service driven radio network management, such as service planning, provisioning, performance monitoring and "optimisation". As a matter of fact, all simulations in this thesis were run using this tool. Secondly, using the simulator proposed in this work for radio interface dimensioning, it is shown how customised user satisfaction criteria may be defined for the implemented services, and how performance results, at a given QoE, can be presented as the maximum number of subscribers a WCDMA cell can satisfactorily accommodate. Such a simulator, in its simplicity, provides an appropriate platform for studying quickly the adverse effects on the existing services in WCDMA networks due to the provisioning of a new service. In the following, the proposed solution is used for analysing the deployment of PoC (Push to talk over Cellular) by means of four case studies. Simulation results show that no extra capacity is required for PoC launch.

5.1 Introduction

The cellular network planning process consists of network dimensioning, detailed network planning, network operation and "optimisation".

Network dimensioning (or initial planning) provides an estimate of the number of radio, transmission and core network elements and capacity of associated interfaces. Calculations thereof are based on operator's requirements for coverage, capacity and quality of service. In the detailed network planning, capacity and coverage are analysed for each relevant part of the network and interfaces between the entities in communication. This requires the knowledge of real traffic estimates and network topology for each analysed area, utilisation of accurate models for signal and user data transmissions, and implementation of the actual network element characteristics, functionalities and parameters. The cellular network "optimisation" can be seen as a process to improve the overall network quality as experienced by the mobile subscribers and to ensure that the network resources are efficiently utilised. This includes performance measurements, analysis of measurement results and updates of the network configuration and parameters. In this work, some aspects of QoS "optimisation" in UTRAN are presented in Chapter 7.

Lots of material covering the basics of radio, transmission and core network planning is available in the literature. For example, a comprehensive description of processes and tools for WCDMA radio network planning and optimisation can be found in [11] and [17]. Several methods for performance improvement of GSM, GPRS and EGDE networks were presented in [79]. Planning and optimisation aspects for 2G, 2.5G, 3G, and evolution to 4G, radio transmission and core were well covered in [85]. However, none of the published analytical methods and tools showed enough flexibility to provide an answer to what kind of service mix the operator is planning to provide, how these services will be implemented and how much money would be needed for the total rollout. The actual methodologies are based on certain required capacity and QoS. In addition crucial aspects of QoS functionalities in UTRAN are not taken into account.

The ability to model voice and data services, as well as the deployment of new service applications integrating voice, data and multimedia, such as presence, multimedia chat, PoC, and conferencing, is one of the key required ingredients for service planning. The traffic mix and volume need to be modelled differently depending on the carried service application characteristics. The RRM functions deployed in the real network, such as Power Control (PC), Admission Control (AC), Load Control (LC), and Handover Control (HC) with QoS differentiation, are essential features for the planning tools. Also, for each of the analysed services customised user satisfaction criteria need to be defined, and performance must be separately monitored and analysed for each of the studied services.

In this chapter, some of the key radio dimensioning and detailed planning issues that may arise from the deployment of multimedia services in WCDMA cellular systems are addressed using the models and assumptions described in Chapter 3. As already pointed out, in addition to what already described in the literature, in this work, plain planning methods and tools for service driven radio network dimensioning and planning are proposed. This encompasses the possibility of taking into account the offered service characteristics, discriminated treatment across the radio access network and service related performance monitoring. The proposed methods and tools allow network planners to assess the provision of the QoS negotiated for each of the deployed services and the maintenance of the data transfer characteristics, as well as to find an "optimum" trade-off between quality, capacity and coverage requirements for any of the services in a network operator's service portfolio.

The remainder of the chapter is organised as follows: Section 5.2 describes a use case on service driven detailed radio network planning using Ares. Section 5.3 presents some use cases on how the impact of PoC deployment on the existing services in a WCDMA cell can be studied using Max. Section 5.4 closes this chapter with a summary of key issues and conclusions upon the simulators utilisation for one cell or multi-cell scenarios.

5.2 Service Driven Radio Network Planning

To validate the capability of the virtual time simulator described in Section 3.5.1, several case studies were investigated and one of those, focusing on the deployment of the services reported in Table 3.1, is presented in the following sections. The provisioning of QoS is provided by service differentiation in admission control, load control, scheduling of bit rates and using service dependent values for the minimum allowed bit rates, maximum queuing times, inactivity timers, and granted DCH allocation times in normal situation and in the case of overload (see Section 4.2).

5.2.1 Specific Models and Simulation Assumptions

This case was introduced to study the benefit for the operator to offer speech, video and SWIS with guaranteed bit rate (GB) and all other services (including streaming (audio and video) and PoC, being in idle for most of the session duration) on best effort (NGB). Besides this, the service performance deterioration at high traffic volume when the bearers start to compete for radio resources was analysed. Therefore, speech and video calls were served as CS-Conversational, whereas SWIS was carried on PS-Streaming class. PoC, streaming and WAP/MMS were mapped onto PS-Interactive. (Note: WAP and MMS should be placed by the operator behind the same APN [73].) Dialup connections, which comprised, for example, ftp, emails and http traffic, were carried on PS-Background class. The RRP values were set such that speech calls had top priority, followed by video calls, and SWIS (real time video sharing). Within the Interactive class, using different traffic handling priorities (THP), PoC was handled first, followed by streaming and WAP/MMS. Dialup, on Background class, was the service with the lowest priority, and hence served last.

The differentiated parameter values to further improve PoC and streaming performance at the expense of lower priority services and the mapping of the services onto the distinct QoS profiles are illustrated in Table 5.1.

The simulation was performed using the macro cellular network depicted in Figure 3.2 over a period of 2 hours with a time step of 200 ms (RRI period). Terminals were randomly evenly distributed over the land area (not on the water). The traffic mix (see Table 3.4) and the traffic intensity were held constant, i.e. 2 call (or session) attempts per second. The corresponding offered traffic was about 750 users per cell over the entire simulated time. Table 5.2 summarises the most important network parameters. Other details on the assumptions and simulated environment can be found in Chapter 3.

Table 5.1. Mapping of services onto QoS classes and parameter values.

QoS Profile		Service	Bit Rate (kb/s)	RRP	Min. All. Bit Rate (kb/s)	AC Max. Queuing Time (s)	Granted Min. DCH Alloc. Time (s)	Granted Min. DCH Alloc.Time in Overload (s)	Buffering delay (s)	Inactivity Timer (s)	CR Max. Queueing Time (s)
GB	CS-conv.	Speech	12.2	1	GB	5	-	-	-	-	-
		Video	64	2	GB	10	-	-	-	-	-
	PS-stream.	SWIS	64	3	GB	10	-	-	5	-	-
NGB	PS-int. THP1	PoC	0, 8	4	8	15	15	10	4	60	4
	THP2	Streaming	0, 64	5	64	15	10	5	16	5	10
	THP3	WAP+MMS	0, 16, 32, 64, 128, 144, 256, 384	6	32	15	5	0.2	-	10	10
	PS-backg.	Dialup	0, 16, 32, 64, 128, 144, 256, 384	7	16	15	1	0.2	-	5	5

Table 5.2. Most important system based parameters.

Parameter	**Value**
Call or session mean arrival rate	0.5 s
Radio resource indication period (*RRI*)	0.2 s
Simulation time (s)	7200 s
Power target for DL AC	3 dB below BTS total power
Overload offset for DL AC	1 dB above power target
Orthogonality (α)	0.5
Period for load control actions	0.2 s (1 RRI)
Period for Packet Scheduling	0.2 s (1 RRI)
Length of AC queue	10 Radio bearers
Dedicated NGB capacity	0 W, i.e. not used
Priority step (x)	0
Power weight for inactive NGB traffic (k)	0.5
Packet Scheduling Algorithm	Fair Throughput

5.2.2 Simulation Results and Discussion

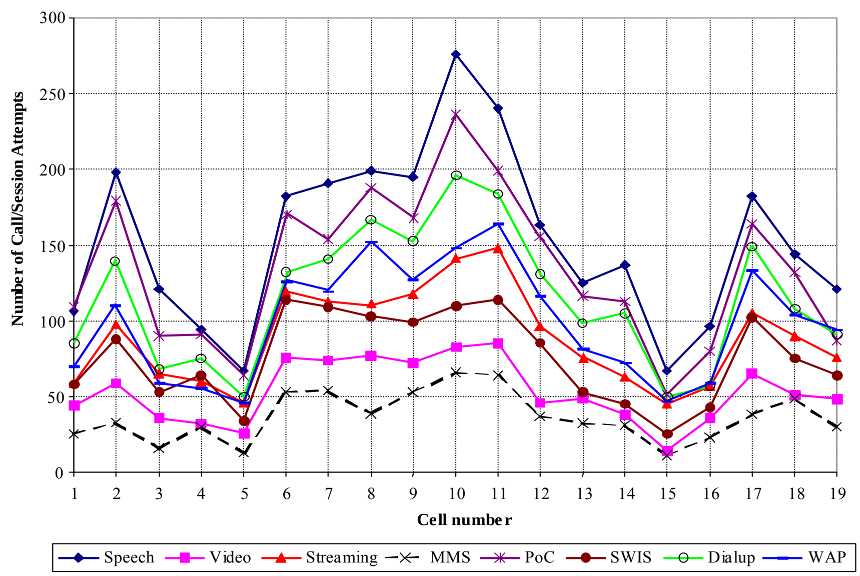
The simulation results are shown through Figure 5.1 to Figure 5.7 and in Table 5.3, where a summary of the system based performance measurements is reported.

The offered traffic mix in number of call or session attempts over the simulated period is illustrated for each of the analysed cells in Figure 5.1 a). Figure 5.1 b) depicts the number of average active connections in the cells during the radio resource indication period (RRI). The total (service based) average cell throughput during the simulation period is illustrated as a function of the planned cells in Figure 5.2. From the graphs, taking into account the traffic models and mix reported in Table 3.1 and Table 3.4, respectively, we can conclude that the traffic generator worked as planned. In fact, the offered traffic mix corresponds to the input share of calls values, and the average cell throughput is in line with the served traffic. Though not shown in the figures, the peak values of the total cell throughput ranged up to 1.5 Mb/s. The results in Figure 5.1 b) show ~4 to 24 connections on average active per cell during the sampling period (cell deployment variation). For CS calls the measure may be connected to the corresponding offered load in erlangs considering the call arrival intensity of Figure 5.1 a), or simply the call/session mean arrival rate and the share of calls of that particular service type, and the mErl per user, which can be derived from Table 3.1. Similar calculations were made in Section 3.5.2.1, for generating the traffic presented in

Section 5.3. As explained in Section 1.2.2, for PS applications, such analytical approach is not possible for an accurate modelling of the problem. This is also visible from the figures, where there is no exact scaling factor between offered and served load over the measured cells. The correlation factor between the two values depends on the traffic pattern, system parameterisation and RRM algorithms. As a matter of fact, Ares is proposed in this work for studying different traffic mix scenarios and/or network configurations. Considering the parameter settings of Table 5.2, with a planned power target (denoted as $P_{TxTarget}$ in (4.1) and (4.2)) 3dB below the maximum transmission power of the base station, the intensity of the offered traffic may be regarded as high for NGB applications. In fact, as shown in the following figures, the QoE of such services was rather poor in most of the simulated cells.

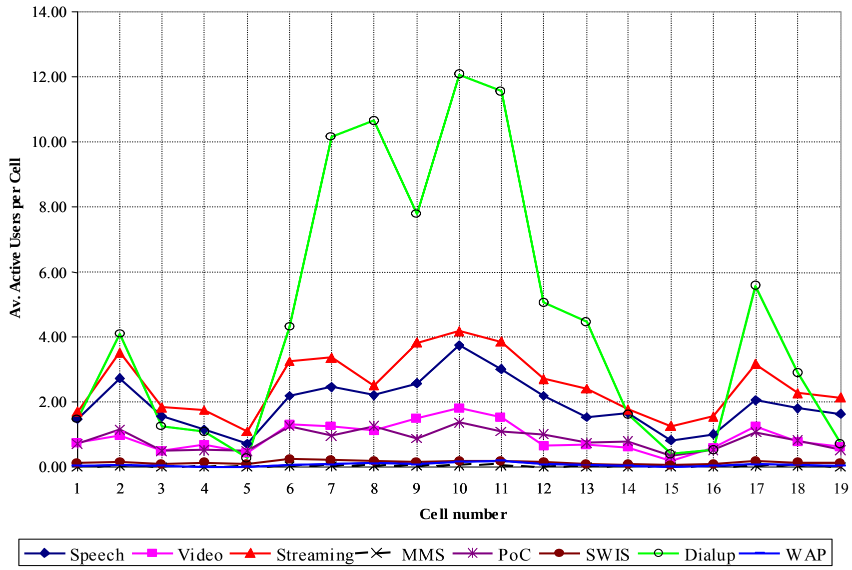
A snapshot (1000 s) of the load status in the Cell 11 (see Figure 3.2), where most of the users resulted unsatisfied, is depicted in Figure 5.3 a). The values plotted against the simulation time are the transmission powers, the power budget (*PB*) and scheduled capacity (*SC*). From this figure, the dynamics of the supported QoS management functions can be observed. At a given NGB traffic volume and available bit rates, the curve of the allocated bit rates (*SC*) by PS closely follows the power budget (*PB*) one. Besides this, when the total load in the cell ($P_{TxTotal}$, i.e. $P_{GB} + P_{NGB}$) reaches the overload threshold, denoted by $P_{TxTotal}$ + *Offset*, the NGB allocated bit rates are accordingly reduced by LC. In turn, throughputs are immediately resumed when the P_{GB} decreases and the system backs off to its normal state of operation. Ultimately, the measured GB load (P_{GB}) hardly ever exceeds the target threshold ($P_{TxTarget}$) and, following the input traffic mix, $P_{TxTotal}$, thanks to the AC and PC functions, always preserves its point of equilibrium ($P_{TxTarget}$). The normalised distribution functions of the radio link power (P_{TxLink}), GB transmission power (P_{GB}), NGB transmission power (P_{NGB}) and total downlink transmission power ($P_{TxTotal}$) are shown in Figure 5.3 b). From these statistics taken over the simulated period analogous conclusions in terms of transmission powers can be drawn.

Let us now examine the performance experienced by the users of the deployed services. As shown in Table 5.3 in terms of call block ratio (CBR), none of the calls/sessions were rejected due to buffer overflow and almost none of the speech, video and SWIS calls were blocked due to the time spent in the AC queue. This is due to the fact that the offered GB traffic hardly ever exploited the available cell capacity (power), $P_{TxTarget}$ in (4.1) and (4.2). As a result, almost all users of the GB services resulted satisfied. The favourable effects of the prioritisation between GB services can be noticed in Figure 5.4 a), where the percentage of satisfied users for each of the deployed services is illustrated as a function of the simulated cells. In Cell 10 and 11, the quality experienced by speech users, in terms of CBR, is better than the accessibility offered to SWIS and video users. In the figure, the differentiated treatment of NGB traffic is more relevant, which reflects the provisioned discrimination between real time (RT) and non-real time (NRT) services of Interactive and Background classes. In fact, in each of the simulated cells, the percentage of satisfied users of Dialup is the lowest, followed by WAP and MMS; and for the RT services, PoC performance is always better than the streaming one. Analogous results for NGB services are depicted in Figure 5.5, where the position of the dissatisfied users is marked with an asterisk of different colour depending on the used service. This performance overview enables the operator to capture the benefit provided by the service differentiation and limitations thereof in high traffic scenarios. In addition, the most critical areas in the network where more capacity is needed may be identified. In this case, the NGB QoE is definitely unacceptable in Cell 11; can be tolerable in Cell 7, 10, and 13; results more or less satisfactory in all the other cells. A more detailed analysis upon the reason why the users of NGB services were not satisfied is possible based on the raw performance indicators illustrated in Figure 5.4 b), Figure 5.6, and in Figure 5.7 a). In these figures, the intended differentiation between services in terms of the metrics characterising the QoE of each of the deployed services is also visible. In particular, Figure 5.4 b) and Figure 5.6 a) show, respectively, the 10th percentile of the average active session throughput (AST) and the capacity request rejection ratio (CRRR) collected for each of the above services during the measurement period, as a function of the deployed cells. The throughput experienced by Dialup users is lower than the corresponding one offered to WAP/MMS users, which underwent the same treatment. Conversely, the accessibility offered to PoC and streaming services, while requesting capacity to PS, is better than the corresponding blocking experienced by WAP/MMS and Dialup users. Figure 5.6 b) reveals how the throughput deterioration adversely affects the PoC and streaming performance in terms of re-buffering in the UE. As expected, the re-buffering ratio is higher for streaming, whereas the dissatisfaction due to "too long time to refill up the buffer" depicted in Figure 5.7 a) is higher for PoC. This is due to the fact that the tolerance for streaming users (up to sixteen seconds) was higher than for PoC ones, which were not supposed to wait for more than four seconds (see Table 5.1). The differentiation between WAP/MMS and Dialup connections and the benefit thereof is also shown in Figure 5.7 b) and Table 5.3, where the transfer delays (system based statistics upon all simulated time) of the WAP, MMS and Dialup objects during the simulation period are presented.



a)

Figure 5.1. a) Traffic distribution over the 19 cells: Offered load in call arrivals.



b)

Figure 5.1. b) Traffic distribution over the 19 cells: Served load on average active connections taken over each of the simulated radio resource indication period.

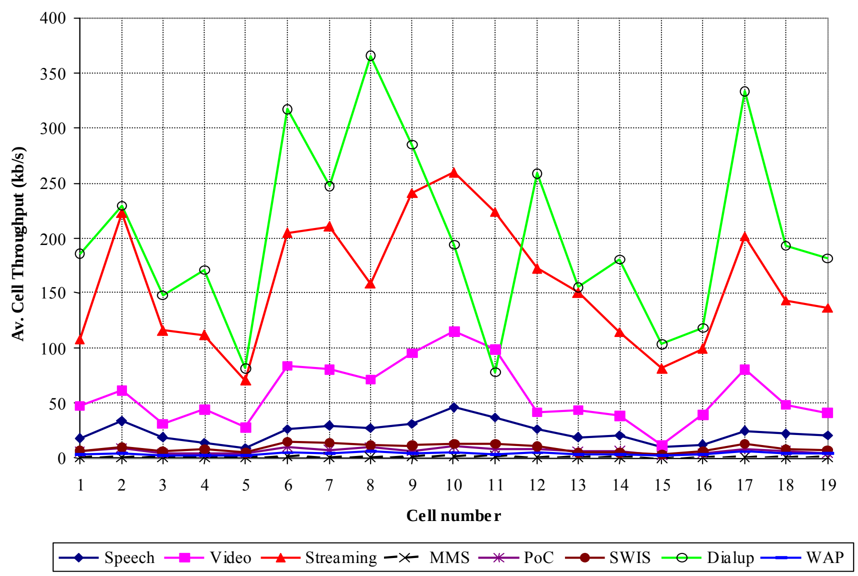
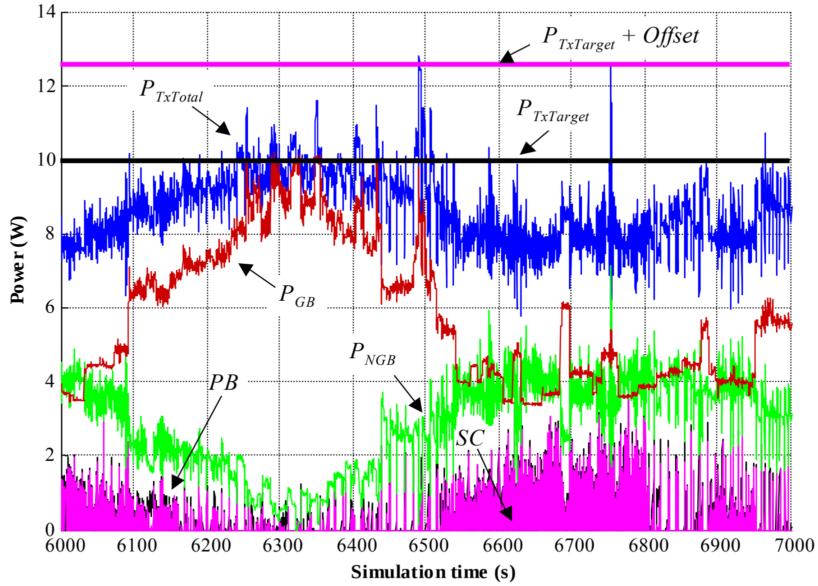
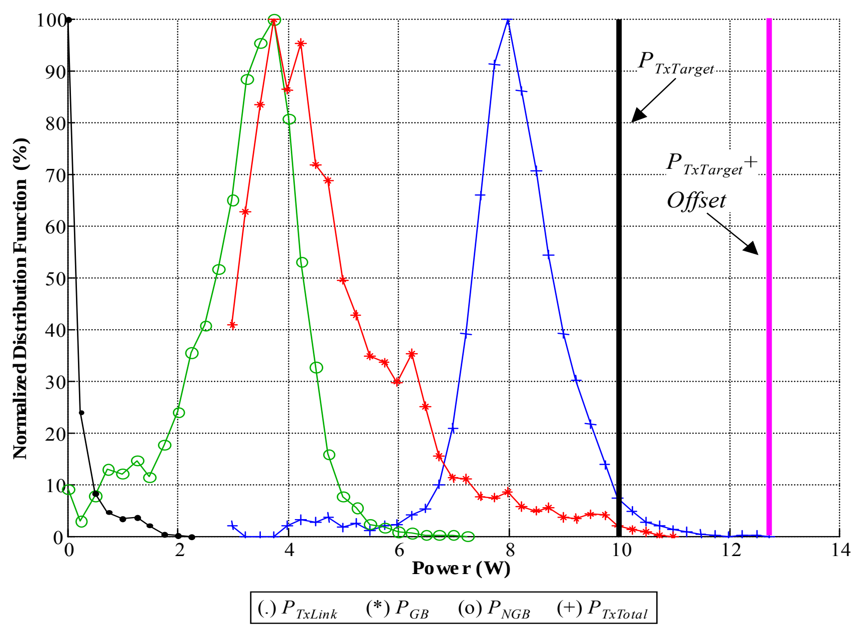


Figure 5.2. Traffic distribution over the 19 cells: Average cell throughputs (service based values in kb/s for each of the simulated cells).



a)



b)

Figure 5.3. Cell 11: a) Snapshot of the simulation period; b) normalised distribution functions of the radio link (.), GB (*), NGB (o), and total (+) downlink transmission power.

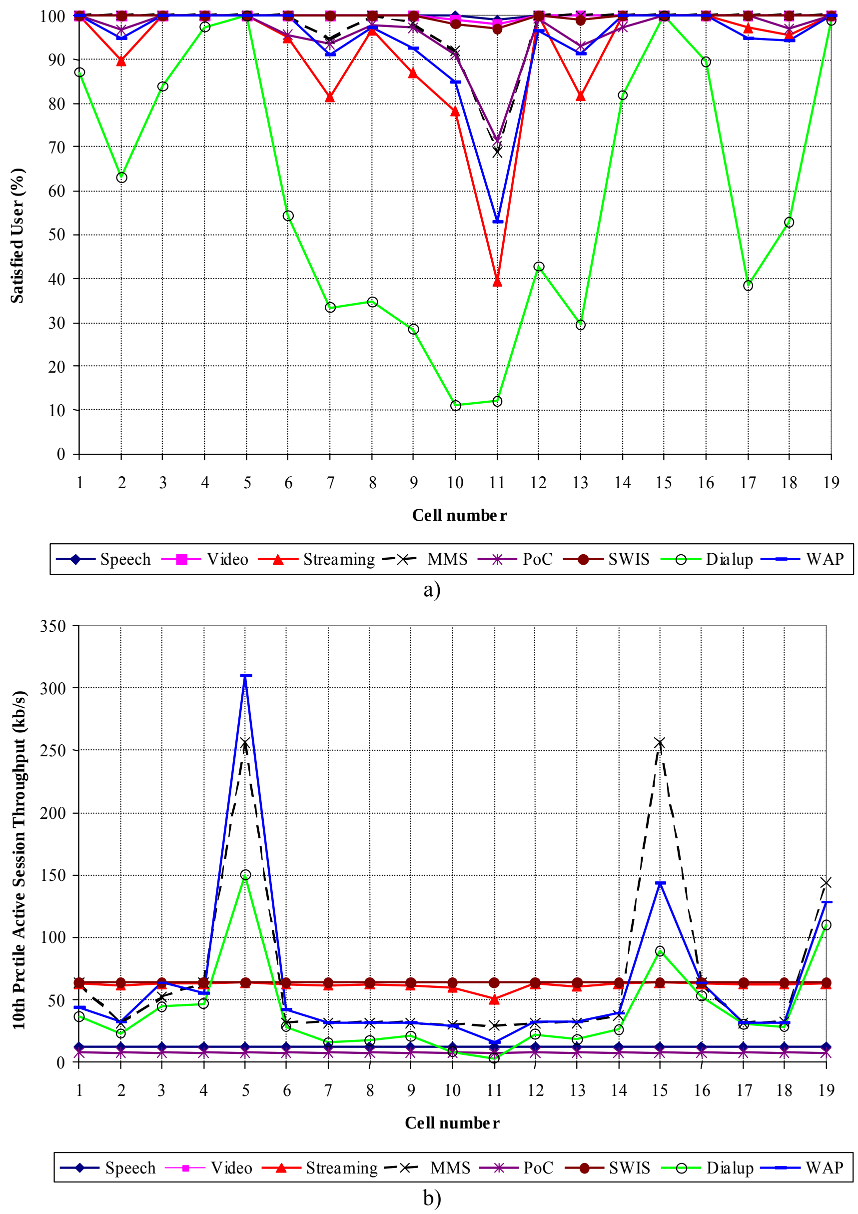


Figure 5.4. Service based indicators for each of the simulated cells: a) Percentage of satisfied users; b) 10^{th} percentile of the average active session throughput during the simulated time.

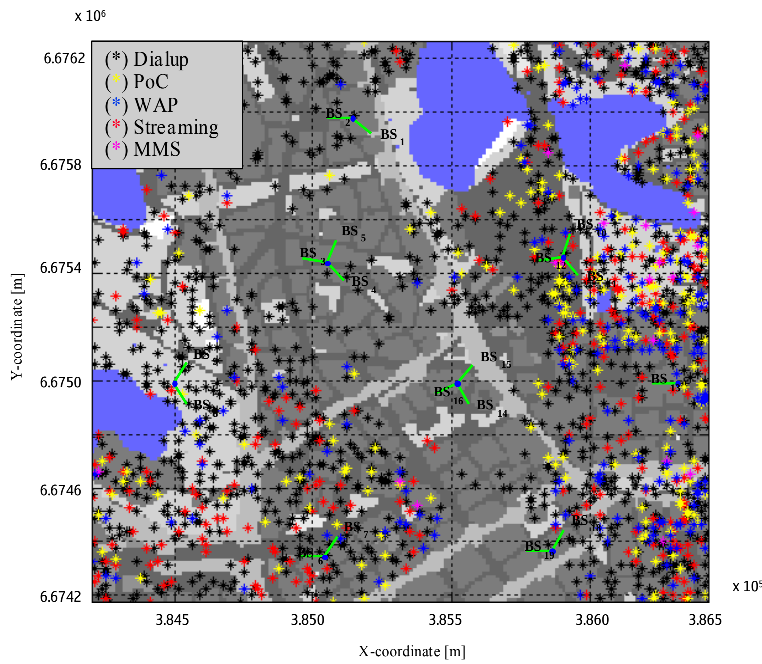
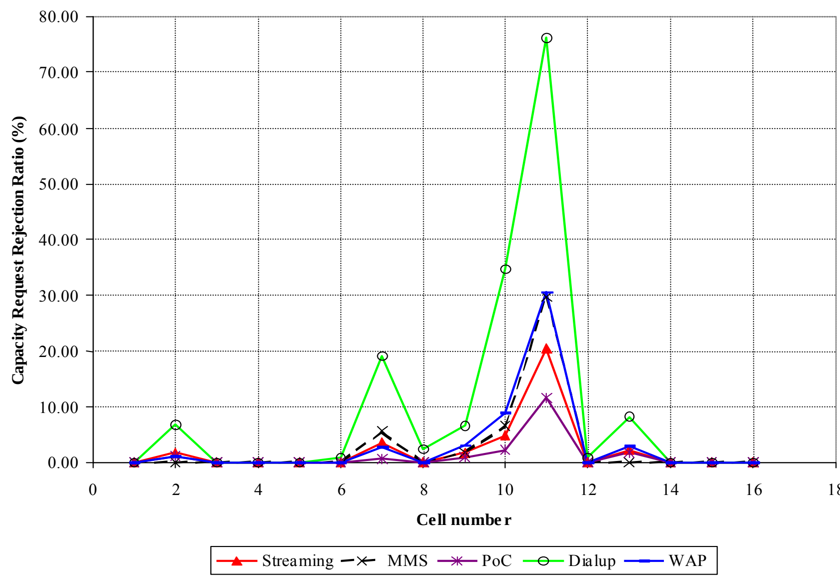
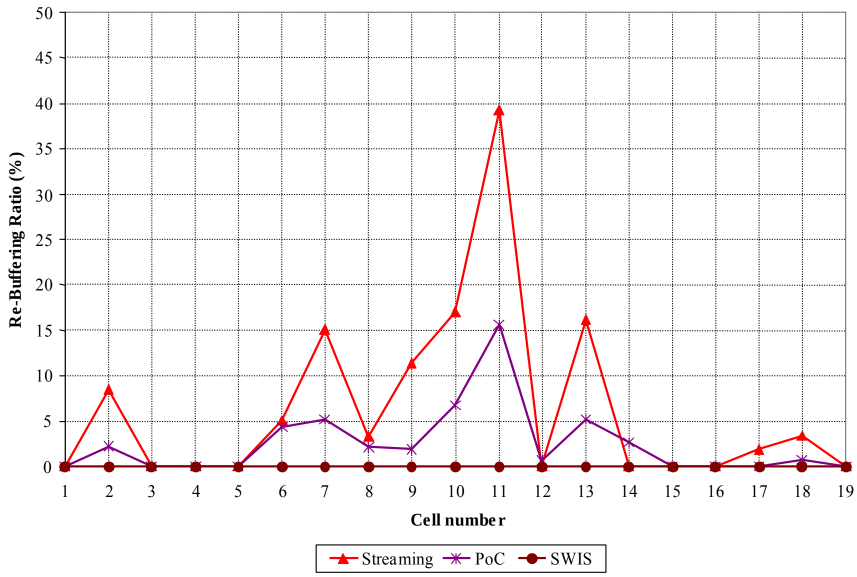


Figure 5.5. Dissatisfied users of Dialup, PoC, WAP, streaming and MMS services.



a)

Figure 5.6. a) Service based performance indicators for each of the simulated cells: Capacity request rejection ratio (CRRR).

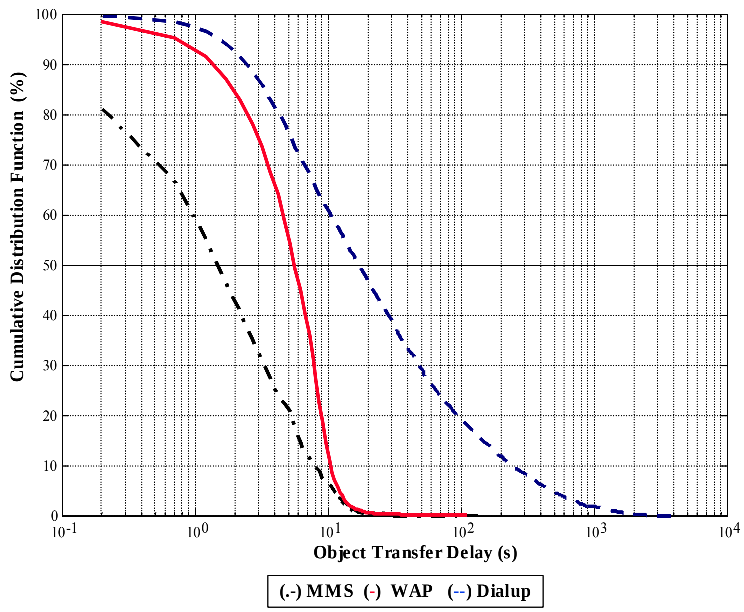


b)

Figure 5.6. b) Service based performance indicators for each of the simulated cells: Re-buffering ratio (RBR).

a)

Figure 5.7. a) Too long time needed for re-buffering ratio (service based performance indicators for each of the simulated cells).



b)

Figure 5.7. b) MMS, WAP and Dialup object transfer delays (system based statistics upon all simulated time).

Table 5.3. System based measurement results.

Service Type	CBR (%)	CDR (%)	CRRR (%)	RBR (%)	DBR (%)	Median AST (kb/s)	Median Object Size (kB)	Calculated Object Delay (s)	SU (%)
Speech	0.05	0.00	-	-	-	12.2	-	-	99.95
Video	0.16	0.00	-	-	-	64.0	-	-	99.84
Streaming	0.00	0.00	1.93	6.35	0.05	63.4	1682	212.2	91.67
MMS	0.00	0.00	2.29	-	-	70.5	15	1.7	97.55
PoC	0.00	0.03	1.08	2.49	0.19	8.0	4	4.0	96.31
SWIS	0.32	0.00	-	0.00	0.00	64.0	89	11.1	99.68
Dialup	0.00	0.00	8.44	-	-	51.4	120	18.7	59.94
WAP	0.00	0.05	2.57	-	-	66.0	48	5.8	94.24

Note: RBR = Re-Buffering Ratio, DBR = Delay Buffering Ratio; SU = Satisfied Users

5.3 WCDMA Cell Dimensioning

In this section, four case studies using the simplified version of Ares (denoted as *Max*, in the thesis), described in Section 3.5.2, are presented. In the case studies, Max is used for evaluating rapidly the influence of the deployment of PoC on the maximum number of subscribers a WCDMA cell can accommodate in the downlink within the constraints defined by minimum percentages of satisfied users.

5.3.1 Specific Models and Simulation Assumptions

In this section, four case studies concerning the effects of PoC deployment on the quality of the offered services in a WCDMA cell are presented. The services are mapped onto QoS profiles with different priorities as described in Section 3.5.2.1, i.e. Speech calls have top priority, followed by CS video and SWIS (RTVS) applications. Within the NGB traffic, the bit rate to PoC users is allocated (scheduled) first, followed by streaming, WAP and MMS, and Dialup connections, which are served last. The most important cell based parameters are listed in Table 5.4 and the traffic models and mix are reported Table 5.5.

5.3.1.1 Case 1

This case focuses on the deployment of PoC carried by a bearer with a maximum bit rate of 8 kb/s. The average PoC group size is one, which means that one user is allowed to call only one person at a time. The only modified parameters of the distinct services are the allocated priorities.

5.3.1.2 Case 2

This case uses the same parameter settings of Case 1, but PoC is offered with a maximum allowed bit rate of 16 kb/s (See Section 4.2.2).

5.3.1.3 Case 3

Parameters are set as in Case 2, but now the PoC group size is on average composed by 4 users. It is further assumed that a given PoC user will press the PoC talk key as often and speak as long as in Case 2. As a result, a PoC user will generate on average four times more traffic in the cell than in Case 2, and eight times more than in Case 1.

5.3.1.4 Case 4

Parameters are set as in Case 3, but the input load for PoC is gradually increased, whereas for all other services the offered traffic volume in the cell is fixed to 500 users. This case illustrates which are the services that are likely to be affected by the PoC penetration.

5.3.2 Simulation Results and Discussion

This section presents the simulations results of the case studies introduced above. For Case 1-3, the served PoC traffic is monitored in terms of absolute and relative cell throughput, whereas, the effects of PoC deployment on the other existing services are characterised by the maximum number of satisfactory non-PoC subscription in the cell and the variation thereof in percentage. In Case 4, we display the course of the average cell throughput and the percentage of satisfied users as a function of the number of PoC subscribers.

Table 5.4. Most important cell based parameters.

Parameter	**Value**			
Number of iterations (inner loop)	1000			
Downlink Load target	70 %			
Overload offset	10%			
Orthogonality (α)	0.5 (ITU Vehicular A)			
Soft handover overhead (*SHO*)	20%			
Other-to-own cell interference ratio (*i*)	0.55			
Chip rate (*W*)	3.84 Mchip/s			
Offered services - Traffic class	**DL DCH bit rates (kb/s)**	**Priority**	**E_b/N_0 (dB)**	**Activity Factor (ν)**
Speech - CS Conversational (GB)	12.2	1	7	0.67
Video - CS Conversational (GB)	64	2	6	1
SWIS - Streaming (GB)	64	3	6	1
PoC - Interactive THP1 (NGB)	0, 8, 16	4	6	1
Streaming - Interactive THP2 (NGB)	0, 64	5	7	0.6
WAP/MMS - Interactive THP3 (NGB)	0, 64, 128, 144, 256, 384	6	5/5.5	1/0.6
Dialup – Background (NGB)	0, 64, 128, 144, 256, 384	7	5.5	0.8

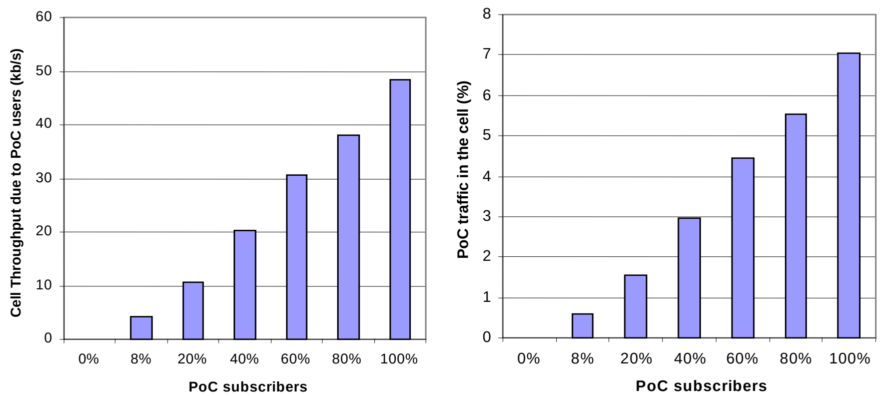
Table 5.5. Adopted traffic models and mix.

Offered service	**Share Of Subscriptions (%)**	**Mean Service Time (s)**	**Mean Arrival Intensity (Hz)**
Speech (CS)	100	90	1/4800
Video (CS)	3	120	1/24000
Streaming	10	600	1/(5*3600)
MMS	10	10	1/(2*3600)
SWIS (RTVS)	3	180	1/(2*3600)
Dialup	1	1200	1/(2*3600)
WAP browsing	20	600	1/(4*3600)
PoC	Varies*	60	1/(2*3600)

* The volume is increased from 0 to 100%, whereas the average PoC group size is held constant: 1 user in Case 1 and Case 2, 4 users in Case 3 and Case 4.

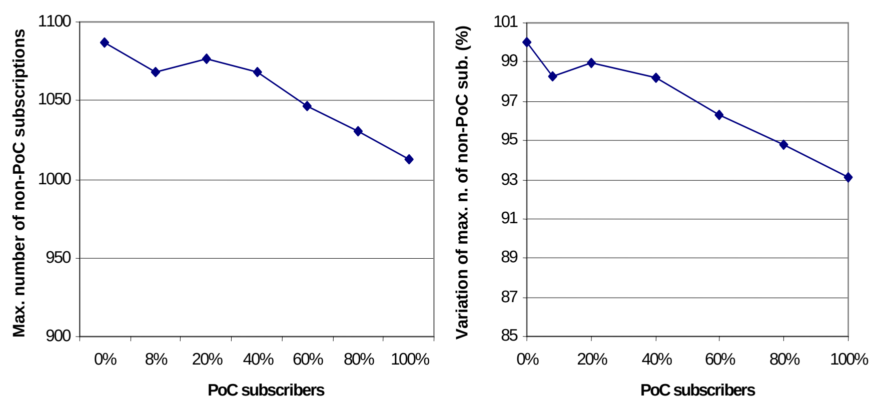
5.3.2.1 Case 1

Simulation results are illustrated in Figure 5.8 and Figure 5.9. The former depicts the served PoC traffic in the cell, whereas the latter shows the impact of PoC on other services. From Figure 5.8 a) and b), we can notice that if all subscribers made use of the PoC service, the average cell throughput due to PoC traffic (7% of the total traffic in the cell) would be less than 50 kb/s. Hence, as depicted in Figure 5.9, the impact of PoC usage on the other services is really small: The performance deterioration in the cell would be noticeable only when more than 40% of the end-users subscribed to PoC, and only 7% of the non-PoC subscriptions would not be satisfactory if all users made use of PoC.



b)

Figure 5.8. Case 1 – a) Average cell throughput due to PoC users, and b) Percentage of PoC traffic in the cell as a function of the percentage of PoC subscribers in the cell. (Note: 100% means that all subscribers would subscribe to PoC service.)

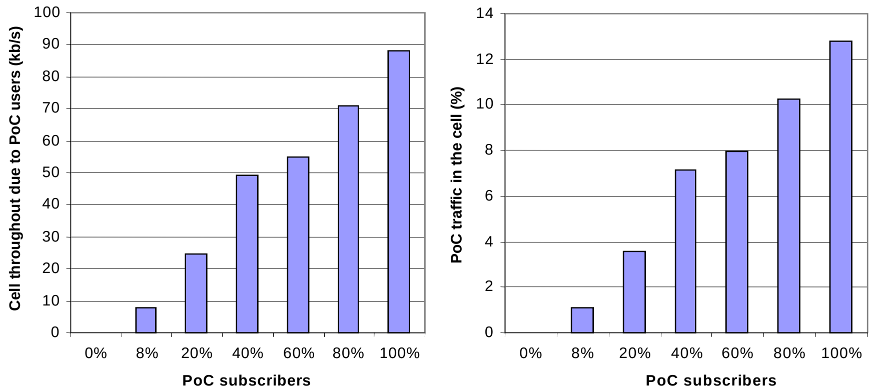


b)

Figure 5.9. Case 1 – a) Maximum number of satisfactory non-PoC subscriptions, and b) variation in percentage thereof as a function of the percentage of PoC subscribers in the cell. (Note: 100% means that all subscribers would subscribe to PoC service.)

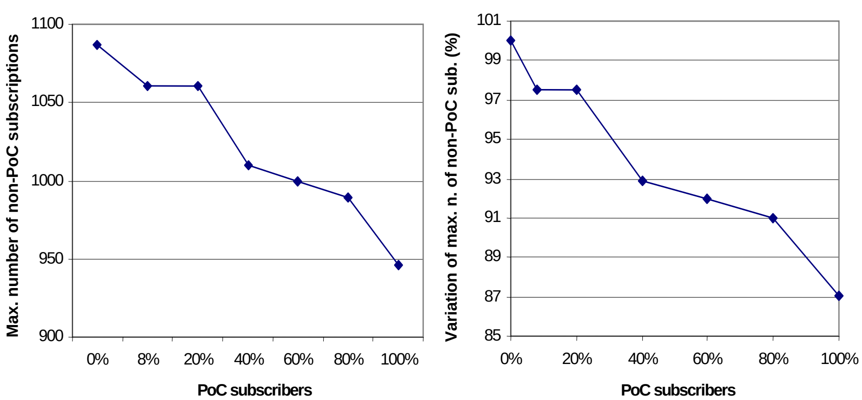
5.3.2.2 Case 2

Simulation results are illustrated in Figure 5.10 and Figure 5.11. The former depicts the served PoC traffic in the cell, whereas the latter shows the impact of PoC on other services. Figure 5.11 reveals that, also in this case, PoC slightly deteriorates the performance of other services. If all users subscribed to PoC the corresponding generated traffic would require on average only ~13% of the cell capacity; and the maximum number of non-PoC subscribers satisfactorily handled in the cell would decrease by ~13%, which is, as expected, almost twice as much the corresponding traffic served in Case 1.



a) b)

Figure 5.10. Case 2 – a) Average cell throughput due to PoC users, and b) Percentage of PoC traffic in the cell as a function of the percentage of PoC subscribers in the cell. (Note: 100% means that all subscribers would subscribe to PoC service.)



a) b)

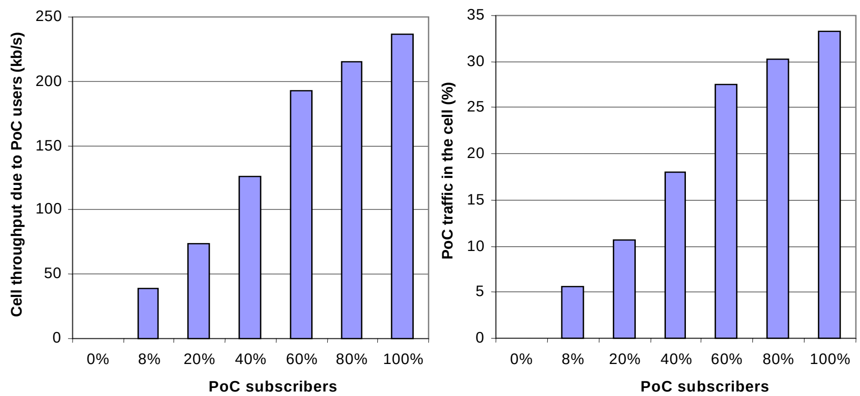
Figure 5.11. Case 2 – a) Maximum number of satisfactory non-PoC subscriptions, and b) variation in percentage thereof as a function of the percentage of PoC subscribers in the cell. (Note: 100% means that all subscribers would subscribe to PoC service.)

5.3.2.3 Case 3

Simulation results are illustrated in Figure 5.12, Figure 5.13 and Figure 5.14. Figure 5.12 depicts the served PoC traffic in the cell, whereas Figure 5.13 and Figure 5.14 display the influence of PoC on other services. In this case, PoC service has a significant impact on the maximum number of non-PoC subscribers the WCDMA cell can accommodate. If less than 20% of the subscribers subscribed to PoC the corresponding generated traffic would require ~10% of the cell capacity, and the maximum number of non-PoC subscribers satisfactorily handled in the cell would be decreased by only 10%. Furthermore, if all users subscribed to PoC, on average, one third of the cell capacity would be exploited by PoC traffic. Figure 5.14 shows that the course of the curves displayed separately for each of the services deployed in the cell is rather similar to the slop of the graph depicted in Figure 5.13. This is due to the fact that the simulation ends when any of the dissatisfaction criteria presented in Section 3.5.2.3 is verified. Case 3 is the "worst scenario", where the mean PoC group size is fairly high (4) and the group members are all and always located in the same cell.

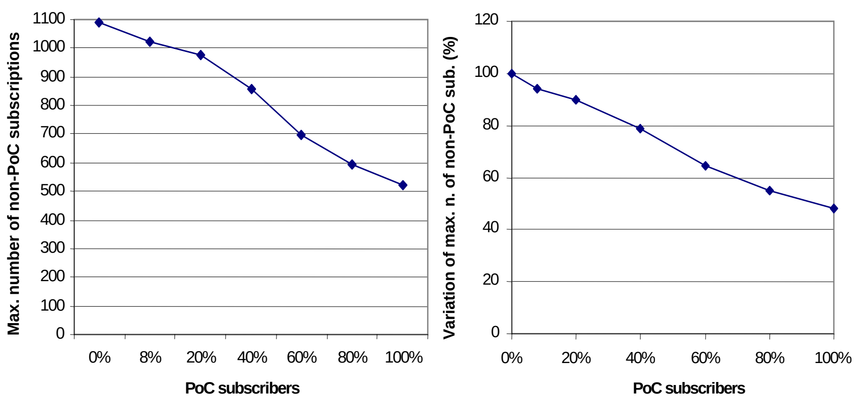
5.3.2.4 Case 4

As already pointed out, whereas in previous cases the goal was to compute the maximum number of subscribers a cell can properly accommodate with or without the PoC presence, the main objective of this case is to quantify the impact of the PoC penetration when the number of subscribers of other services is kept constant. In general, the traffic volume in the cell without PoC is for the operator to determine, yet, in this study, it was chosen such that all percentages of satisfied users of the offered services were above the quality thresholds defined in Section 3.5.2.3. Figure 5.15 depicts the average cell throughput for GB, NGB and PoC traffic as a function of PoC subscribers' percentage. As expected, when the PoC traffic increases, the NGB load decreases, whereas the load due to GB services remains constant. This is due to the fact that PoC has the highest priority among NGB services (see Table 5.4) and from (3.7) has no means to affect the admission of GB services. However, from this figure we cannot draw any conclusion on the influence of PoC on the performance of other services, as the quality experienced by the users of NGB applications is not shown. These effects can be notice in Figure 5.16, where the percentage of satisfied users is depicted separately for each of the offered services in the cell. When the percentage of PoC subscribers goes beyond 50%, the bit rate allocated to WAP connections is too low and thus resulting in too many WAP dissatisfied users. This would indicate the operator to add extra radio capacity in the cell, e.g. introduce one more carrier.



a) b)

Figure 5.12. Case 3 – a) Average cell throughput due to PoC users, and b) Percentage of PoC traffic in the cell as a function of the percentage of PoC subscribers in the cell. (Note: 100% means that all subscribers would subscribe to PoC service.)



a) b)

Figure 5.13. Case 3 – a) Maximum number of satisfactory non-PoC subscriptions, and b) variation in percentage thereof as a function of the percentage of PoC subscribers in the cell. (Note: 100% means that all subscribers would subscribe to PoC service.)

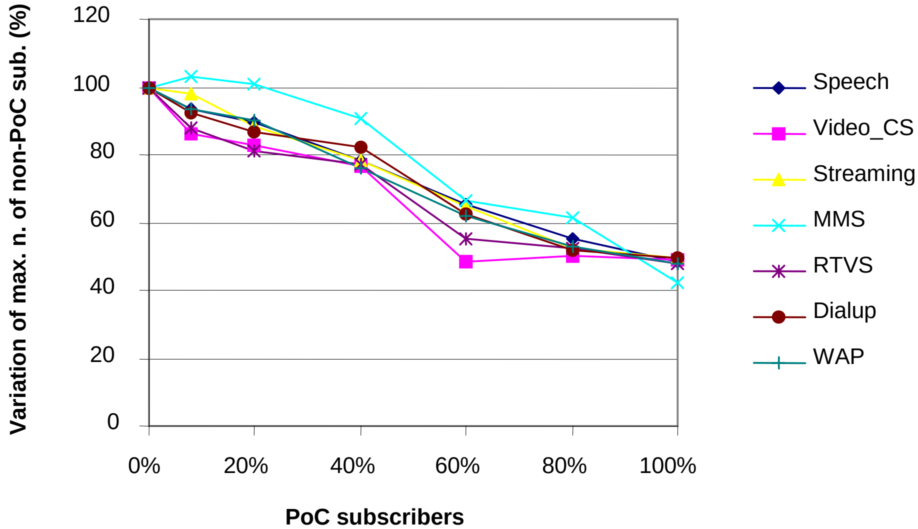


Figure 5.14. Case 3 – Variation of the maximum number of satisfactory non-PoC subscriptions as a function of the percentage of PoC subscribers in the cell. (Note: 100% means that all subscribers would subscribe to PoC service.)

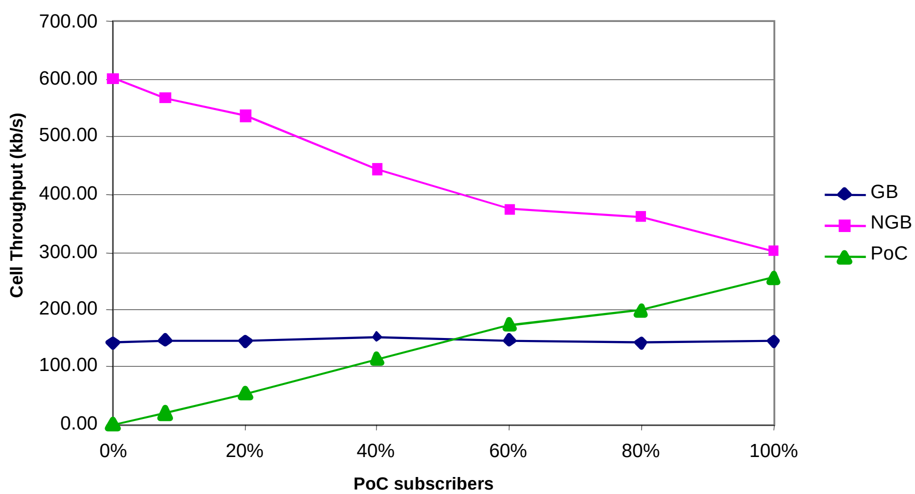


Figure 5.15. Case 4 – (GB) Guaranteed Bit Rate, (NGB) Non Guaranteed Bit Rate (without PoC) and PoC average cell throughput (kb/s) as a function of the percentage of PoC subscribers in the cell. (Note: 100% means that all subscribers would subscribe to PoC service.)

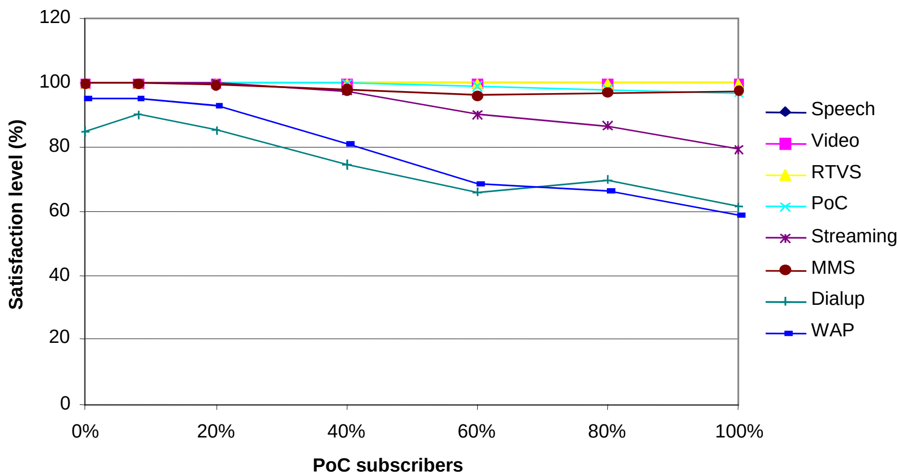


Figure 5.16. Case 4 – Satisfaction levels per service as a function of the percentage of PoC subscribers in the cell. (Note: 100% means that all subscribers would subscribe to PoC service.)

5.4 Conclusions

A use case on how to study the provisioning of QoS before the deployment of new services throughout UTRAN was presented using Ares. The supported power control, admission control, load control, and packet scheduler functions, based on priorities and differentiated parameters, were explored in terms of: Offered and served traffic mix, transmission powers, throughputs, queuing times, object transfer delays, call block ratios (CBR), call drop ratios (CDR), capacity request rejection ratios (CRRR), and QoE by means of simulations. From the simulation results, the proposed tool turns out to perform as intended, and appears to be a good trade off between the complexity of an advanced dynamic simulator and the simplicity of a static tool. Thus, the proposed simulator has the potential for investigating any QoS management function and multimedia service provisioning in UTRAN, before its deployment throughout a real WCDMA radio access network. As a part of this framework, a simplified version of tool (limited to one cell and based on throughput and snapshots of the cell status) was used to analyse the impact of PoC on the performances of the other deployed services in a WCDMA cell. Simulation results show that the maximum number of subscribers a cell can accommodate, at a given QoE, depends heavily on the assumed subscriptions percentage of each of the offered services. Nevertheless, the radio interface does not appear to be a bottleneck when the PoC service is launched. In the worst scenario, additional radio resources may be required when the number of PoC subscriptions exceeds the 20% of the existing customers. Using the proposed simulator, the influence of the penetration of any service can be rapidly evaluated by keeping the subscription levels for all other services constant and increasing the traffic volume of the application in question.

6

QoS Monitoring

UMTS provides an opportunity for operators to offer new services to their potential and existing subscribers. Unlike circuit switched services, packet switched services are dependent on a multitude of factors to guarantee the QoS of the content delivery. To ensure that the contracted QoS is sustained, it is not sufficient to overprovision resources. In addition, operators need to have tools for monitoring and analysing the QoS experienced by diverse packet flows in different network segments. This chapter presents a simple mechanism to monitor the quality of distinct services across UMTS radio access networks, which enables a proactive and timely supervision of service level agreements and quality. The approach is based on the concept of UMTS bearer, which maps services with different performance requirements onto distinct QoS profiles. By means of a subset of bearer service attributes, e.g. bit rates, priorities and traffic classes, it is possible to formulate metrics to measure separately the performance of distinct services within the RNS, in the uplink (UL) and downlink (DL) directions, without any visibility of the content carried by upper layer protocols. From the classified performance indicators is then possible to derive the system load and the amount of satisfied users each cell currently accommodates.

6.1 Introduction

Although QoS monitoring is a major aspect of differentiation of service providers, little material is available in the literature to describe how essential performance metrics can be measured for a proactive management of an end-to-end GPRS service performance. In [38], only the challenges involved in providing QoS distribution monitoring were discussed, and the proposed approaches to meet these challenges were far from any practical cost-effective service assurance solution. An interesting approach to end-to-end service assurance on the mobile Internet was presented in [39]. Yet, only a management model and architecture of the service performance management system and related capabilities were introduced. In [40], Telia-Sonera presented an interesting performance evaluation of their GPRS network accomplished by a combination of measurements at the end hosts and tracing user and control plane data at U_m (radio interface), A_{bis} (interface between the Base Station, BS, and its Controller, BSC) and G_b (interface between BSC and Serving GPRS Support Node, SGSN). For throughput measurements, they used tools generating bulk transfers over TCP and dumped data at the end hosts. Ultimately, standardisation forums define QoS attributes and specify what to measure, but not how to monitor the actual QoS management functions in the user and control plane: This is left to the vendor and operator's choice.

This chapter describes how the uplink and downlink exploited capacity and delivered QoS can be measured in UTRAN, and the classification of counters to assess performance of protocols carrying service applications through the radio interface. As a part of this framework, we explain how the performance of dedicated uplink transport channels, such as E_b/N_0, block and bit error rate, can be derived by the radio access controller starting from some of the mandatory measurements specified for UTRAN [58]. This method makes it also possible for the network administrator to avoid drive tests as e.g. done in [86] to measure indirectly the uplink performance from the UE transmission power.

Our target is to provide means and methods to monitor separately and in a cost-effective way higher layer service applications, without tracing and/or explicitly disclosing the characteristics of upper layer protocols by means of any tool or protocol entity in UTRAN. The proposed measurements collected by the radio network subsystem (RNS) provide an essential input to ensure quality compliance to service layer management commitments. A similar work was presented in [59] for EGPRS networks, where the proposed method was also experimentally validated in the laboratory.

The reminder of this chapter is organised as follows: Section 6.2 defines the application performance monitoring based on bearer service characteristics. Section 6.3 describes the essential counters to be collected in the RNC and the classification thereof. Besides this, in this section, we derive some of the most important indicators for the uplink and downlink differentiated performance monitoring and how the quality of upper layer protocols can be assessed through a proper categorisation of the proposed performance metrics. Conclusions are drawn in Section 6.4.

6.2 Performance Monitoring Based on Bearer Characteristics

This section defines the concept of performance monitoring of application protocols based on the characteristics of the UMTS bearer service, i.e. a subset of attributes of the PDP context QoS profile.

When a PDP context is established, the attributes of the negotiated QoS profile for that particular UMTS bearer are available in the UE, RNC, SGSN and corresponding GGSN. (For more information see Section 2.7.) To assess the carried applications, at the different network elements, i.e. RNC, SGSN and GGSN, performance of bearer services needs to be collected based on a subset of attributes of the subscribed QoS profile, e.g. Traffic Class (TC), Traffic Handling Priority (THP) for Interactive class, Allocation Retention Priority (ARP), and bit rates (maximum and guaranteed), which unambiguously relate to the offered service (one to one mapping between a subset of the PDP context attributes values and upper layer protocols). In this work, the term *management class* (MC) denotes such a limited combination of attributes.

(Note: For interworking purposes between different releases, the following mapping rules between GPRS Release 97/98 (R97/98) and GPRS Release 99 (R99) as well as UMTS are defined [1]: Delay Class = 1 corresponds to Interactive THP = 1, Delay Class = 2 corresponds to Interactive THP = 2, Delay Class = 3 corresponds to Interactive THP = 3 and Delay Class = 4 corresponds to TC = Background. Furthermore, for consistent traffic treatment through the mobile network, one needs to set ARP = THP = Precedence Class = Delay Class. Background class (Delay Class = 4) is mapped by the 2G SGSN to ARP = Precedence Class = 3.)

Hence, in theory, without taking into account the bit rate and quality attributes, 10 MCs (PS Conversational ARP1-3, PS Streaming ARP1-3, Interactive THP1-3, and Background) can be defined for legacy and new terminals. The practical MC definition and mapping of services onto MCs is for the operator to manage.

Figure 6.1 illustrates an example of 7 MCs and a possible mapping of distinct services onto different MCs. See What I See (SWIS) and Push to talk over Cellular (PoC) can be offered with the highest priority on MC 1 (PS Streaming class, ARP=1) and MC 2 (PS Streaming class, ARP=2), respectively. Audio and video streaming applications can be carried using either MC 3 (PS Streaming class, ARP=3) with guaranteed bit rate or using MC 4 (Interactive class, THP=ARP=1) on best effort. Corporate and Internet traffic can be offered with lower priority on MC 5 (Interactive class, THP=ARP =2). WAP and MMS, mostly placed behind the same access point name (APN), can be carried on MC 6 (Interactive class, THP=ARP=3). Ultimately, Dialup connections, in this case with the lowest priority, can be mapped onto MC 7 (Background class, ARP=3).

This means that counters should be classified by the NEs with a granularity that allows the NMS to make statistics based on those. Figure 6.2 shows an example of how different counters collected by SGSN and RNC can be classified to support the PDP context (management class) based monitoring of throughput by the management layer. The classification of measurements by RNC and SGSN includes attributes as TC and THP (in the case of Interactive class) that enable NMS to filter out the indicators and compute throughput per bearer service as described in the following section, where some metric for the differentiated QoS analysis are presented. From these measured metrics, taking into account the one to one mapping of applications onto distinct QoS profiles (management classes), it is possible for the operator to draw conclusions upon the performance of each provisioned service.

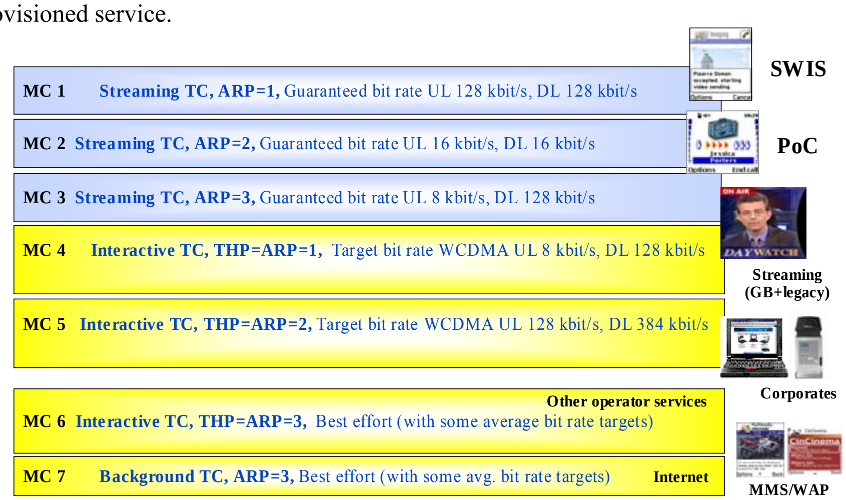


Figure 6.1. Example of management class (MC) definition and mapping of services onto MCs.

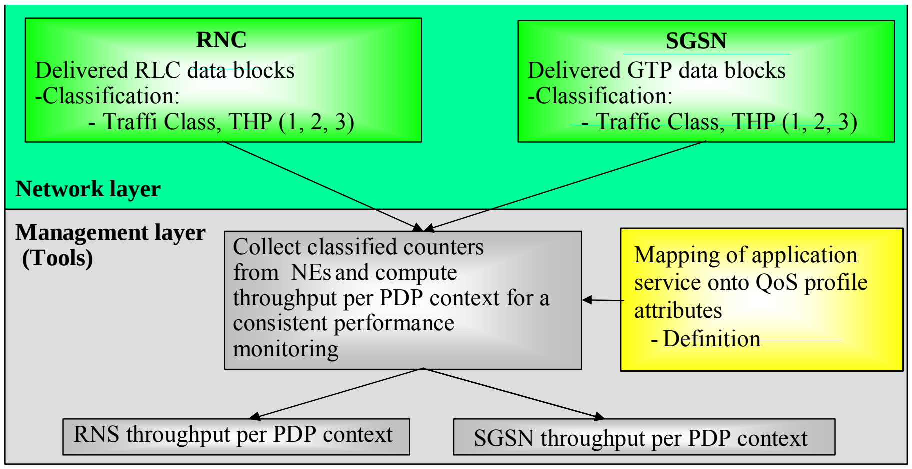


Figure 6.2. Example of how classified counters at the different NEs can be retrieved by the management layer to compute, consistently, throughput for a particular bearer service.

However, if more applications (as for example WAP and MMS in Figure 6.1) are carried by the same PDP context, or placed behind the same Access Point Name (APN) for legacy terminal, i.e. mapped onto the same sub set of QoS attributes, upon which the metrics are classified, the QoS monitoring of the services using this method is not possible, because only the overall performance of the traffic aggregate would be assessed, and the integrity, retainability and accessibility of the different services can no longer be distinguished from each other.

6.3 Differentiated Performance Monitoring

This section presents theoretically how the QoS of differentiated packet switched (PS) and circuit switched (CS) services can be assessed through counters collected and classified in the RNS. The analytical approach assumes that the network topology where the service performances are analysed is already defined (or selected within a wider *scope*) together with the measurement period (*history*), and user *satisfaction criteria*. The identified area may encompass radio network controllers (RNC), base stations (BSs) or Node Bs, cells and the interface between the base station and the radio network controller (I_{ub}). Our target is to define essential counters and key performance indicators (KPIs) that need to be retrieved, and/or derived from measurements in NEs, for a capacity and QoS status view in the RNS and/or a detailed performance analysis. In the latter case, for example, performance results may be compared directly to the target values or, since only the QoS perceived by end-user matter, expressed in terms of satisfied users. The network administrator may then compare the number of satisfied users to the related target thresholds defined a priori.

6.3.1 Classification of Counters

Cumulative counters or gauges (when data being measured can vary up or down during the measurement period) presented in the following sections are collected in the RNC. For a differentiated performance monitoring, the measurements, wherever possible, should be collected in UL and DL directions and classified based on the following attributes (for more information see Chapter 2):

- CS Conversational
  - Call type: Speech or Transparent (T) data
  - Guaranteed bit rate
  - Transport channel type: e.g. DCH
- CS Streaming
  - Non Transparent (NT) data
  - Guaranteed bit rate
  - Transport channel type: e.g. DCH
- PS Conversational
  - Guaranteed bit rate
  - Allocation Retention Priority (ARP)
  - Transport channel type: e.g. DCH
- PS Streaming
  - Guaranteed bit rate
  - Allocation Retention Priority (ARP)
  - Transport channel type: e.g. HS-DSCH, DCH
- PS Interactive
  - Maximum bit rate
  - Traffic Handling Priority (THP)
  - Allocation Retention Priority (ARP)
  - Transport channel type: e.g. HS-DSCH, DCH, FACH, RACH
- PS Background
  - Maximum bit rate
  - Allocation Retention Priority (ARP)
  - Transport channel type: e.g. HS-DSCH, DCH, FACH, RACH
- Cell, RNC, URA, RA and LA identifiers

This make it possible for the operator to filter out measurements for consistent service performance monitoring across network domains, as pointed out in Section 6.2, or for gathering performance metrics according to specific needs. In practice, only a subset of the above attributes may be required. Such a combination of attributes is denoted by m, in the following sections. (Note: Currently there are not any 3GPP specifications, including [1], in which may be found restrictions to possible combinations between TC and CN domain. Consequently, the proposed approach leaves open every possible combination between TC and CN domain as [1] does.)

6.3.2 QoS Integrity Monitoring

This section presents how some useful performance metrics for QoS integrity monitoring such as throughput, delay (and delay variation or jitter) and data loss of bearer services can be derived.

6.3.2.1 Uplink E_b/N_0, BLER and BER Derivation

This section enables the determination of BLER (Block Error Rate), BER (Bit Error Rate), and E_b/N_0 associated to a selected TrCH multiplexed with more TrCHs to a DPCH based on the available target Signal to Interference Ratio (SIR) in the RNC. The mathematical derivation of the following formulas is described in Appendix D.

The desired E_b/N_0 for each transport channel can be put into relation to the determined SIR according to the following equation:

$$\frac{E_{b,DCH}}{N_0} = SIR_{DPCCH} - 20\log(r) - 20\log\left(\frac{R_{DCH}}{R_{DPDCH}(N + r^2)}\right) - 10\log\left(\frac{\sum_{DCH\in RL} RM_{DCH} N_{DCH}^C}{RM_{DCH} N_{DCH}^C}\right) - 10\log\left(\frac{SF_{DPCCH}}{SF_{DPDCH}}\right), \tag{6.1}$$

where SIR_{DPCCH} is the SIR target per symbol on the DPCCH estimated in the RNC or the actual SIR_{DPCCH} value measured at Node B [88], r is the uplink DPCCH overhead in terms of power, it may be expressed as a function of the 3GPP specified *gain factors* (amplitude offsets) β_c and β_d [89], that is:

$$\frac{RSCP_{DPCCH}}{RSCP_{DPDCH}} = \left(\frac{\beta_{c,TFC_{Max}}}{\beta_{d,TFC_{Max}}}\right)^2 = r^2, \tag{6.2}$$

R_{DCH} is the bit rate prior to CRC attachment, R_{DPDCH} is the DPDCH symbol rate, N is the number of codes employed in the uplink transmission, RM_{DCH} is the rate matching attribute for a specific dedicated channel belonging to the radio link (RL) in question [71], N_{DCH}^C is the number of bits per radio frame subsequent to radio frame equalisation, i.e. prior to rate matching, and SF is the spreading factor. SF_{DPCCH} is always 256 in the uplink direction.

Equation (6.1) defines the calculation of the E_b/N_0 requirement for a specific transport channel. The composite E_b/N_0 figure for all transport channels belonging to a physical channel (denoted as Code Composite Transport Channel [71], CCTrCH) can be obtained by summing the individual E_b/N_0 values.

In the case of HSDPA, (6.1) needs to take into account the overhead introduced by the UL HS-DPCCH. Let Δ be the power offset between the uplink associated DPCCH and the HS-DPCCH [89], the energy per bit for a specific dedicated channel can be expressed as:

$$E_{b,HSDPA} = E_{b,DCH}\left(1 + \frac{r^2}{(N + r^2)}\Delta\right). \tag{6.3}$$

Hence, (6.1) for HSDPA becomes:

$$\frac{E_{b,HSDPA}}{N_0} = \frac{E_{b,DCH}}{N_0} + 10\log\left(1 + \frac{r^2}{(N + r^2)}\Delta\right). \tag{6.4}$$

Using (6.1) or (6.4), for HSDPA, the RNC is thus able to determine an average E_b/N_0 for this TrCH as follows:

$$\overline{\frac{E_b}{N_0}} = \frac{\sum_{SW_1}\sum_{RP} 10^{\frac{E_b/N_0}{10}}}{RP \cdot SW_1}, \tag{6.5}$$

where SW_1 denotes the sliding window size for the average computation and RP the reporting period of the counter in question, which should be classified based on the bearer attributes of the transported user data (see Section 6.3.1).

Further a BLER can be determined for the transport channel by:

$$\overline{BLER_{DCH}} = \frac{\sum_{SW_2} \frac{\sum_{RP} CRC\_NOK}{\sum_{RP}(CRC\_OK + CRC\_NOK)}}{SW_2}, \tag{6.6}$$

where SW_2 denotes the BLER sliding window size, and *CRC_NOK*, *CRC_OK* is the result of the CRC check the RNC receives from the BS together with the received transport block. This counter should be also classified based on the bearer attributes of the transported user data (see Section 6.3.1). In addition or alternatively, also a BER can be determined for the monitored DCH by:

$$BER_{DCH} = \frac{\sum_{SW_3} \frac{\sum_{RP} QE}{RP}}{SW_3}, \tag{6.7}$$

where SW_3 denotes the BER sliding window size and QE is the estimation of the average BER of the DPDCH data of a Radio Link Set at Node B [89], which is reported to the RNC after the end of each TTI of the TrCH. The BER computation is only possible, however, when turbo coding is used. This counter should be also classified as stated in Section 6.3.1. Furthermore, the sliding window content for quality computations needs to be reset when the target SIR is changed and thus sent to the WCDMA Node B.

The connection-based counters should be updated cell by cell, since the RNC is aware of the cell participating in diversity handover (DHO), and based on the bearer attributes of the transported user data (see Section 6.3.1). If the measurements are needed for an online and/or trace of a specific radio connection, the counters should be delivered together with the actual connection frame number (CFN) every counters update period.

6.3.2.2 Dowling BLER Computation

The downlink transport channel block error rate (BLER) is based on evaluating the CRC of each transport block associated with the measured transport channel after RL combination. The BLER is computed over the measurement period as the ratio between the numbers of received transport blocks resulting in a CRC error and the number of received transport blocks. The mobile when explicitly ordered by the network may report such a measurement periodically on event based [69]. In the case of AM RLC, the BLER can be estimated in the RNC, reducing the signalling load in the cell, as follows:

$$BLER^m = \frac{\sum_{i=1}^{N} \overline{B}_i^m}{\sum_{i=1}^{N}\left(B_i^m + \overline{B}_i^m\right)}, \tag{6.8}$$

where $\overline{B}_i^m$ and B_i^m denotes, respectively, the number of RLC blocks unsuccessfully and successfully transmitted during the sampling period *s*, and *N* is the total number of samples collected by the RNC for the management class *m* during the measurement period *S*.

6.3.2.3 Throughput Computation

The throughput relates only to the correctly received bits during a predefined measurement period (observation time), denoted by *S* in the following, where RLC buffers are not empty. In the literature the metric is presented as Active Session Throughput or Circuit Switched Equivalent Bit-Rate, which is an essential indicator to assess the QoE and how spectral efficient the provisioned QoS is [74]. Throughput measurements in the UL may be made as described in the next paragraph for the DL direction considering the correctly received RLC blocks instead of the transmitted ones by the corresponding protocol entities. (In the RNC, accurate measurements are only possible in the downlink direction, since the exact status of the uplink buffers is known only at UE.)

Let $B_i^{m,k}$ be the number of delivered RLC blocks (PDUs without retransmissions, in the case of AM RLC) by the RLC entity *i*, where *m* denotes management class, and *k* the related PDU size, and d_i be the corresponding transmission duration during the sampling period *s*, in number of transmission time intervals (TTIs). The throughput for a particular management class *m* in a cell of the WCDMA network (SAP 6 in Figure 2.20), can be calculated as:

$$t^m = \frac{\sum_{i=1}^{N}\sum_{k=1}^{C} r_k B_i^{m,k}}{\sum_{i=1}^{N} d_i^m}, \tag{6.9}$$

where the RLC PDU size in bits without header r_k depends on the amount of data that can be transferred to MAC during each TTI, *C* denotes all possible combinations of dynamic parts of the set of Transport Formats (TFs) associated to the transport channel [70], and *N* is the total number of samples collected by the RNC for that particular management class, during the measurement period *S*. (Note: For Conversational and Streaming traffic classes there are no needs to measure the bit rate per connection, since services carried on those QoS classes are offered with guaranteed bit rate. However, (6.9) may be used to compute the throughput per cell as follows.)

Ultimately, for each management class *m* the mean cell throughput T^m can be derived by summing up all correctly delivered bits in the cell over the monitored period *S*, and dividing the attained value by *S*, i.e.

$$T^m = \frac{\sum_{i=1}^{N}\sum_{k=1}^{C} r_k B_i^{m,k}}{S}. \tag{6.10}$$

The total cell throughput can be obtained from (6.10) summing up all contributions.

Despite the high computational time, the advantage of measuring at RLC, instead of e.g. at MAC layer, stands on the fact that the obtained measurements are independent from the transport channel employed in the user data transmission. Statistics based on transport channels may be derived filtering the collected data based on *m*, as explained in Section 6.3.1.

6.3.2.4 RLC Retransmission Rate

This indicator relates to number of retransmissions required to transmit an RLC PDU when the first transmission was not successful through the radio interface. The metric can be used to set the maximum number of allowed link layer retransmissions without compromising the load of the cell for the global quality of service requirements [17].

Let F_i^m be the number of RLC blocks (PDUs) to be transmitted (or received, in the UL case) for the first time after segmentation by the AM RLC entity *i*, where *m* denotes the management class, and let P_i^m be the total number of transmitted (or received, in the UL) blocks (including retransmissions) during the sampling period *s*. The RLC retransmission rate *U* for *m*, in a cell of the WCDMA network (SAP 6 in Figure 2.20), can be calculated as:

$$U^m = 1 - \frac{\sum_{i=1}^{N} F_i^m}{\sum_{i=1}^{N} P_i^m}, \tag{6.11}$$

where *N* is the total number of samples collected by the RNC for the management class *m* during the measurement period *S*.

6.3.2.5 Service Data Unit Error Ratio

The SDU error ratio is defined as the fraction of SDUs lost or detected as erroneous [1]. This metric is an essential indicator for assessing the performance of protocols, algorithm configurations and error detection schemes.

Let Y_i^m be the total number of SDUs obtained from upper layer (PDCP or RRC) by the AM RLC entity *i* for transmission, where *m* denotes the management class, and X_i^m be the number of SDUs detected as erroneous or discarded (e.g. due to error detection, too many retransmissions or due to the fact that the timer discard expired [66]), during the sampling period *s*. The SDU error ratio for *m*, in a cell of the WCDMA network (SAP 6 in Figure 2.20), can be calculated as:

$$Q^m = \frac{\sum_{i=1}^{N} X_i^m}{\sum_{i=1}^{N} Y_i^m}, \tag{6.12}$$

where N is the total number of samples collected by the RNC for the measurement m during the measurement period S.

6.3.2.6 Downlink Transfer Delay Computation

The transfer delay is defined as maximum delay for 95th percentile of the distribution of delay for all delivered SDUs during the lifetime of a bearer service, where delay for an SDU is defined as the time from a request to transfer an SDU at one SAP to its delivery at the other SAP [1]. In practice, in these terms, the statistical measurement of the transfer delay would require to measure the delay of all delivered SDUs during the lifetime of one bearer service and save the corresponding values separately so that the distribution of delay could be derived. The transfer delay would then be the delay that is greater than or equal to the delays of 95% of the delivered SDUs during the lifetime of the bearer service. Hence, the transfer delay measurement may become an issue when a statistical analysis is required. Assuming the above definition, a sensible statistic would require the collection of data on all bearer services being active during the measurement interval, which may last several hours for an accurate and precise convey of the information. Therefore, in practice, such a solution is not adequate due to the amount of memory and processing power consumption it would require in the network elements involved in the computation. This section proposes a simple technique by means of statistics, such as mean and standard deviation of the SDU delay, and introduces a confidence interval assuming the distribution of delay is normal. The method is valid for AM RLC in the downlink direction.

Let D_i be the time interval from a request to transfer an SDU at SAP 6 to its delivery at the SAP 4 for the bearer service #i (see Figure 2.20). This may be computed as the difference between the times when the SDU is received from upper layer and the last PDU containing data from that SDU is correctly received with acknowledgement by the UE. (Note: If the transmission time is included together with the packet waiting time, since packets may vary in size, the transfer delay should be always related to the maximum SDU size, which is derived from the RAB attributes.) The expectation of the delay can be estimated as:

$$\mathrm{E}(D_i) = \frac{Di_1 + \ldots + Di_{M_i}}{M_i}, \tag{6.13}$$

where M_i is the number of transferred SDUs for that particular bearer service. Furthermore, the standard deviation of the delay when M_i is sufficiently large can be estimated as:

$$\sigma(D_i) = \sqrt{\mathrm{E}(D_i^{\,2}) - \mathrm{E}(D_i)^2}, \tag{6.14}$$

where:

$$\mathrm{E}(D_i^{\,2}) = \frac{\sum_{j=1}^{M_i} (Di_j)^2}{M_i}. \tag{6.15}$$

A confidence interval is an interval in which a measurement or trial falls corresponding to a given probability [91]. Usually, the confidence interval of interest is symmetrically placed around the mean, so a 50% confidence interval for a symmetric probability function would be the interval [-*a, a*] such that

$$\frac{1}{2} = \int_{-a}^{a} p(x)dx\,. \tag{6.16}$$

For a normal distribution, the probability that *a* measurement falls within *n* standard deviations ($n\sigma$) of the mean μ (i.e., within the interval [μ-$n\sigma$, μ+$n\sigma$]) is given by

$$P(\mu - n\sigma < x < \mu - n\sigma) = \mathrm{erf}\left(\frac{n}{\sqrt{2}}\right), \tag{6.17}$$

where erf(x) is the so-called error function.

Conversely, to find the probability-P confidence interval centred about the mean for a normal distribution in units of σ, solve equation (6.17) for *n* to obtain

$$n = \mathrm{erf}^{-1}(P), \tag{6.18}$$

where $\mathrm{erf}^{-1}(x)$ is the inverse error function.

Hence, assuming that the distribution of delay is normal, the maximum delay for X^{th} percentile TD_i of the distribution of delay for all delivered SDUs during the lifetime of the bearer service is given by:

$$TD_i = \mathrm{E}(D_i) + n \cdot \sigma(D_i). \tag{6.19}$$

Under this assumption, when setting the parameter $n = 1.95$, (6.19) yields the transfer delay of bearer service #*i* with respect to the 3GPP definition, i.e. the maximum delay for 95$^{\mathrm{th}}$ percentile of the distribution of delay for all delivered SDUs during the lifetime of bearer service #*i*.

The statistical value of the transfer delay *TD* for a particular management class *m* is the average of the transfer delay values upon the N_B different bearer services of class *m* during the measurement interval *S*, i.e.

$$\overline{TD}^m = \frac{TD_1^m + \ldots + TD_{N_B^m}^m}{N_B^N}. \tag{6.20}$$

By inserting (6.14) and (6.15) in (6.19), (6.20) can be rearranged as:

$$\overline{TD}^m = \frac{\mathrm{E}(D_1^m) + \ldots + \mathrm{E}(D_{N_B^m}^m)}{N_B^m} + n \cdot \frac{\sigma(D_1^m) + \ldots + \sigma(D_{N_B^m}^m)}{N_B^m}. \tag{6.21}$$

Therefore, the statistical value of the transfer delay of a particular management class *m* can be expressed as the average of the mean values of the SDU delays plus *n* times the average value of the corresponding standard deviations, where *n* relates to the level of confidence for the operator to choose.

In practice, by adopting this method, only three different counters need to be collected by each of the RLC entities in the RNC, namely:

- The number of transferred SDUs, i.e. *M*;

- The occurred SDU delays, i.e. M•E(D) in (6.13);
- The squared of the occurred SDU delays, i.e. M•E(D^2) in (6.15).

6.3.2.7 Downlink Jitter Computation

The jitter of a specific bearer service is defined as the difference between the one-way delays of the selected packet pair, e.g. consecutive packets. This section proposes a method for assessing the delay variation as the difference between maximum and minimum one-way delay for a predefined percentile of the distribution of delay for all delivered SDUs during the lifetime of the bearer in question. Assuming that the distribution of delay is normal, this is exactly the confidence interval corresponding to a given probability. Hence, for the bearer service #i the delay variation for X^{th} percentile of the distribution of delay for all delivered SDUs during the lifetime of bearer service #i is:

$$J_i = 2 \cdot n \cdot \sigma(D_i), \tag{6.22}$$

where n relates to the level of confidence for the operator to choose.

The statistical value of the delay variation J for a particular management class m is the average of the jitter values upon all bearer services N_B of type m during the measurement interval S, i.e.

$$\overline{J}^m = \frac{J_1^m + \ldots + J_{N_B^m}^m}{N_B^m}. \tag{6.23}$$

By inserting (6.14) in (6.22), (6.23) can be rearranged as:

$$\overline{J}^m = 2 \cdot n \cdot \frac{\sigma(D_1^m) + \ldots + \sigma(D_{N_B^m}^m)}{N_B^m}. \tag{6.24}$$

Therefore, the statistical value of the delay variation of a particular management class m can be expressed as $2n$ times the average value of the standard deviation of the SDU delay, where n relates to the level of confidence for the operator to choose.

In practice, by adopting this method, only three different counters need to be collected by each of the RLC entities in the RNC, as explained in Section 6.3.2.6.

6.3.3 QoS Accessibility and Retainability Monitoring

Accessibility and retainability measurements are based on the success/failure of procedures needed to setup, modify or maintaining a certain bearer service or signalling connection. Hence, the proposed measurements are attached either to the successful or the unsuccessful issue of a procedure for RAB or signalling connection management. The procedures of interest are depicted in Figure 6.3 and in Figure 6.4, other relevant procedures, such as radio link and handover management, for a more detailed performance analysis, may be found in [30].

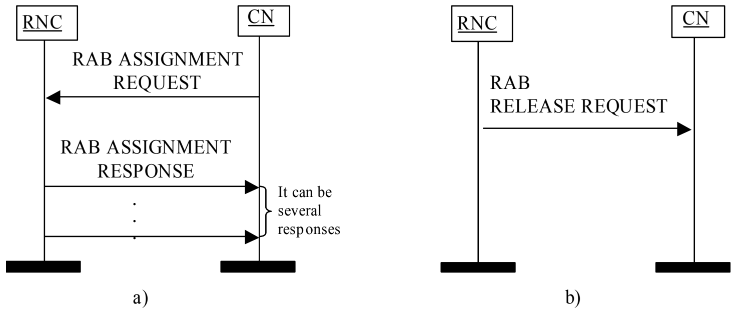


Figure 6.3. a) RAB assignment procedure; b) RAB release request procedure.

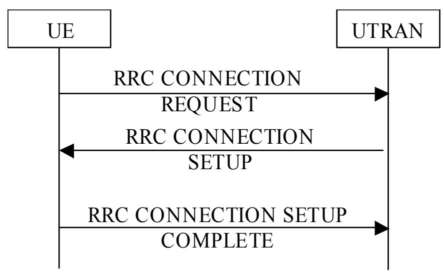


Figure 6.4. RRC connection set-up procedure.

6.3.3.1 RAB Management

Five measurement types may be defined for CS and PS domains. The measurements need to be split into sub-counters according to the classification of Section 6.3.1. The subset of relevant attributes is denoted by *m* in the extension of the following performance indicators. The possible failure causes are specified in [92].

- **Number of RAB assignment attempts**: On receipt by the RNC of a RANAP RAB ASSIGNMENT REQUEST message for CN, each RAB assignment request is added to *RAB.AttEstab.m* counter.
- **Number of successfully established RABs**: On transmission by the RNC of a RANAP RAB ASSIGNMENT RESPONSE message to the CN, each successfully established RAB is added to *RAB.SuccEstab.m* counter.
- **Number of RAB establishment failures:** On transmission by the RNC of a RANAP RAB ASSIGNMENT RESPONSE message to the CN, each RAB failed to establish is added to *RAB.FailEstab.Cause.m* counter according to the failure cause.
- **RAB connection set-up time (mean)**: This measurement is obtained by accumulating the time intervals *RAB.SuccEstabSetupTimeMean.m* for each successful RAB establishment, which is then divided by the number of successfully established RABs observed in the granularity period to give the arithmetic mean.

- **RAB connection set-up time (maximum)**: This measurement may be obtained by the high tide mark *RAB.SuccEstabSetupTimeMax.m* of the monitored time intervals for each successful RAB establishment.
- **Number of RAB releases**: On transmission by the RNC of a RANAP RAB RELEASE REQUEST message, each RAB requested to be released is added to the relevant per cause measurement *RAB.Rel.Cause.m*.

From the above measurements, for instance, the following key performance indicators may be derived. The bearer service accessibility relates to the percentage of the attempts that have been followed by an assignment of a bearer service:

$$RAB.SetupSuccRatio.m = \frac{\sum RAB.SuccEstab.m}{\sum RAB.AttEstab.m}. \tag{6.25}$$

While the reliability of the bearer service relates to the number of RABs that have been terminated by the UE (Cause: UE generated signalling connection release) and the number of RABs that have been started correctly:

$$RAB.RelSuccRatio.m = \frac{\sum RAB.Rel.UE.m}{\sum RAB.SuccEstab.m}. \tag{6.26}$$

Both the accessibility and reliability of the bearer service can be assessed considering the product of the previous performance indicators, which reduces to the *RAB success ratio*:

$$RAB.SuccRatio.m = \frac{\sum RAB.Rel.UE.m}{\sum RAB.AttEstab.m}. \tag{6.27}$$

6.3.3.2 Signalling Connection Management

In order to assess the establishment and release of a signalling connection the following measurements can be made. The possible causes are defined in [92].

- **Attempted signalling connection establishments**: This measurement provides the number of attempts *SIG.AttConnEstab.m* by RNC to establish an I_u control plane connection with the CN. In this case, *m* may simply denote the PS or CS domain. The trigger point is the transmission of a RANAP Initial UE message by the RNC to the CN, which is sent by the RNC on receipt of an RRC Initial Direct Transfer message from the UE.
- **Attempted RRC connection establishments**: This measurement provides the number of RRC connection establishment attempts for each establishment cause. On receipt of an RRC Connection Request message by the RNC from the UE, each received RRC Connection Request message is added to the relevant per cause measurement *RRC.AttConnEstab.Cause.*
- **Failed RRC connection establishments**: This measurement provides the number of RRC establishment failures for each rejection cause. On transmission of an RRC Connection Reject message by the RNC to the UE, or an expected RRC CONNECTION SETUP COMPLETE message not received by the RNC, each RRC Connection Reject message received is added to the relevant per cause measurement *RRC.FailConnEstab.Cause.*

- **Successful RRC connection establishments**: This measurement provides the number of successful RRC establishments for each establishment cause. On receipt by the RNC of a RRC CONNECTION SETUP COMPLETE message following a RRC establishment attempt, each RRC Connection Setup Complete message received is added to the relevant per cause measurement *RRC.SuccConnEstab.Cause*.
- **RRC connection set-up time (mean)**: This measurement is obtained by accumulating the time intervals for every successful RRC connection establishment per establishment cause between the receipt by the RNC from the UE of a RRC CONNECTION REQUEST and the corresponding RRC CONNECTION SETUP COMPLETE message over a granularity period. The end value of this time, denoted as *RRC.AttConnEstabTimeMean.Cause*, is then divided by the number of successful RRC connections observed in the granularity period to give the arithmetic mean. The measurement is split into sub counters per establishment cause.
- **RRC connection set-up time (max)**: This measurement is obtained by monitoring the time intervals for each successful RRC connection establishment per establishment cause between the receipt by the RNC from the UE of a RRC CONNECTION REQUEST and the corresponding RRC CONNECTION SETUP COMPLETE message. The high tide mark of this time, *RRC.AttConnEstabTimeMax.Cause*, is the collected value.
- **Attempted RRC connection releases**: This measurement provides the number of RRC connection release attempts per release cause sent from UTRAN to the UE. On transmission of an RRC CONNECTION RELEASE message by the RNC to the UE, each RRC Connection Release message sent is added to the relevant per cause measurement *RRC.AttConnRel.Cause*.

From the above counters, for example, the following key performance indicator may be derived for each cause of the RRC connection establishment:

$$RRC.SetupAccessCompleteRatio.Cause = \frac{\sum RRC.SuccConnEstab.Cause}{\sum RRC.AttConnEstab.Cause}\,. \qquad (6.28)$$

6.4 Conclusions

Measuring performance by means of management classes is a promising technique for assessing performance of services through WCDMA networks, without tracing upper layer protocols at the different interfaces or dumping data in the mobile terminals. The proposed performance metrics, namely: E_b/N_0, BLER, BER, throughput, RLC retransmissions, SDU error ratio, transfer delay and jitter, may be used for a proactive and timely supervision of service level agreements and quality in the radio network subsystem. Yet, the application of the proposed method is not limited to RNS, but it can be deployed in any radio access and packet core networks supporting PDP (packet data protocol) contexts.

7

QoS "Optimisation"

This chapter presents a genetic approach to "optimal" QoS provisioning for WCDMA networks. Using a genetic algorithm, several parameters are changed simultaneously to maximise the percentage of satisfied users of the offered services upon different input traffic mixes. The gains provided by the "optimal" parameter values with respect to default network configurations are expressed in terms of spectral efficiency. Simulation results show clear values for the network administrator to adopt different QoS provisioning for distinct input traffic mixes and the proposed genetic algorithm to be a feasible *research* method for multi-parameter optimisation. The research methodology applies to estimate good parameter settings for real networks and/or to study whether there is any need for the network administrator to change them, e.g. by other means, when the offered traffic varies.

7.1 Introduction

The QoS optimisation in mobile networks is a process that aims at maximising the spectral efficiency of the system at a given QoE, i.e. percentage of satisfied users of the offered services, by choosing the best values for a selected set of crucial parameters. Traditionally, technical experts, who have access to live networks, perform the network optimisation. In order to automise the "optimisation" process and thus reduce the operational expenditures (OPEX), several methods are available in the literature to search for the solutions within a part of the possible combinations of network configurations; some of those were presented in [41]-[50]. Yet, some useful methodologies of QoS "optimisation" were never proposed.

This chapter presents a genetic approach to QoS "optimisation" for WCDMA networks that performs well in terms of accuracy and speed of convergence. A *genetic algorithm* is an optimisation algorithm that makes use of a population of artificially reproducing test solutions and works by enforcing the condition that a fitness function for this population is equal to an objective function. The *fitness function* is a way to describe the dynamics of gene or genotype frequencies in populations of reproducing individuals. The fitness function measures the (potential for) reproductive success of any individual in a given environment. The *objective function* is input to the genetic algorithm, while process of forcing the fitness function to be equal to the objective function is a part of the algorithm itself. (Because of the widespread popularity of genetic algorithms, the term fitness function has taken on the same meaning as the term objective function.) In this work, this solution space search algorithm is applied to find the “optimal” values of a number of radio resource management (RRM) parameters that have influence on the performance of bearer services with different QoS requirements. The parameter settings are tuned to improve the service performance of five different traffic mixes. From the optimised parameters, two default settings are derived (differentiated and undifferentiated parameters), which may be used in all deployment scenarios.

Our ultimate aim is to quantify the spectral efficiency gains provided by the service differentiation with “optimal” parameter values with respect to default configurations at a given QoE and input traffic mix [60].

The extensive simulations were run implementing a feedback loop in the tool presented in Section 3.5.1. More insights into test parameters, radio resource management functions, and traffic models were given in Chapter 3 and 4.

The remainder of this chapter is organised as follows: Section 7.2 describes the test parameters and constraints imposed in the corresponding ranges. Section 7.3 presents the genetic approach used in this work to search the “optimal” combination of parameter values within the range of a selected set of the crucial parameters described in Section 7.2. Section 7.4 reports the specific models and case studies for “optimal” QoS provisioning and spectral efficiency gains computation. The simulation results are illustrated and discussed in Section 7.5. Conclusions are drawn in Section 7.6.

7.2 Test Parameters

In this section, we present the radio resource management parameters to be optimised for non-guaranteed bit rate bearer services, i.e.: *Minimum and maximum allowed bit rates*, and *inactivity timer*, which are differentiated based on Interactive (traffic handling priority THP = 1, 2, and 3) and Background classes, and the undifferentiated *DCH minimum allocation time* and *maximum capacity request queuing time* in the packet scheduler queue. Such parameters were found in [61] to be a crucial set to provide discriminated treatment among the services carried with distinct QoS profiles.

For the sake of limiting the solution space search size of our optimisation problem, the ranges of the differentiated parameter values in this study were customised based on the requirements (source bit rate and user satisfaction criterion) of the carried application with the constraint that the maximum allowed bit rate would be always greater or at least equal to the corresponding minimum allowed one (see Section 7.4). The ranges of the test parameters and their mapping onto service applications are listed in Table 7.1.

(Note: In this study, the parameters have the same value for all cells of the analysed network, where statistics are collected. This reduces the computation time and complexity of evaluating a solution and provides a method to optimise one cell or a cluster of cells at a time.)

7.2.1 Minimum and Maximum Allowed Bit Rates

When a bearer service of the Interactive or Background class is admitted, Admission Control in UTRAN produces the transport format set (TFS) for the transport channel (TrCH) used to transfer the user data from the layer 2 to layer 1 (protocol entities located in the base station or Node B). As illustrated in Figure 7.1, this is done by initially comparing the available bit rates (vendor specific parameters) for the TrCH in question and the maximum bit rate specified in the QoS profile (see Table 2.3) the RNC receives when the radio bearer is set up or modified. The maximum bit rate of the TFS is then calculated as the minimum between the one in the QoS profile and the *maximum allowed bit rate*, which is a radio network planning parameter. The minimum bit rate of the TFS is zero, to allow DTX transmission, followed by the *minimum allowed bit rate*, which is also a parameter for the operator to set. Hence, the bit rate allocated to that particular TrCH is then within the constructed TFS. In the example of Figure 7.1, the available bit rates for the TFS construction are: 0, 8, 16, 32, 64, 128, 144, 256 and 384 kb/s. The peak bit rate from the RAB attributes is supposed to be 256 kb/s. Now, considering the fact that the minimum and maximum allowed bit rates are set to 32 and 64 kb/s, respectively, the TFS for the bit rate allocation is therefore 0, 32, and 64 kb/s. This means that if the available power is not enough to accommodate the minimum allowed bit rate, 32 kb/s in this case, during that particular scheduling period, the capacity request (CR) of the bearer service in question remains in the PS queue, with no possibility of data transmission. An example of bit rate allocation for up to five capacity requests in PS queue was illustrated in Figure 4.2 and in Figure 4.3. In this study, the minimum and maximum allowed bit rates were differentiated parameters, which means that different values could be set for those parameters depending on a subset of the QoS profile attributes, i.e. Interactive class, THP 1, 2 and 3, and Background class. Each parameter thus allows 4 distinct treatments.

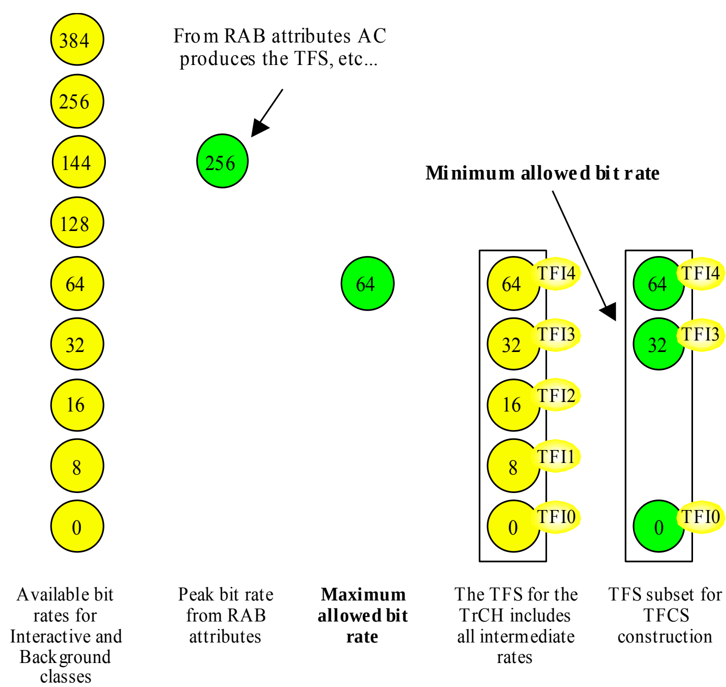


Figure 7.1. Example of Transport format set (TFS) construction in the packet-scheduling algorithm.

7.2.2 Inactivity Timer

During the packet data transmission, bearer services that have been longer in DTX than the corresponding *inactivity timer* are moved to Cell_FACH (forward access channel) state and the corresponding allocated resources are released. In this work, the transmission using FACH was disregarded: When the new data arrives in the buffer of the radio network controller (RNC), a new CR is sent to PS and subsequently another DCH is allocated to the bearer in question based on the FT scheduling algorithm (see Section 4.2.2).

Figure 7.2 shows how the transport channel switching (from Cell_DCH to Cell_FACH) was modelled, where the inactivity timer was also tested as differentiated parameter. This means that the inactivity timer could be set differently depending on the Interactive and Background class attributes of the established PDP context. When the communication is ongoing and there is no data to transfer from the RNC buffer, a timer (denoted by T_{in} in Figure 7.2) is started. When the counter exceeds the corresponding inactivity timer value, the connection is moved to Cell_FACH, radio resources are released and the bit transmission interrupted. Subsequently, when the *reading time* (randomly generated depending on the corresponding traffic model) expires, and new data arrives in the RNC buffer, a new CR is sent to PS. At the same time, the measurements of the active session throughput and UE buffer level for the QoE evaluation are started.

7.2.3 DCH Granted Minimum Allocation Time

Allocated bit rates may be rescheduled if the ongoing communication has lasted longer than the corresponding *(granted) minimum allocation time* (see Figure 4.1, where an example of bit rate modification is depicted). In this study, such a parameter was not differentiated for the different bearer services.

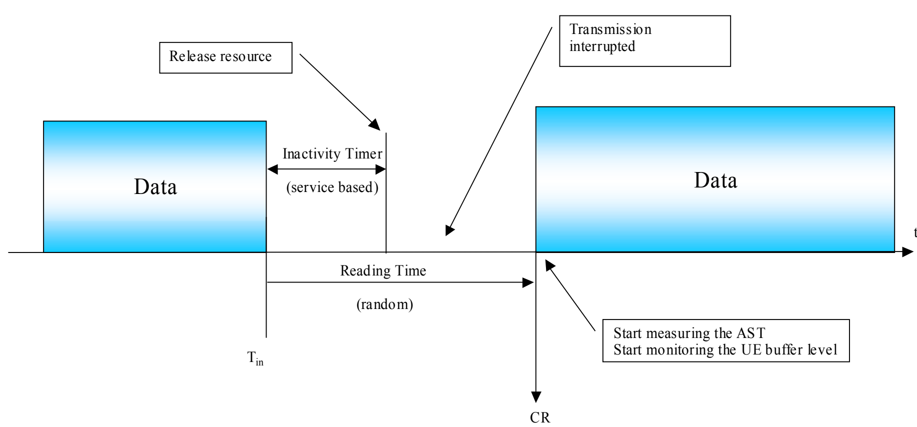


Figure 7.2. Example of TrCH switching from DCH to FACH.

Table 7.1. Test parameters and corresponding ranges.

Parameter	Management Class	Range	Service
Minimum allowed bit rate	MC 1	8, 16 kb/s	PoC
	MC 2	64, 128 kb/s	Streaming
	MC 3	32, 64, 128, 144, 256, 384 kb/s	WAP/MMS
	MC 4	64, 128, 144, 256, 384 kb/s	Dialup
Maximum allowed bit rate	MC 1	8, 16 kb/s	PoC
	MC 2	64, 128 kb/s	Streaming
	MC 3	32, 64, 128, 144, 256, 384 kb/s	WAP/MMS
	MC 4	64, 128, 144, 256, 384 kb/s	Dialup
Inactivity timer	MC 1	1, 2, 5, 10, 20, 30 s	PoC
	MC 2	1, 2, 5, 10, 20, 30 s	Streaming
	MC 3	1, 2, 5, 10, 20, 30 s	WAP/MMS
	MC 4	1, 2, 5, 10, 20, 30 s	Dialup
DCH minimum allocation time	Not differentiated	1, 2, 5, 10, 15, 20 s	All services
CR Maximum queuing time	Not differentiated	1, 2, 5, 10, 15, 20 s	All services

7.2.4 Capacity Request Maximum Queuing Time

Capacity requests are rejected if they stay in the PS queue longer than the value specified by *CR maximum queuing time*. In this study, this parameter was not differentiated, since in [61] the benefit of its differentiation remained questionable.

7.3 Genetic Approach to QoS "Optimisation"

The problem of finding the parameter settings that maximise the percentage of satisfied users at a given traffic mix and volume has not an analytical solution. The advantage of a genetic algorithm stands on the fact that it processes the potential solutions directly, rather than manipulating mathematical formulations of a cost function (object functions). Another important property of the genetic method is that it maintains a set of potential solutions; whereas other algorithms process a single point of the space search and thus only one potential solution to the problem is made available. Genetic algorithms do not guarantee the optimal parameter settings, but by randomly choosing sufficiently many candidates the probability of error may be as small as defined by the operator/user. The genetic algorithm is not a special purpose optimisation technique, it is easy to implement and was adopted in this study to investigate its feasibility for our optimisation problem.

7.3.1 Genetic Algorithm

The first step in the execution consists in the creation of an initial population (combination of distinct parameter settings). In our case, the initial population consisted of 10 vectors (reasonable number of individuals) with 14 components (see Table 7.1). The components are selected randomly within the range of the parameters presented in the previous section.

7.3.1.1 Fitness (Objective Function) Computation

From the initial population the fitness (objective function) of each individual is computed. The objective function is defined as the difference between the actual ratio of satisfied users of the worst performing service and the corresponding target performance (90% for all applications). A graphical derivation of this performance metric in the case of three distinct bearer services, at a given traffic volume, is reported in Figure 7.3. In this study, the actual fitness computation is obtained as a standard statistic of testing the similarity of two ratios: The observed percentage of satisfied users of the service *i*, and the corresponding expected ratio target; namely:

$$Fitness = -\min_{\{i=1,\ldots,N\}}\left\{\frac{\dfrac{SatisfiedUsers(i)}{\max\{ServiceType(i),1\}} - TargetSatisfaction(i)}{\sqrt{\dfrac{TargetSatisfaction(i)\cdot(1-TargetSatisfaction(i))}{\max\{ServiceType(i),1\}}}}\right\}, \tag{7.1}$$

where N is the number of offered services, *SaisfiedUsers*(i) is the observed number of satisfied users (calls) of the service type i, *TargetSatisfaction*(i) is the corresponding target performance threshold (90% for all application), and *ServiceType*(i) denotes the total number of users of that particular service type. The fitness function is a statistic that yields positive or negative values if the ratio of satisfied calls is higher or lower than the target set for the ratio. The lower the fitness value the better the network configuration. Fitness values below zero mean that the service performance is above the target, which is when the fitness function is equal to zero. The numerator of (7.1) is the difference between the ratios and the denominator of (7.1) is the expected mean error of the ratio difference, which decreases when the number of calls increases. The mean error is the standard deviation of the sample ratio or tested proportion, which follows the binomial distribution ($\sigma^2 = p(1\text{-}p)/n$, where n is the number of trials and p the probability of a success on each trial). If the number of calls is sufficiently high, the argument of the minimum function in (7.1) is normally distributed. To be reasonably conservative, this requires that np and $n(1\text{-}p)$ be greater than or equal to 5 [91], which in this case implies that:

$$ServiceType(i) \geq \frac{5}{1-TargetSatisfaction(i)} \geq 50, \tag{7.2}$$

with *TargetSatisfaction*(i) = 90%. This makes it possible to evaluate the confidence interval on the simulation results as presented in Appendix C.

(Note: There are no needs to define a fitness cost (measure) for excessive QoE. Using the propose approach, services with excessive fraction of satisfied users give automatically resources to services with lower quality during the optimisation process. In other words, the performance of the service with lowest QoE is always ameliorated at the expense of the quality experience by other applications, within the constraints (limits) imposed by the QoS management functions presented in Section 4.2. In this work only parameters influencing the differentiation of NGB traffic are optimised. The tuning of the dedicated NGB power, $P_{TxNGBcapacity}$ in (4.2), was already studied by the author though dynamic policies adaptation. A part of the performance results were presented in [56]. This parameter was not included in this work, since the soft capacity (or *statistical multiplexing gain* due to the statistical sharing of resources that happens automatically in WCDMA networks, which is defined as the difference between the power requirement of the aggregate and the sum of the per-stream power requirements) of each cell in the system would be reduced.)

7.3.1.2 Offspring Generation

Each individual is then linked to a weight based on the fitness value, the lower the fitness value the higher the weight. The probability of an individual to be selected for generating an offspring (the set of parameters to be tested) is defined as the weight of the individual in question divided by the sum of all weights associated to all individuals. Table 7.2 summarises the process. The weight is a number from 1 to 10, 10 being the size of the population. The probability is the weight itself divided by the sum of these.

Then two individuals are randomly selected based on the associated probability. Let **Ind_i = [$p_{1,i}$ $p_{2,i}$ ... $p_{14,i}$]** and **Ind_j = [$p_{1,j}$ $p_{2,j}$ ... $p_{14,j}$]** be the two selected individuals (parents). The components of the offspring are randomly chosen between the corresponding components of the two parents. Hence, the offspring can be express as:

$$\mathbf{Ind\_k = [p_{1,k}\ p_{2,k} \ldots\ p_{14,k}] = [p_{1,rand(i,j)}\ p_{2,rand(i,j)} \ldots\ p_{14,rand(i,j)}]}.$$

To compensate for the limited number of individuals (population size), the offspring is subsequently mutated. The mutation (changes in the components of the offspring) is made randomly with a probability of 10% for each parameter to be changed within its range. During the testing phase of the genetic algorithm, 10% resulted a good trade-off between an effective representation of individuals and speed of convergence.

Table 7.2. Example of probability computation for each individual of the population.

Individuals	Fitness (%)	Weight	Probability
Ind 1	-5	10	10/55
Ind 2	-2	9	9/55
…	…	…	…
Ind 10	5	1	1/55

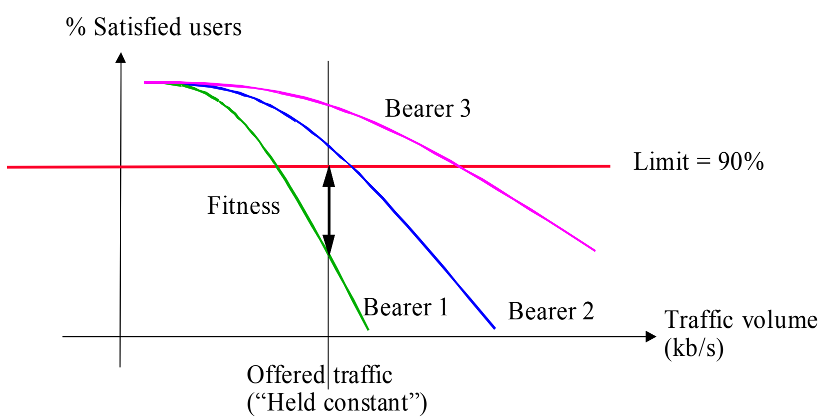


Figure 7.3. Example of fitness computation for three distinct bearer services.

An example of muted offspring is:

$$\textbf{Mutated Offspring} = [\mathbf{p_{1,k}\ p'_{2,k}\ p_{3,k}\ p_{4,k} \ldots\ p'_{14,k}}],$$

where $\mathbf{p'_{2,k}}$ and $\mathbf{p'_{14,k}}$ are the altered components.

The new individual is then checked. If it is a duplicate or any of the components is inconsistent (e.g. MinAllowedBitRate > MaxAllowedBitRate), the offspring is regenerated repeating the above process. If the individual is accepted, its fitness is computed (a new simulation with these parameters is run) and the new candidate is put in the above table, where all individuals are collected. The table is then sorted based on the fitness values and cropped taking out the worst performing individual.

If the best fitness improves then we repeat the process for a maximum of 100 times, otherwise the best candidate in the table is selected being the set of "optimal" parameters for that particular traffic mix and volume. The above algorithm is presented with a flow chart in Figure 7.4. The corresponding pseudo code is described in Appendix E.

7.4 Specific Models and Simulation Assumptions

This section describes the case studies for QoS optimisation and spectral efficiency gains computation. More information on the adopted traffic, system and environment models can be found in Sections 3.2-3.4. The mapping of services onto QoS profiles is explained in Section 7.4.1. The user satisfaction criteria thereof were defined in Section 3.6.

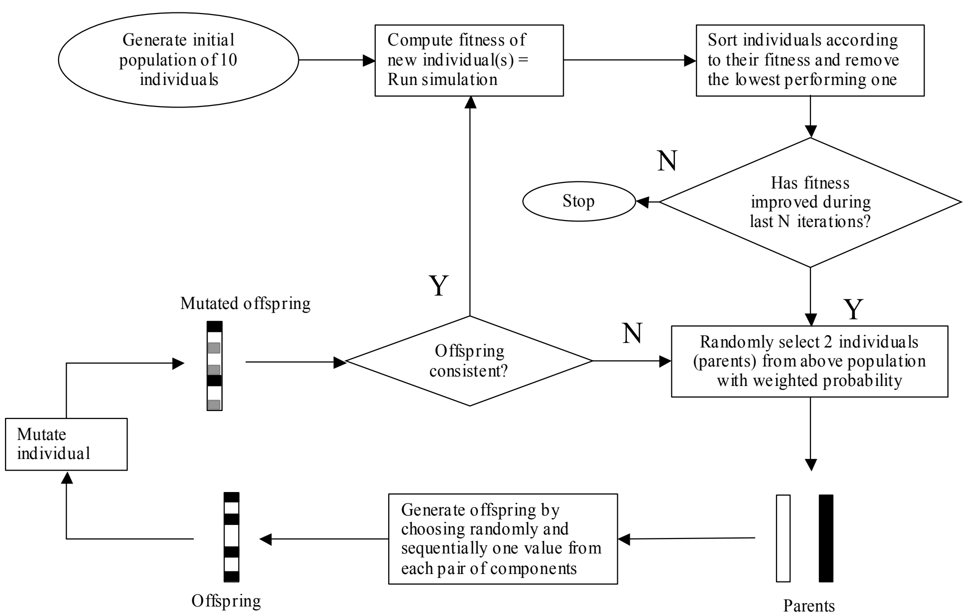


Figure 7.4. Genetic approach flow chart.

7.4.1 Offered Services and QoS Profiles

In this chapter, the performances of the following services are investigated: CS speech and video calls; PS SWIS, PoC, WAP, MMS and dialup connections. Speech, video and SWIS were offered with guaranteed bit rate (GB), and all other services were run on the best effort (NGB). Speech and video calls were served as CS-Conversational, whereas SWIS was carried on PS-Streaming class. PoC, streaming and WAP/MMS were mapped onto PS-Interactive. Dialup connections, which comprised, for example, ftp, emails and http traffic, were carried on PS-Background class. Hence, in the admission control queue management, speech calls had top priority, followed by video and SWIS calls. In the packet scheduler queues, within the Interactive class, using different traffic handling priorities, PoC was handled first, followed by streaming and WAP/MMS. Dialup was served in the end.

The specific traffic models are reported in Table 7.3. Traffic mix, in share of calls, and the traffic intensity, in arrivals per second, were set differently depending on the case study (see Section 7.4.3 and 7.4.4).

7.4.2 Simulation Environment

Performance results were analysed using a macro cellular network located in the downtown of Helsinki (see Figure 3.2), where terminals were uniformly randomly distributed, but not on the water. The simulations were performed over a period of 2500 s and 5000 s for the QoS optimisation and spectral efficiency gains computation, respectively; using a time step of 200 ms. Statistics were collected over the entire simulation periods. The impact of the transient time and statistical confidence on the simulation results is discussed in Appendix C. Table 7.4 reports the most important network parameters (see also Table B.1).

Table 7.3. Adopted traffic models.

Service type	**Data rate (kb/s)**	**Buffer size (s)**	**Object size (k octets)**	**Off time (s)**	**Session length (Objects)**	**Mean session size (k octets)**
PoC	8	1	Exponential 6 mean, 0.5 min, 40 max	Exponential 60 mean, 1 min, 1200 max	Geometric 8 mean, 1 min, 30 max	42
Streaming	64	8	Uniform 160 min, 3200 max	-	1	1680
MMS	Best effort	-	Exponential 20 mean, 3 min, 200 max	-	1	20
Dialup	Best effort	-	Log-normal (μ=5, σ=1.8) 0.1 min, 20000 max	Pareto (k=2, α=1) 2 min, 3600 max	Inv. Gaussian (μ=3.8, λ=6) 1 min, 50 max	600
SWIS	64	1	Exponential 80 mean, 32 min, 2400 max	-	1	80
WAP	Best effort	-	Log-normal (μ=2, σ=1) 0.1 min, 50 max	Exponential 20 mean, 1 min, 600 max	Geometric 3 mean, 1 min, 50 max	200
Speech	12.2	-	-	-	Exponential 90 s	135
Video	64	-	-	-	Exponential 120 s	960

Table 7.4. Most important system based parameters.

Parameter	Value
Radio resource indication period (*RRI*)	0.2 s
Simulation time (s)	7200 s
Power target for DL AC	3 dB below BTS total power
Overload offset for DL AC	1 dB above power target
Orthogonality (α)	0.5
Period for load control actions	0.2 s (1 RRI)
Period for Packet Scheduling	0.2 s (1 RRI)
Length of AC queue	10 Radio bearers
Dedicated NGB capacity	0 dB, i.e. not used
Power weight for inactive NGB traffic (k)	0.5

7.4.3 Case 1 – Parameter Optimisation

Table 7.5 reposts the traffic mixes upon which the test parameters of Table 7.1 were optimised. In brief: Mix 1 was derived from a traffic prediction for 2007, whereas all other cases were made-up for this study. Mix 2 takes into account PoC launch in ´05. SWIS is not considered, because we assumed people would mainly make use of video calls in this year. Mix 3 excludes the novel services, such as PoC and SWIS. Mix 4 proposes the same share of calls for all services offered on best effort. Mix 5 assumes all services in place (deployed in the network) and PoC to be the most used packet switched application. The network optimisation is performed for each traffic mix at a given traffic volume. The mean call (or session) arrival rate in the system was set equal to 0.2 s for Mix 1 and equal to 0.4 s for all the other mixes. Table 7.6 and Table 7.7 report the corresponding offered traffic in the system and per cell, respectively.

7.4.4 Case 2 – Spectral Efficiency Computation

The gain provided by the "optimal" parameter values with respect to the undifferentiated ones is expressed in terms of spectral efficiency. Statistics are collected at a given traffic volume. The volume of traffic is varied for each of the input traffic mixes from a minimum, where almost all users are satisfied, to a maximum, where the services performance are no longer tolerable. In practice, this would mean for the operator to retrieve measurements separately during distinct hours of the day (week) where the traffic mix and volume is similar. The spectral efficiency is the overall performance indicator derived upon the observed traffic as the cell load where each of the deployed services has at least 90% of the users satisfied. A graphical example of the spectral efficiency computation for three distinct bearer services was shown in Figure 3.9. The spectral efficiency gain (or loss) is the difference between the two spectral efficiency values before and after the change of the parameter settings in the analysed network.

In this case study, the offered traffic volumes in terms of call or session attempts in the radio access network are listed in Table 7.8. The corresponding offered loads in terms of arrivals, erlangs and octets (8 bits) for each of the analysed traffic mixes and services are reported in Table 7.9-Table 7.18.

Table 7.5. Input traffic mixes for “optimal” parameter derivation.

Service type	Traffic mixes (%)				
	Mix 1	**Mix 2**	**Mix 3**	**Mix 4**	**Mix 5**
Speech	32	44	40	40	20
Video	0	3	0	2	7
SWIS	3	0	10	5	10
PoC	14	9	10	25	18
Streaming	2	10	10	4	12
WAP	38	17	10	10	13
MMS	8	5	10	7	5
Dialup	3	12	10	7	15

Table 7.6. Offered traffic volumes in the all system (19 cells) during the simulated time (2500 s) for “optimal” parameter derivation.

Service type	Offered load (System call or session attempts)				
	Mix 1	**Mix 2**	**Mix 3**	**Mix 4**	**Mix 5**
Speech	4000	2750	2500	2500	1250
Video	0	188	0	125	438
SWIS	375	0	625	313	625
PoC	1750	563	625	1563	1125
Streaming	250	625	625	250	750
WAP	4750	1063	625	625	813
MMS	1000	313	625	438	313
Dialup	375	750	625	438	938
Total	**12500**	**6250**	**6250**	**6250**	**6250**

Table 7.7. Offered traffic volumes per cell during the busy hour for "optimal" parameter derivation. Rough estimates from the traffic models of Table 7.3.

Service type	**Offered load (Busy hour/cell)**					
	Mix 1	**Mix 2**	**Mix 3**	**Mix 4**	**Mix 5**	**Unit**
Speech	7.58	5.21	4.74	4.74	2.37	Erl
Video	0.00	**0.47**	0.00	0.32	**1.11**	Erl
SWIS	2274	0	3789	1895	3789	k octets
PoC	5464	1756	1952	4879	3513	k octets
Streaming	**31832**	**79579**	79579	31832	**95495**	k octets
WAP	72000	16105	9474	9474	12316	k octets
MMS	1516	474	947	663	474	k octets
Dialup	17053	34105	28421	19895	42632	k octets
Total CS	7.58	5.68	4.74	5.05	3.47	Erl
Total PS	130.14	132.02	124.16	68.64	158.22	M octets
Total	**171.75**	**174.27**	**150.17**	**103.74**	**203.05**	**M octets**

Table 7.8. Offered traffic volumes for spectral efficiency gains derivation.

Offered load	**Vol 1**	**Vol 2**	**Vol 3**	**Vol 4**	**Vol 5**	**Vol 6**
MeanArivalRate (s)	1.00	0.50	0.33	0.25	0.20	0.17
N of calls (system)	5000	10000	15000	20000	25000	30000
N of calls/cell	263	526	789	1053	1316	1579
N of Calls/Cell/hour	**189**	**379**	**568**	**758**	**947**	**1137**

Table 7.9. Mix 1: Call or session attempts in the system during the simulated time.

Offered load	Vol 1	Vol 2	Vol 3	Vol 4	Vol 5	Vol 6
Speech	1600	3200	4800	6400	8000	9600
Video	0	0	0	0	0	0
SWIS	150	300	450	600	750	900
PoC	700	1400	2100	2800	3500	4200
Streaming	100	200	300	400	500	600
WAP	1900	3800	5700	7600	9500	11400
MMS	400	800	1200	1600	2000	2400
Dialup	150	300	450	600	750	900

Table 7.10. Mix 1: Offered traffic per cell (rough estimate from traffic models for the busy hour).

Offered load	Vol 1	Vol 2	Vol 3	Vol 4	Vol 5	Vol 6	Unit
Speech	1.5	3.0	4.5	6.1	7.6	9.1	Erl
Video	0.0	0.0	0.0	0.0	0.0	0.0	Erl
SWIS	455	909	1364	1819	2274	2728	k octets
PoC	1093	2186	3279	4372	5464	6557	k octets
Streaming	6366	12733	19099	25465	31832	38198	k octets
WAP	14400	28800	43200	57600	72000	86400	k octets
MMS	303	606	909	1213	1516	1819	k octets
Dialup	3411	6821	10232	13642	17053	20463	k octets
Total CS traffic	**1.5**	**3.0**	**4.5**	**6.1**	**7.6**	**9.1**	Erl
Total PS traffic	**26**	**52**	**78**	**104**	**130**	**156**	M octets

Table 7.11. Mix 2: Call or session attempts in the system during the simulated time.

Offered load	**Vol 1**	**Vol 2**	**Vol 3**	**Vol 4**	**Vol 5**	**Vol 6**
Speech	2200	4400	6600	8800	11000	13200
Video	150	300	450	600	750	900
SWIS	0	0	0	0	0	0
PoC	450	900	1350	1800	2250	2700
Streaming	500	1000	1500	2000	2500	3000
WAP	850	1700	2550	3400	4250	5100
MMS	250	500	750	1000	1250	1500
Dialup	600	1200	1800	2400	3000	3600

Table 7.12. Mix 2: Offered traffic per cell (rough estimate from traffic models for the busy hour).

Offered load	**Vol 1**	**Vol 2**	**Vol 3**	**Vol 4**	**Vol 5**	**Vol 6**	**Unit**
Speech	2.1	4.2	6.3	8.3	10.4	12.5	Erl
Video	0.2	0.4	0.6	0.8	0.9	1.1	Erl
SWIS	0	0	0	0	0	0	k octets
PoC	703	1405	2108	2810	3513	4215	k octets
Streaming	31832	63663	95495	127326	159158	190989	k octets
WAP	6442	12884	19326	25768	32211	38653	k octets
MMS	189	379	568	758	947	1137	k octets
Dialup	13642	27284	40926	54568	68211	81853	k octets
Total CS traffic	**2.3**	**4.5**	**6.8**	**9.1**	**11.4**	**13.6**	Erl
Total PS traffic	**53**	**106**	**158**	**211**	**264**	**317**	M octets

Table 7.13. Mix 3: Call or session attempts in the system during the simulated time.

Offered load	**Vol 1**	**Vol 2**	**Vol 3**	**Vol 4**	**Vol 5**	**Vol 6**
Speech	2200	4400	6600	8800	11000	13200
Video	150	300	450	600	750	900
SWIS	0	0	0	0	0	0
PoC	450	900	1350	1800	2250	2700
Streaming	500	1000	1500	2000	2500	3000
WAP	850	1700	2550	3400	4250	5100
MMS	250	500	750	1000	1250	1500
Dialup	600	1200	1800	2400	3000	3600

Table 7.14. Mix 3: Offered traffic per cell (rough estimate from traffic models for the busy hour).

Offered load	**Vol 1**	**Vol 2**	**Vol 3**	**Vol 4**	**Vol 5**	**Vol 6**	**Unit**
Speech	1.9	3.8	5.7	7.6	9.5	11.4	Erl
Video	0.0	0.0	0.0	0.0	0.0	0.0	Erl
SWIS	1516	3032	4547	6063	7579	9095	k octets
PoC	781	1561	2342	3123	3903	4684	k octets
Streaming	31832	63663	95495	127326	159158	190989	k octets
WAP	3789	7579	11368	15158	18947	22737	k octets
MMS	379	758	1137	1516	1895	2274	k octets
Dialup	11368	22737	34105	45474	56842	68211	k octets
Total CS traffic	**1.9**	**3.8**	**5.7**	**7.6**	**9.5**	**11.4**	Erl
Total PS traffic	**50**	**99**	**149**	**199**	**248**	**298**	M octets

Table 7.15. Mix 4: Call or session attempts in the system during the simulated time.

Offered load	Vol 1	Vol 2	Vol 3	Vol 4	Vol 5	Vol 6
Speech	2000	4000	6000	8000	10000	12000
Video	0	0	0	0	0	0
SWIS	500	1000	1500	2000	2500	3000
PoC	500	1000	1500	2000	2500	3000
Streaming	500	1000	1500	2000	2500	3000
WAP	500	1000	1500	2000	2500	3000
MMS	500	1000	1500	2000	2500	3000
Dialup	500	1000	1500	2000	2500	3000

Table 7.16. Mix 4: Offered traffic per cell (rough estimate from traffic models for the busy hour).

Offered load	Vol 1	Vol 2	Vol 3	Vol 4	Vol 5	Vol 6	Unit
Speech	1.9	3.8	5.7	7.6	9.5	11.4	Erl
Video	0.1	0.3	0.4	0.5	0.6	0.8	Erl
SWIS	758	1516	2274	3032	3789	4547	k octets
PoC	1952	3903	5855	7806	9758	11709	k octets
Streaming	12733	25465	38198	50931	63663	76396	k octets
WAP	3789	7579	11368	15158	18947	22737	k octets
MMS	265	531	796	1061	1326	1592	k octets
Dialup	7958	15916	23874	31832	39789	47747	k octets
Total CS traffic	**2.0**	**4.0**	**6.1**	**8.1**	**10.1**	**12.1**	Erl
Total PS traffic	**27**	**55**	**82**	**110**	**137**	**165**	M octets

Table 7.17. Mix 5: Call or session attempts in the system during the simulated time.

Offered load	Vol 1	Vol 2	Vol 3	Vol 4	Vol 5	Vol 6
Speech	2000	4000	6000	8000	10000	12000
Video	100	200	300	400	500	600
SWIS	250	500	750	1000	1250	1500
PoC	1250	2500	3750	5000	6250	7500
Streaming	200	400	600	800	1000	1200
WAP	500	1000	1500	2000	2500	3000
MMS	350	700	1050	1400	1750	2100
Dialup	350	700	1050	1400	1750	2100

Table 7.18. Mix 5: Offered traffic per cell (rough estimate from traffic models for the busy hour).

Offered load	Vol 1	Vol 2	Vol 3	Vol 4	Vol 5	Vol 6	Unit
Speech	0.9	1.9	2.8	3.8	4.7	5.7	Erl
Video	0.4	0.9	1.3	1.8	2.2	2.7	Erl
SWIS	1516	3032	4547	6063	7579	9095	k octets
PoC	1405	2810	4215	5621	7026	8431	k octets
Streaming	38198	76396	114594	152792	190989	229187	k octets
WAP	4926	9853	14779	19705	24632	29558	k octets
MMS	189	379	568	758	947	1137	k octets
Dialup	17053	34105	51158	68211	85263	102316	k octets
Total CS traffic	**1.4**	**2.8**	**4.2**	**5.6**	**6.9**	**8.3**	Erl
Total PS traffic	**63**	**127**	**190**	**253**	**316**	**380**	M octets/cell

7.5 Simulation Results and Discussion

In this section, we present the simulation results of the QoS optimisation using the genetic algorithm presented in Section 7.3.1 and the spectral efficiency gains attained with respect to reference cases. The reference cases are network configurations with default parameter values defined as the most used ones in the differentiated and undifferentiated settings. The simulation assumptions of these case studies were described in Section 7.4.3.

7.5.1 Case 1 – QoS Optimisation Results

Simulation results are illustrated from Figure 7.5 through Figure 7.10 and in Table 7.19. Figure 7.5-Figure 7.9 show the actual fitness value and the related percentage of satisfied users of the deployed services as a function of the iteration number, for each of the traffic mixes reported in Table 7.5. Figure 7.10 presents the corresponding best fitness values over the actual number of iterations. From this figure conclusions upon the convergence (efficiency) of the genetic approach and the effectiveness of QoS optimisation upon the five cases can be drawn. The "optimal" parameter values for each of the input traffic mixes using the traffic volumes listed in Table 7.6 are reported in Table 7.19, where the rows denoted by Ref 1 and 2 collect the most common values obtained with the mixes (reference cases). Ref 1 is the differentiated network configuration and Ref 2 is the network configuration with no parameter differentiation (common cell-based values).

7.5.1.1 Mix 1: Simulation Results

Figure 7.5 depicts the performance results for Mix 1. The resulting traffic volume in this case is rather small, hence the percentage of satisfied users for each of the analysed services result above the target threshold (90%) in most of the iterations. Despite the slight margin of amelioration, from the graphs, the parameter values achieved their "optimum" after about twenty iterations, where the optimisation process could have been stopped. This is also clear from Figure 7.10, where the actual best values of the objective function are displayed. However, to understand the behaviour of the algorithm, the simulation continued up to fifty iterations, where a new point of equilibrium was achieved after several adverse mutations. The performance deterioration in the intermediate steps was worse than the one attained deploying the initial population, which was selected randomly within the parameter ranges. This appears to be a characteristic of this particular solution space search algorithm, which should be controlled in a real network by limiting the range of values of the mutated parameter set.

7.5.1.2 Mix 2: Simulation Results

The performance results for Mix 2 are shown in Figure 7.6 and in Figure 7.10. After thirty iterations, the improvement of the objective function may be neglected. In this case, the resulting traffic volume for Mix 2 is fairly high for the differentiation to provide the target performance for all the offered services: 20% of the dialup users remained unsatisfied. The optimisation process resulted however effective for more than 20% of streaming and dialup users. Considering the user satisfaction criteria of Section 3.6 and the not differentiated target thresholds (90% for all services), we conclude that, at this traffic volume and for this particular traffic mix, by adopting the QoS differentiation proposed in this work, the overall service performances would remain tolerable and the network administrator would not need to provision additional radio resources.

7.5.1.3 Mix 3: Simulation Results

The optimisation results for Mix 3 are depicted in Figure 7.7 and in Figure 7.10. In this case, we can notice that the target performance in terms of percentage of satisfied users is achieved in about thirty iterations. After that, the fitness improvement is rather small and the optimisation process could have been stopped. This is the scenario (traffic mix and volume) that mostly benefitted from the optimisation process. In this case, the parameter adaptation could well compensate for the missing performance: The reduction of the traffic congestion in the packet scheduling queues made it possible for all the offered services to achieve the related target performance.

7.5.1.4 Mix 4: Simulation Results

Simulation results for Mix 4 are shown in Figure 7.8 and in Figure 7.10. Even in this case the traffic volume is rather small and the performance of each of the offered services were already above the target with the initial random parameter settings. Figure 7.10 revels that after twenty iterations the improvement of the fitness function can have been neglected and the simulation stopped.

7.5.1.5 Mix 5: Simulation Results

Figure 7.9 and Figure 7.10 depict the performance results for Mix 5. As we can see from the figures, this is the worst scenario, where the QoS differentiation is not sufficient to compensate for the loss in performance of streaming and dialup connections due to the congestion of capacity requests in the packet scheduler queues. The optimisation process still benefits 30% of the streaming and dialup users, but in practice this would be the case for the network administrator to provision more radio recourses.

7.5.1.6 Mix 1-5: “Optimal” Parameter Values

Let us now examine the “optimal” parameter values reported in Table 7.19.

As expected, the minimum allowed bit rate for PoC (MC 1) and streaming (MC 2) is equal to the related source bit rate (see Table 7.3) in most of the analysed cases. However, in some of the analysed scenarios, better performance were attained allocating to the transport channel in question a higher bit rate. From Table 7.6, this appears to be when the % of arrivals of that particular application is rather low compared to the other services. For WAP/MMS (MC 3) and dialup (MC 4) the most used minimum allowed bit rate is equal to the target active session throughput of the corresponding user satisfaction criterion (see Section 3.6). However, even for these applications, in some of the investigated traffic scenarios, the performances improved when a higher transport channel bit rate was utilised.

Only in one case the “optimal” maximum allowed bit rate for PoC (MC 1) and streaming (MC 2) did not correspond to the maximum bit rate of the related TFS (see Table 7.1). This is due to the fact that a transmission rate higher than the source bit rate at layer 2 compensated for the data gathering in the RNC (data deficit in the UE buffer) caused by the temporary congestions of capacity requests in the packet scheduler queues. For WAP/MMS (MC 3) and dialup (MC 4) the “optimal” maximum allowed bit rate depended strongly on the traffic conditions, and for better performance, in practice, could assume any value within the allowed range.

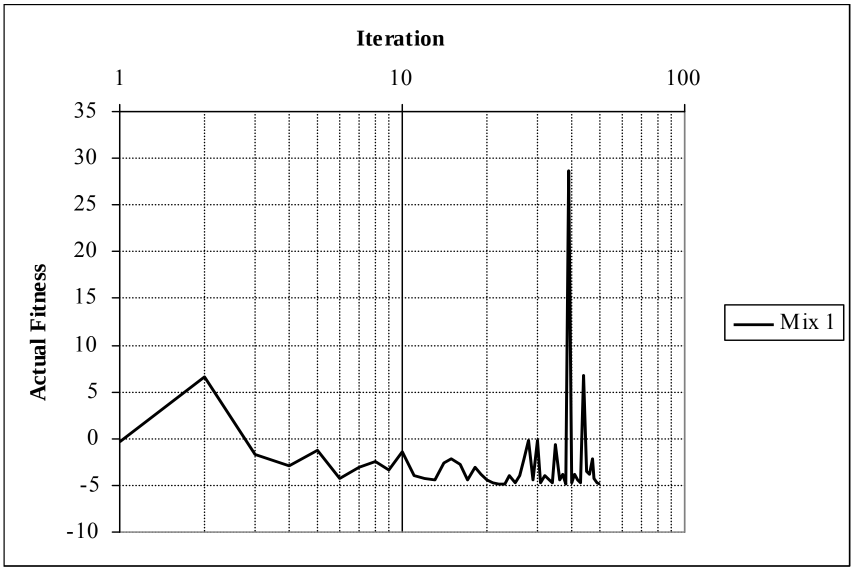
The inactivity timer had no influence on streaming (MC 2) performance, since the session consisted of only one packet call (see Table 7.3). For PoC (MC 1) the "optimal" parameter value varies from 1 to 2 s. In the adopted traffic models, such a short time reduced the number of rejections of new capacity requests in the PS, without compromising the performance of the ongoing communications being in idle (DTX) for much longer periods (see Table 7.3). In practice, a longer inactivity timer may be beneficial for the PoC performance, which may be adversely affected by the delays in the transport channel (DCH-FACH) commutations. For the other service applications carried on MC 3 and MC 4, the "optimal" inactivity timer depended on the traffic conditions; thought 5 s turned out to be a good compromise.

In this study, the DCH minimum allocation time and the CR maximum queuing time were not differentiated. The "optimal" parameter values ranged from 5 to 20 s depending on the traffic situations. Hence, for the reference case an intermediate value was selected.

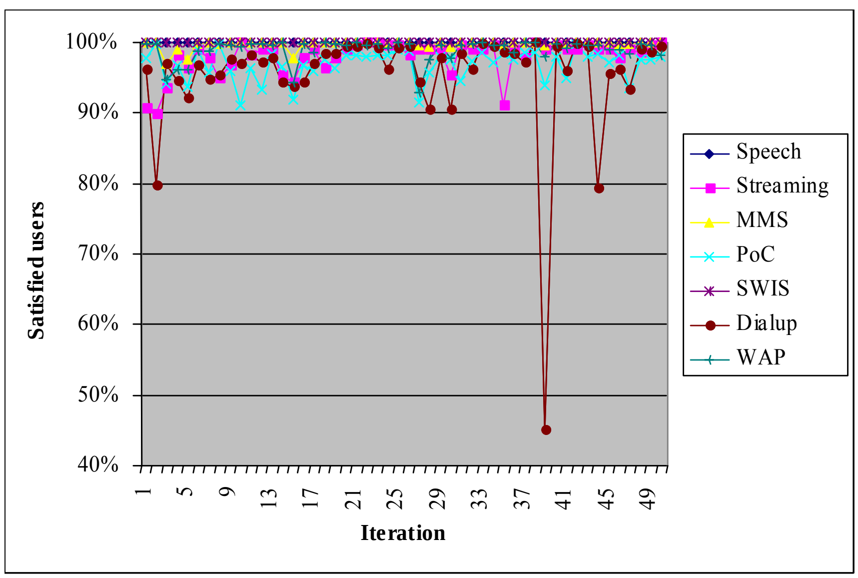
Simulation results showed that "optimal" parameter settings depend on the input traffic mix, and the genetic algorithm proved to be a feasible automated multi-parameter-tuning component. In the following section, some practical considerations regarding the research methodology are reported. The spectral efficiency gains provided by the "optimal" settings with respect to the most common values obtained with the mixes (denoted by Ref 1 and Ref 2 for the differentiated and undifferentiated case, respectively) are presented in Section 7.5.2. On these grounds, conclusions upon the business opportunity for the network administrator to optimise the QoS differentiation in UTRAN are drawn.

7.5.1.7 Practical Considerations

The genetic algorithms have been shown to be robust for online parameter adaptation of various dynamic systems. The usage of optimising WCDMA network parameters in a simulator environment was described above. Though infrequent, Figure 7.5-Figure 7.9 showed that some combinations of parameter settings during the optimisation process yielded poor user satisfaction of some of the simulated services. In online optimisation of a running WCDMA network such performance drops are naturally disadvantageous. The network performance needs thus to be monitored actively during the whole process. In practice, performance data for the fitness function calculations are collected for a sufficiently long period (for instance, one or few days after each configuration change) in order to get statistically a significant sample. However, substantial decrease in performance can be detected with a smaller sample size than the one required for making differences among high satisfaction ratios. In the occurrence of problem performance, the cycle can be cancelled and new parameter values can be generated for evaluation. It is beneficial to analyse the performance decrease for potential reasons that created it, and utilise findings to modify the rules on how the new offspring is recombined from the population members before continuing with the next cycles. Such manual adaptations rule out a fully automated process but the benefits can easily outweigh the needed effort. Another way to maintain the stability of optimised network is to allow small changes in the offspring generation only. However, the convergence is likely to slow down as a trade-off. A possibility is therefore to implement the genetic algorithm in a planning tool connected to the NMS of the WCDMA network. The optimisation process would then be much faster at the expense of accuracy, which could be improved by importing measures made periodically in the real network.

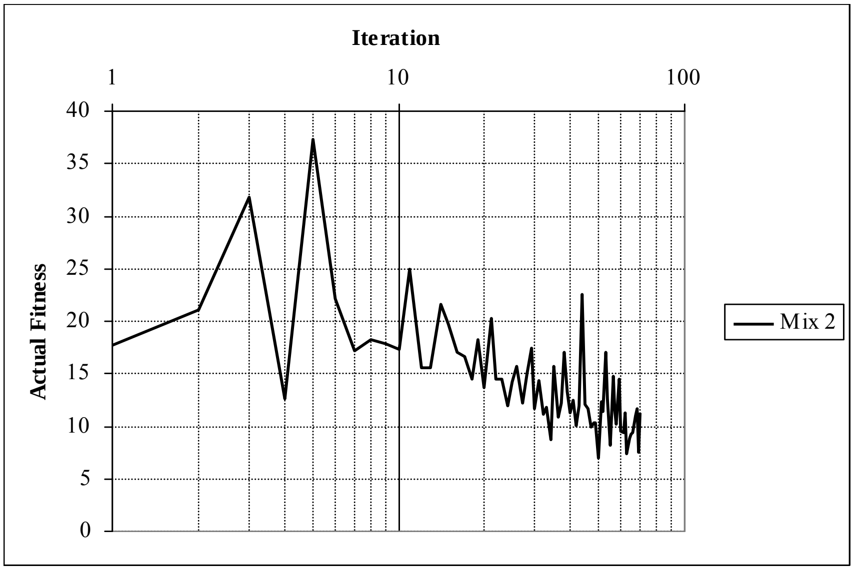


a)

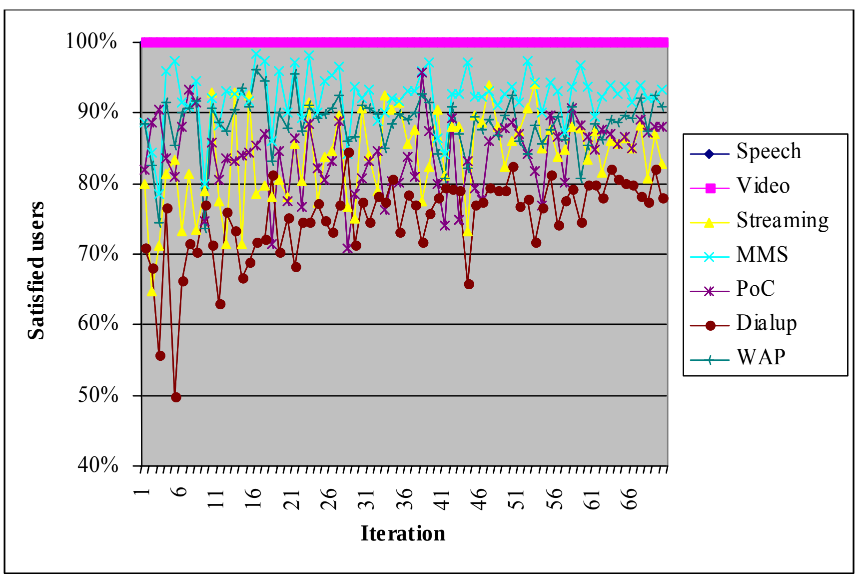


b)

Figure 7.5. Mix 1: a) Fitness values as a function of the iteration number. b) Percentage of satisfied users as a function of the iteration number.

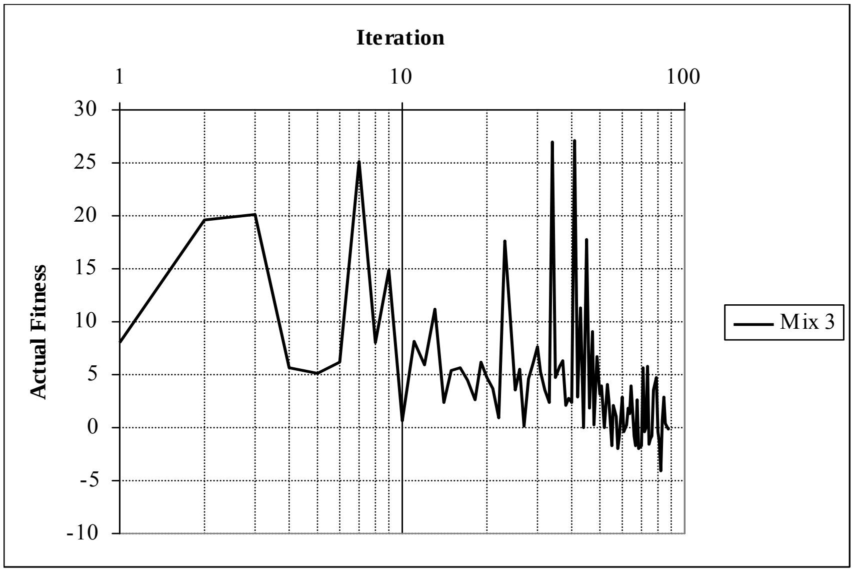


a)

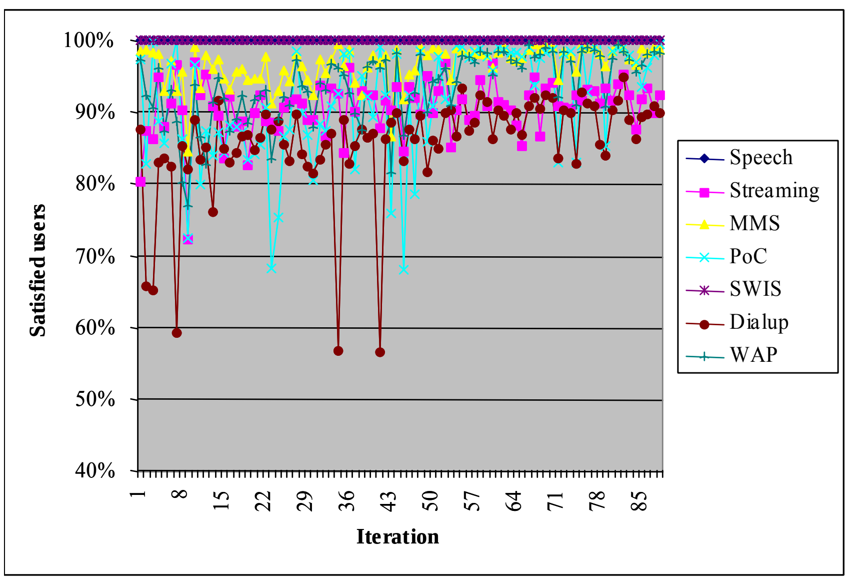


b)

Figure 7.6. Mix 2: a) Fitness values as a function of the iteration number. b) Percentage of satisfied users as a function of the iteration number.

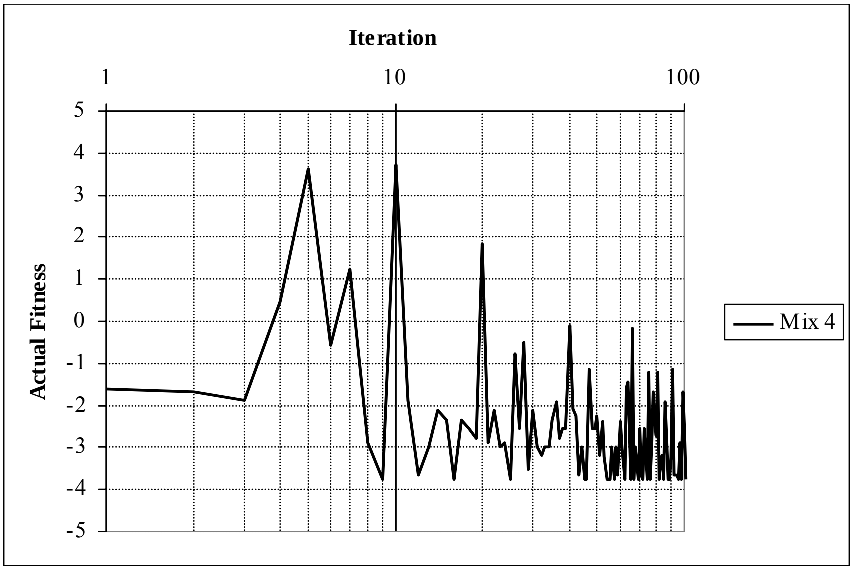


a)

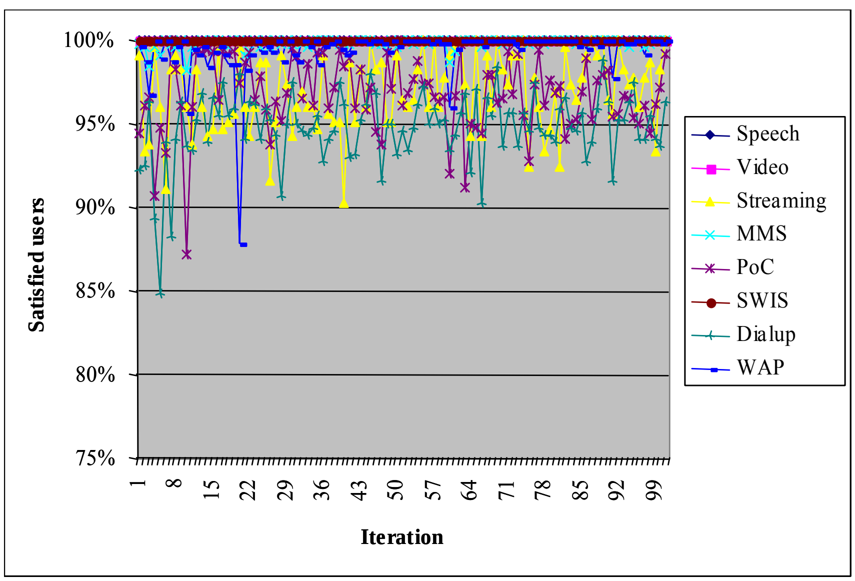


b)

Figure 7.7. Mix 3: a) Fitness values as a function of the iteration number. b) Percentage of satisfied users as a function of the iteration number.

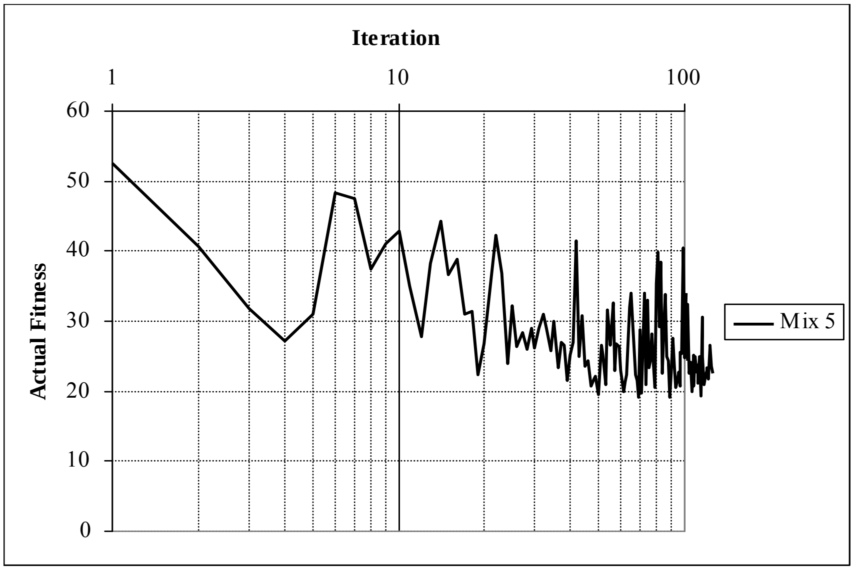


a)

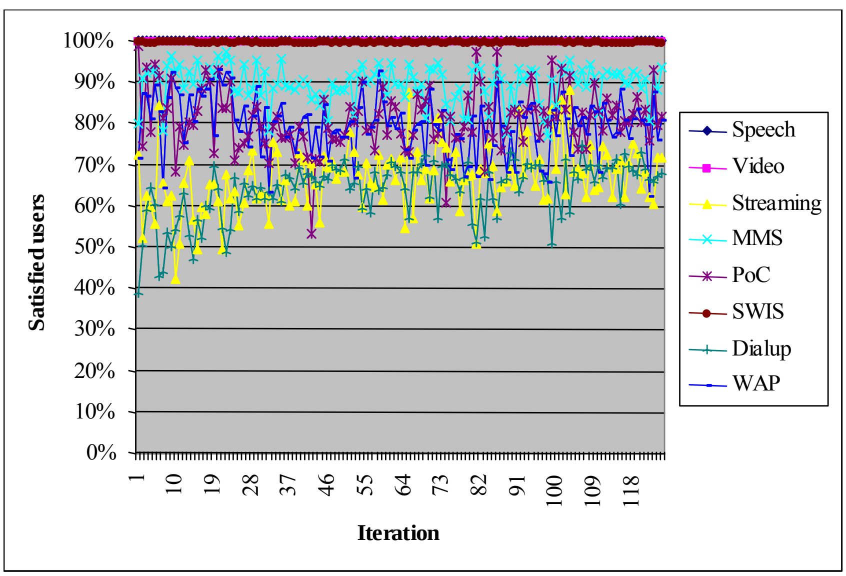


b)

Figure 7.8. Mix 4: a) Fitness values as a function of the iteration number. b) Percentage of satisfied users as a function of the iteration number.



a)



b)

Figure 7.9. Mix 5: a) Fitness values as a function of the iteration number. b) Percentage of satisfied users as a function of the iteration number.

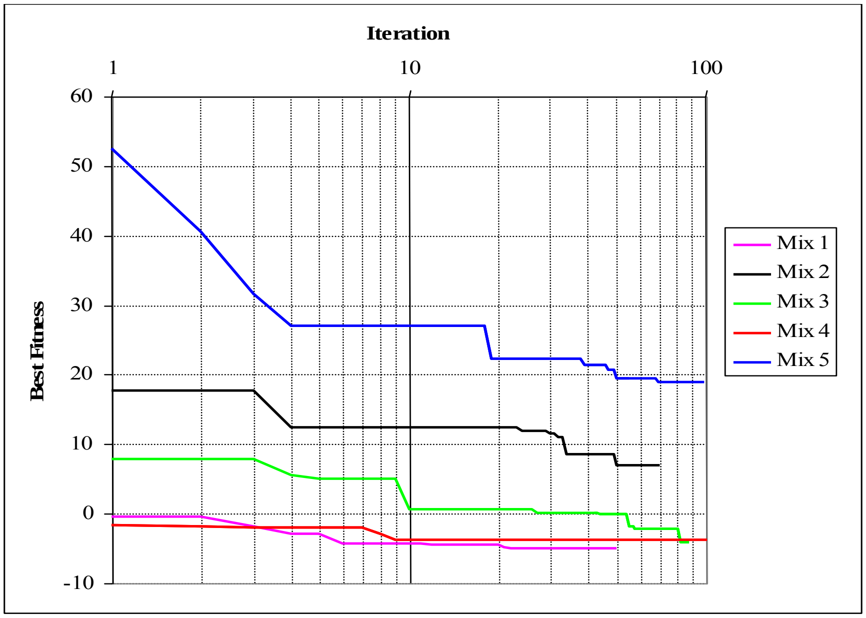


Figure 7.10. Best fitness values as a function of the iteration number for all analysed traffic mixes.

Table 7.19. “Optimal” parameter values for the different traffic mixes and reference settings for differentiated (Ref 1) and undifferentiated provisioning (Ref 2).

Traffic	**Min Allowed Bit Rate (kb/s)**				**Max Allowed Bit Rate (kb/s)**				**Inactivity Timer (s)**				**DCH Min. Alloc.**	**CR Max. Queuing**
Mix	**MC1**	**MC2**	**MC3**	**MC4**	**MC1**	**MC2**	**MC3**	**MC4**	**MC1**	**MC2**	**MC3**	**MC4**	**Time (s)**	**Time (s)**
1	8	**128**	32	64	16	128	64	144	1	1	2	5	5	10
2	8	64	32	64	16	**64**	384	128	1	5	30	1	20	5
3	**16**	64	32	64	16	128	144	144	2	1	5	1	20	20
4	8	64	**64**	**128**	16	128	256	144	1	2	5	5	2	10
5	8	64	32	**128**	16	128	144	256	2	1	1	20	15	5
Ref 1	8	64	32	64	16	128	144	144	1	1	5	5	15	10
Ref 2	64	64	64	64	64	128	144	144	5	5	5	5	15	10

7.5.2 Case 2 – Spectral Efficiency Gains

This section discusses the spectral efficiency gains that would be attained by an operator, at a given traffic mix, if the QoS optimisation process presented in Section 7.3 was adopted to provision the QoS instead of using default parameter settings. Conclusions are drawn for each traffic mix based on the difference between the system loads (average cell throughput taken over the all base stations and simulated time. The cell capacity is by definition the system load divided by the exploited radio spectrum, i.e. 3.84 Mchip/s) where any of the offered services has at least 90% of users satisfied using the reference network configurations (Ref 1 and Ref 2 in Table 7.19, for the differentiated and undifferentiated case, respectively) and the "optimal" QoS provisioning, as illustrated in Figure 3.9 for three bearer services. Simulation results for Mix 1-5 are reported in Table 7.20-Table 7.25 and shown through Figure 7.11 to Figure 7.23, where G_{21} denotes the gain provided by Ref 1 with respect to Ref 2; G_{32} is the gain provided by the "optimal" parameter settings with respect to Ref 1; and G_{31} is the gain given by the "optimal" parameter settings with respect to Ref 2. The simulation assumptions of this case study were described in Section 7.4.4.

7.5.2.1 Mix 1: Simulation Results

Simulation results for Mix 1 are reported in Table 7.20 and in Figure 7.11 and Figure 7.12. Table 7.20 collects the percentage of satisfied users for each of the analysed services as a function of the measured average cell throughput over all cells in the system during the total simulation period. The results relate to Ref 1, Ref 2 and "optimal" parameter settings in the network. A graphical representation thereof is depicted in Figure 7.11 a), b) and Figure 7.12 a). Figure 7.12 b) shows the worst performing service in the case of Ref 1, Ref 2, and "optimal" QoS provisioning as a function of the exploited cell capacity.

As expected, real time services, i.e. speech and SWIS, offered with guaranteed bit rate were not affected by the different network configurations: In all cases, 100% of the users was satisfied. The offered load for such applications was rather low (see Table 7.10) and the total downlink transmission power hardly ever exceeded the target threshold. Among the NGB services, dialup users, followed by streaming, PoC, WAP and MMS connections, experienced the worst quality. PoC was the service that mainly took advantage (benefitted) from the differentiated (Ref 1) and "optimal" parameter settings.

The spectral efficiency gains can be derived graphically from Figure 7.12 b), where the worst performing services of the three cases (Ref 1, Ref 2 and "optimal" QoS provisioning) are confronted. The attained spectral efficiency values and the corresponding gains are reported in Table 7.25, Figure 7.21, Figure 7.22 and Figure 7.23. In this case, the differentiated configuration (Ref 1) performed slightly better than the "optimal" settings. This means that the traffic quantity plays also a role in the optimisation process and, in this case, the multi-parameter tuning should have been performed at a higher traffic volume. The spectral efficiency gains (denoted by G_{21} and G_{31} in Table 7.25) provided by the differentiated settings (Ref 1, and "optimal" parameter values) with respect to the undifferentiated case, Ref 2, were around 5%. Higher gains would have been attained if streaming had been considered the worst performing service.

7.5.2.2 Mix 2: Simulation Results

Simulation results for Mix 2 are reported in Table 7.21 and in Figure 7.13 and Figure 7.14. Table 7.21 collects the percentage of satisfied users for each of the analysed services as a function of the measured average cell throughput over all cells in the system during the total simulation period. As in the previous case, these numbers relate to Ref 1, Ref 2 and "optimal" parameter settings. A graphical representation thereof is depicted in Figure 7.13 a), b) and Figure 7.14 a). Figure 7.14 b) shows the worst performing service in the case of Ref 1, Ref 2, and "optimal" QoS provisioning as a function of the exploited cell capacity.

As for Mix 1, real time services, i.e. speech and video, offered with guaranteed bit rate were not affected by the different network configurations: In all cases, 100% of the users was satisfied. In this case, dialup and streaming users, followed by PoC, WAP and MMS connections, experienced the worst quality. PoC and dialup were the services that mainly profited from the differentiated (Ref 1) and "optimal" parameter settings.

The spectral efficiency gains can be derived graphically from Figure 7.14 b), where the worst performing services of the three cases (Ref 1, Ref 2 and "optimal" QoS provisioning) are compared. The spectral efficiency values and the corresponding gains are reported in Table 7.25, Figure 7.21, Figure 7.22 and Figure 7.23. In this case, the differentiated configuration Ref 1 performed worse than the undifferentiated settings Ref 2. The spectral efficiency gains (denoted by G_{32} and G_{31} in Table 7.25) provided by the "optimal" parameter settings with respect to the reference configurations, Ref 2 and Ref 1, were around 15%. The same numbers would have been attained if streaming and PoC had been considered the worst performing services.

7.5.2.3 Mix 3: Simulation Results

Simulation results for Mix 3 are reported in Table 7.22 and in Figure 7.15 and Figure 7.16. Table 7.22 collects the percentage of satisfied users for each of the analysed services as a function of the measured average cell throughput over all cells in the system during the total simulation period. As in the previous cases these numbers relate to Ref 1, Ref 2 and "optimal" parameter settings. A graphical representation thereof is depicted in Figure 7.15 a), b) and Figure 7.16 a). Figure 7.16 b) shows the worst performing service in the case of Ref 1, Ref 2, and "optimal" QoS provisioning as a function of the exploited cell capacity.

As for Mix 1 and 2, real time services, i.e. speech and SWIS, offered with guaranteed bit rate were not affected by the different network configurations: In all cases, 100% of the users was satisfied. In this case, dialup and streaming users, followed by PoC, WAP and MMS connections, experienced the worst quality. Streaming and dialup were the services that mainly benefitted from the differentiated (Ref 1) and "optimal" parameter settings.

The spectral efficiency gains can be derived graphically from Figure 7.16 b), where the worst performing services of the three cases (Ref 1, Ref 2 and "optimal" QoS provisioning) are compared. The spectral efficiency values and the corresponding gains are reported in Table 7.25, Figure 7.21, Figure 7.22 and Figure 7.23. In this case, the differentiated configuration (Ref 1) performed slightly worse than the "optimal" settings. The spectral efficiency gains (denoted by G_{21} and G_{31} in Table 7.25) provided by the differentiated settings (Ref 1, and "optimal" parameter values) with respect to the undifferentiated case, Ref 2, were around 8%.

7.5.2.4 Mix 4: Simulation Results

Simulation results for Mix 3 are reported in Table 7.23 and in Figure 7.17 and Figure 7.18. Table 7.23 collects the percentage of satisfied users for each of the analysed services as a function of the measured average cell throughput over all cells in the system during the total simulation period. As in the previous cases, these numbers relate to Ref 1, Ref 2 and "optimal" parameter settings. A graphical representation thereof is depicted in Figure 7.17 a), b) and Figure 7.18 a). Figure 7.18 b) shows the worst performing service in the case of Ref 1, Ref 2, and "optimal" QoS provisioning as a function of the exploited cell capacity.

As for Mix 1-3, real time services, i.e. speech, video and SWIS, offered with guaranteed bit rate were not affected by the different network configurations: In all cases, 100% of the users was satisfied. Among the NGB bearer services, dialup and streaming users, followed by PoC, WAP and MMS connections, experienced the worst quality. As a result, PoC, streaming and dialup were the services that mainly profited from the differentiated (Ref 1) and "optimal" parameter settings.

The spectral efficiency gains can be derived graphically from Figure 7.18 b), where the worst performing services of the three cases (Ref 1, Ref 2 and "optimal" QoS provisioning) are compared. The spectral efficiency values and the corresponding gains are reported in Table 7.25, Figure 7.21, Figure 7.22 and Figure 7.23. In this case, the differentiated configuration (Ref 1) performed worse than the "optimal" settings. The spectral efficiency gains (denoted by G_{32} and G_{31} in Table 7.25) provided by the "optimal" parameter settings compared to the reference configurations, Ref 1 and Ref 2, were around 8% and 12%, respectively.

7.5.2.5 Mix 5: Simulation Results

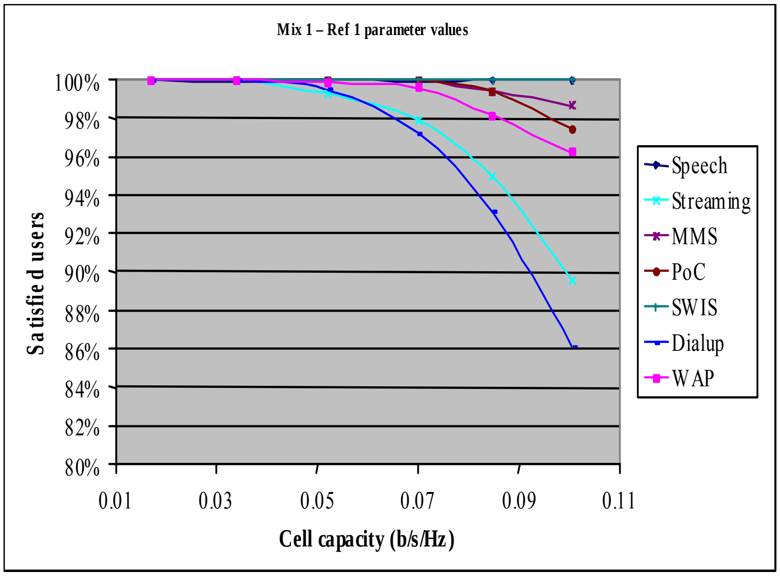
Simulation results for Mix 3 are reported in Table 7.24 and in Figure 7.19 and Figure 7.20. Table 7.24 collects the percentage of satisfied users for each of the analysed services as a function of the measured average cell throughput over all cells in the system during the total simulation period. As in the previous cases, these numbers relate to Ref 1, Ref 2 and "optimal" parameter settings. A graphical representation thereof is depicted in Figure 7.19 a), b) and Figure 7.20 a). Figure 7.20 b) shows the worst performing service in the case of Ref 1, Ref 2, and "optimal" QoS provisioning as a function of the exploited cell capacity.

As for Mix 1-4, real time services, i.e. speech, video and SWIS, offered with guaranteed bit rate were not affected by the different network configurations: In all cases, 100% of the users was satisfied. Among the NGB bearer services, dialup and streaming users, followed by PoC, WAP and MMS connections, experienced the worst quality. As a result, PoC, streaming and dialup were the services that mainly benefitted from the differentiated (Ref 1) and "optimal" parameter settings.

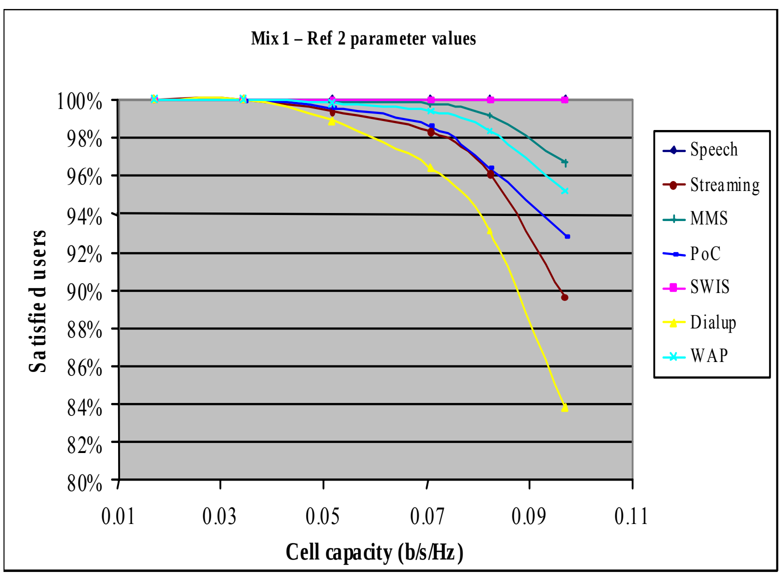
The spectral efficiency gains can be derived graphically from Figure 7.20 b), where the worst performing services of the three cases (Ref 1, Ref 2 and "optimal" QoS provisioning) are compared. The spectral efficiency values and the corresponding gains are reported in Table 7.25, Figure 7.21, Figure 7.22 and Figure 7.23. As for Mix 4, the differentiated configuration (Ref 1) performed worse than the "optimal" settings. The spectral efficiency gains (denoted by G_{32} and G_{31} in Table 7.25) provided by the "optimal" parameter settings compared to Ref 1 and Ref 2 were 5.7% and 7.4%, respectively.

Table 7.20. Mix 1: Percentage of satisfied users for each of the analysed services as a function of the measured average cell throughput taken over all cells in the system.

Parameters (Mix 1)	**Av. Cell throughput (kb/s)**	**Speech**	**Video**	**Streaming**	**MMS**	**PoC**	**SWIS**	**Dialup**	**WAP**
Ref 1	66	100%	0%	100%	100%	100%	100%	100%	100%
	131	100%	0%	100%	100%	100%	100%	100%	100%
	200	100%	0%	99%	100%	100%	100%	100%	100%
	269	100%	0%	98%	100%	100%	100%	97%	100%
	325	100%	0%	95%	99%	99%	100%	93%	98%
	386	**100%**	**0%**	**90%**	**99%**	**97%**	**100%**	**86%**	**96%**
Ref 2	65	100%	0%	100%	100%	100%	100%	100%	100%
	132	100%	0%	100%	100%	100%	100%	100%	100%
	198	100%	0%	99%	100%	100%	100%	99%	100%
	271	100%	0%	98%	100%	99%	100%	96%	99%
	316	100%	0%	96%	99%	96%	100%	93%	98%
	372	**100%**	**0%**	**90%**	**97%**	**93%**	**100%**	**84%**	**95%**
"Optimal"	67	100%	0%	100%	100%	100%	100%	100%	100%
	129	100%	0%	100%	100%	100%	100%	100%	100%
	209	100%	0%	100%	100%	100%	100%	99%	100%
	263	100%	0%	98%	100%	99%	100%	98%	100%
	311	100%	0%	96%	99%	99%	100%	95%	99%
	382	**100%**	**0%**	**91%**	**98%**	**97%**	**100%**	**86%**	**97%**

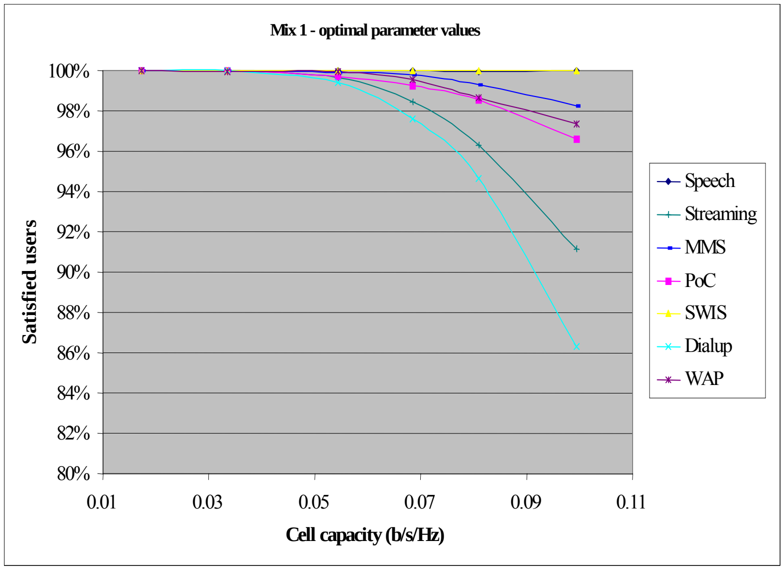


a)

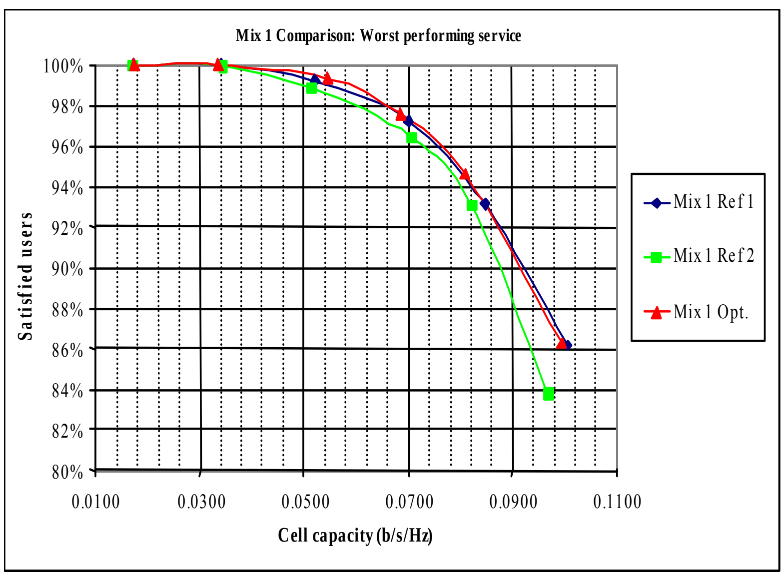


b)

Figure 7.11. Percentage of satisfied users for each of the analysed services as a function of the measured average cell throughput taken over all cells in the system: a) Mix 1, Ref 1 parameter values; b) Mix 1, Ref 2 parameter values.



a)

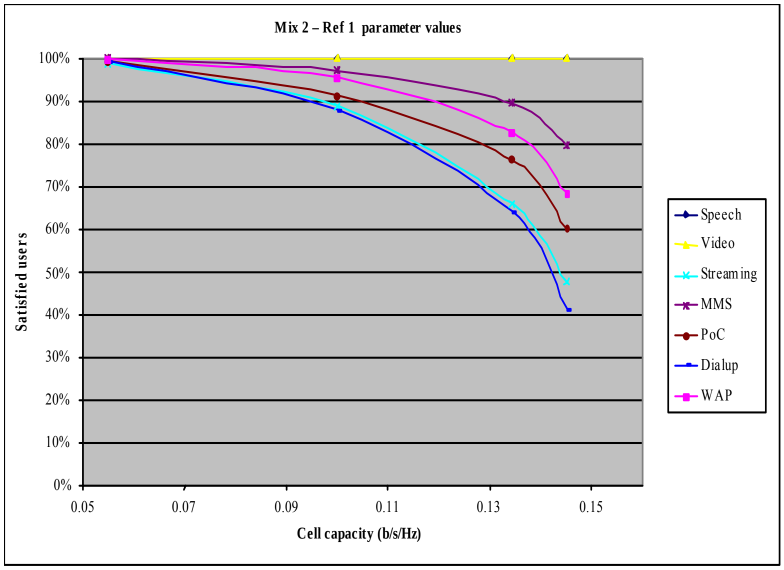


b)

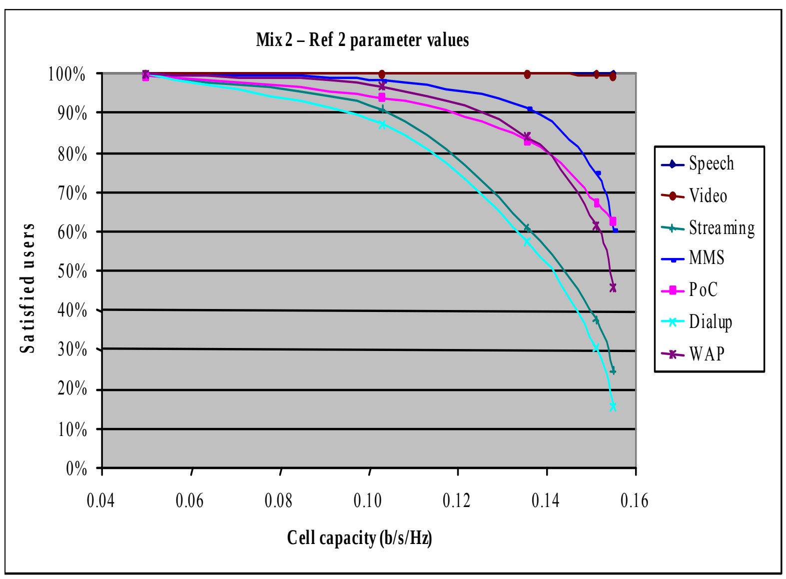
Figure 7.12. a) Mix 1 – "optimal" parameter values: Percentage of satisfied users for each of the analysed services as a function of the measured average cell throughput taken over all cells in the system; b) Mix 1 – Ref 1, Ref 2 and "optimal" parameter values: comparison in terms of the worst performing service for spectral efficiency gain computation.

Table 7.21. Mix 2: Percentage of satisfied users for each of the analysed services as a function of the measured average cell throughput taken over all cells in the system.

Parameters (Mix 2)	Av. Cell throughput (kb/s)	Speech	Video	Streaming	MMS	PoC	SWIS	Dialup	WAP
Ref 1	211	100%	100%	99%	100%	100%	0%	100%	100%
	384	**100%**	**100%**	**89%**	**97%**	**91%**	**0%**	**88%**	**95%**
	516	100%	100%	66%	90%	76%	0%	64%	83%
	558	100%	100%	48%	80%	60%	0%	41%	68%
Ref 2	190	100%	100%	100%	100%	100%	0%	100%	100%
	395	**100%**	**100%**	**91%**	**98%**	**94%**	**0%**	**87%**	**97%**
	520	100%	100%	61%	91%	83%	0%	58%	84%
	580	100%	100%	38%	75%	67%	0%	31%	61%
	595	100%	99%	25%	61%	63%	0%	16%	46%
“Optimal”	195	100%	100%	100%	100%	100%	0%	100%	100%
	401	**100%**	**100%**	**93%**	**98%**	**97%**	**0%**	**93%**	**98%**
	539	100%	100%	71%	95%	91%	0%	60%	90%
	593	100%	100%	45%	82%	85%	0%	33%	73%
	610	100%	99%	30%	70%	78%	0%	21%	58%

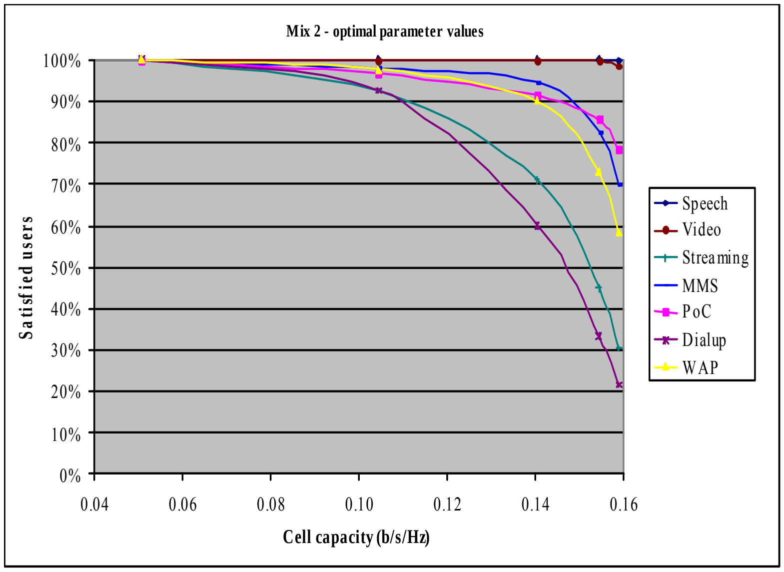


a)

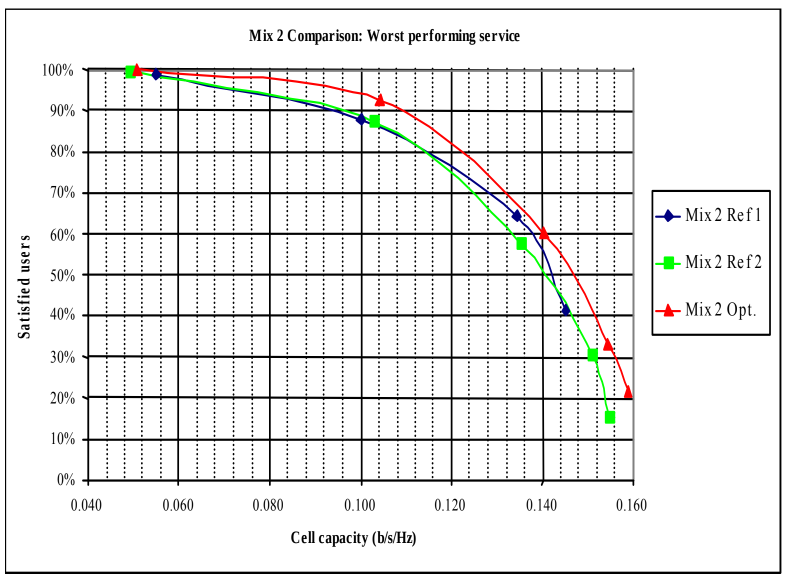


b)

Figure 7.13. Percentage of satisfied users for each of the analysed services as a function of the measured average cell throughput taken over all cells in the system: a) Mix 2, Ref 1 parameter values; b) Mix 2, Ref 2 parameter values.



a)

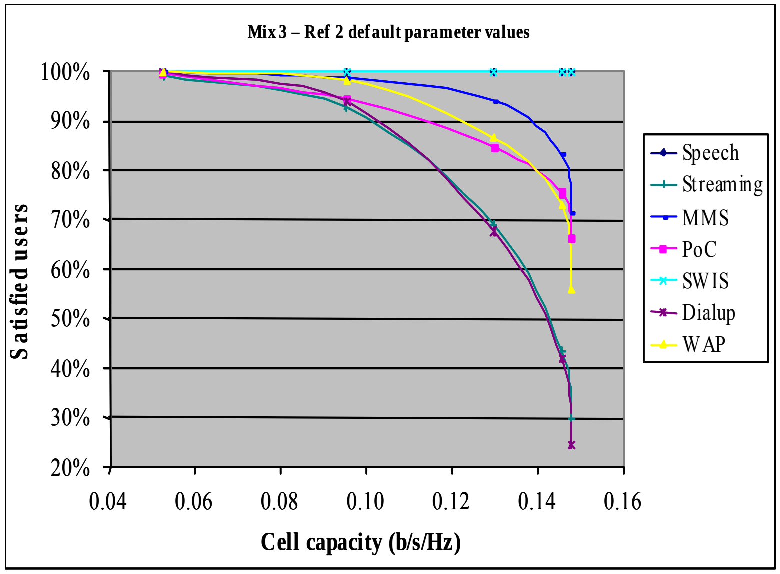


b)

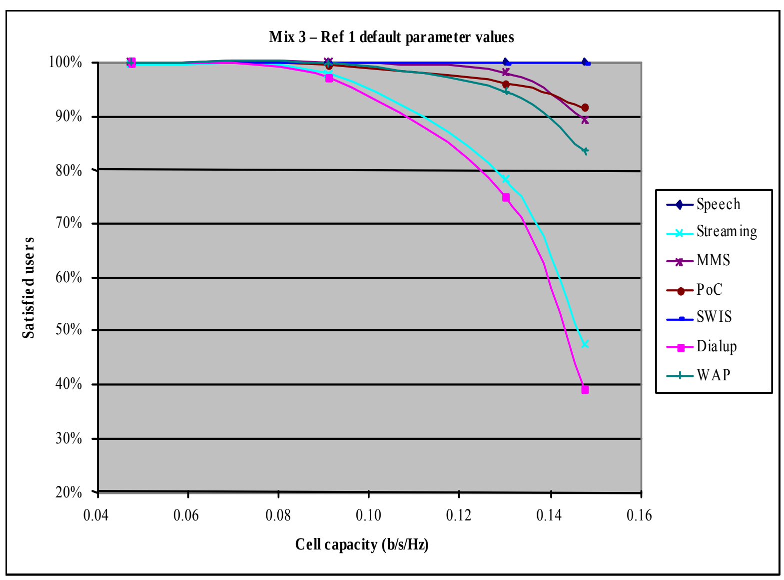
Figure 7.14. a) Mix 2 – "optimal" parameter values: Percentage of satisfied users for each of the analysed services as a function of the measured average cell throughput taken over all cells in the system; b) Mix 2 – Ref 1, Ref 2 and "optimal" parameter values: comparison in terms of the worst performing service for spectral efficiency gain computation.

Table 7.22. Mix 3: Percentage of satisfied users for each of the analysed services as a function of the measured average cell throughput taken over all cells in the system.

Parameters (Mix 3)	Av. Cell throughput (kb/s)	Speech	Video	Streaming	MMS	PoC	SWIS	Dialup	WAP
Ref 1	182	100%	0%	100%	100%	100%	100%	100%	100%
	349	100%	0%	98%	100%	100%	100%	97%	100%
	500	**100%**	**0%**	**78%**	**98%**	**96%**	**100%**	**75%**	**95%**
	567	100%	0%	48%	89%	92%	100%	39%	83%
Ref 2	201	100%	0%	99%	100%	100%	100%	100%	100%
	366	100%	0%	93%	99%	94%	100%	94%	98%
	498	**100%**	**0%**	**69%**	**94%**	**85%**	**100%**	**68%**	**87%**
	559	100%	0%	44%	83%	75%	100%	42%	73%
	567	100%	0%	30%	72%	66%	100%	24%	56%
"Optimal"	183	100%	0%	100%	100%	100%	100%	100%	100%
	367	100%	0%	98%	100%	98%	100%	96%	100%
	496	**100%**	**0%**	**80%**	**95%**	**94%**	**100%**	**76%**	**92%**
	565	100%	0%	56%	90%	89%	100%	49%	82%
	596	100%	0%	40%	82%	82%	100%	30%	68%

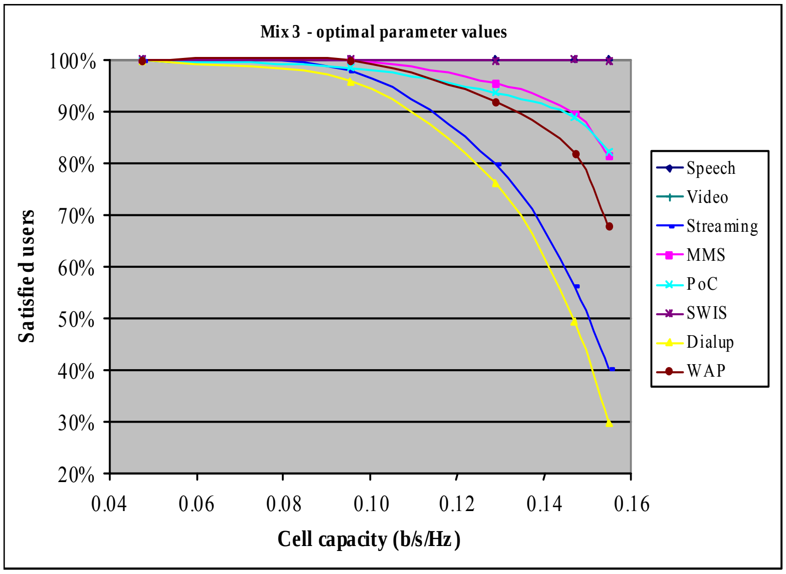


a)

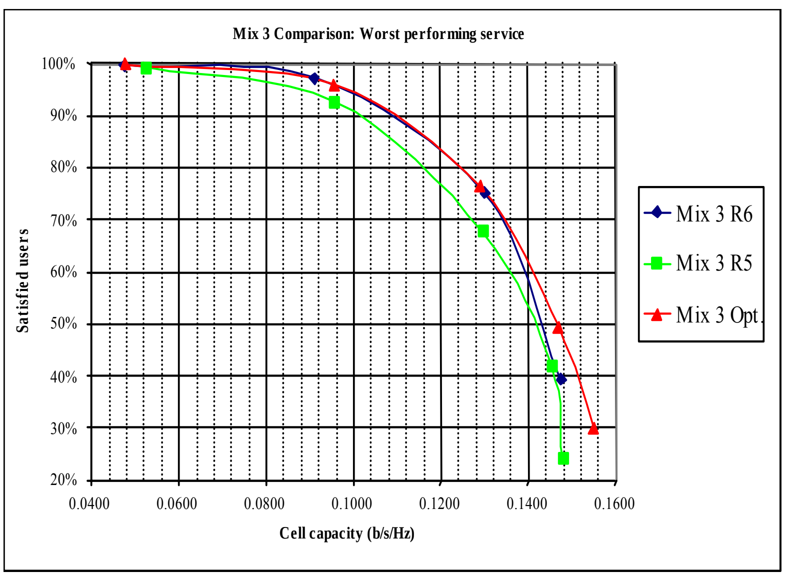


b)

Figure 7.15. Percentage of satisfied users for each of the analysed services as a function of the measured average cell throughput taken over all cells in the system: a) Mix 3, Ref 1 parameter values; b) Mix 3, Ref 2 parameter values.



a)

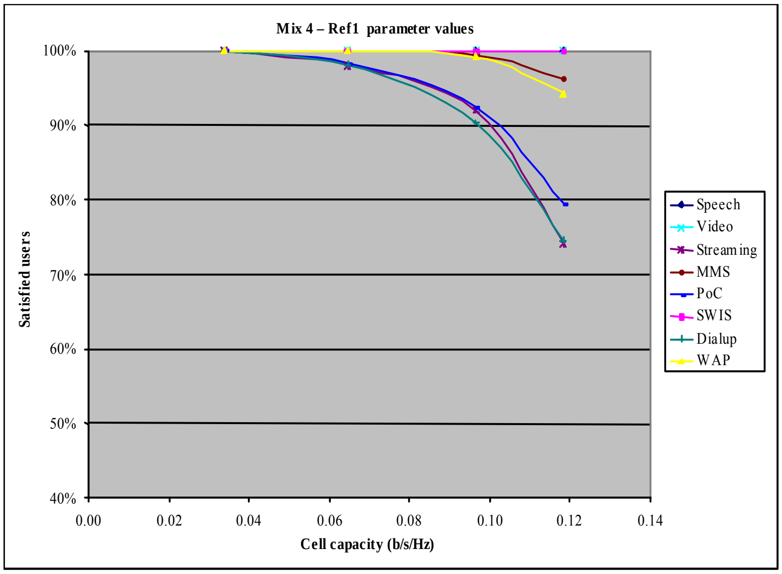


b)

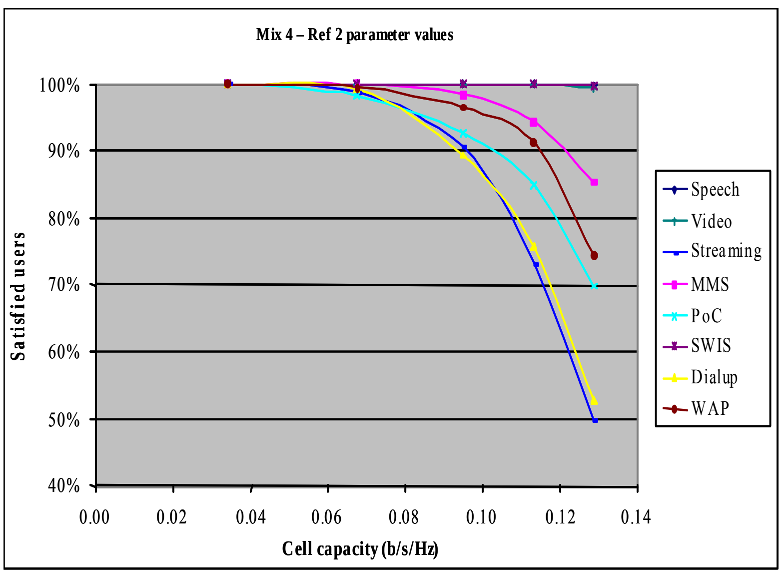
Figure 7.16. a) Mix 3 – "optimal" parameter values: Percentage of satisfied users for each of the analysed services as a function of the measured average cell throughput taken over all cells in the system; b) Mix 3 – Ref 1, Ref 2 and "optimal" parameter values: comparison in terms of the worst performing service for spectral efficiency gain computation.

Table 7.23. Mix 4: Percentage of satisfied users for each of the analysed services as a function of the measured average cell throughput taken over all cells in the system.

Parameters (Mix 4)	**Av. Cell throughput (kb/s)**	**Speech**	**Video**	**Streaming**	**MMS**	**PoC**	**SWIS**	**Dialup**	**WAP**
Ref 1 (1)	130	100%	100%	100%	100%	100%	100%	100%	100%
	248	100%	100%	98%	100%	98%	100%	98%	100%
	371	100%	100%	92%	100%	93%	100%	90%	99%
	455	**100%**	**100%**	**74%**	**96%**	**79%**	**100%**	**75%**	**94%**
Ref 2 (2)	130	100%	100%	100%	100%	100%	100%	100%	100%
	259	100%	100%	99%	100%	98%	100%	99%	100%
	364	100%	100%	91%	98%	93%	100%	89%	97%
	434	**100%**	**100%**	**73%**	**94%**	**85%**	**100%**	**76%**	**91%**
	494	100%	99%	50%	85%	70%	100%	53%	74%
"Optimal" (3)	118	100%	100%	100%	100%	100%	100%	100%	100%
	231	100%	100%	100%	100%	100%	100%	100%	100%
	371	100%	100%	93%	99%	97%	100%	93%	99%
	456	**100%**	**100%**	**82%**	**97%**	**93%**	**100%**	**83%**	**95%**
	516	100%	100%	56%	92%	87%	100%	55%	85%

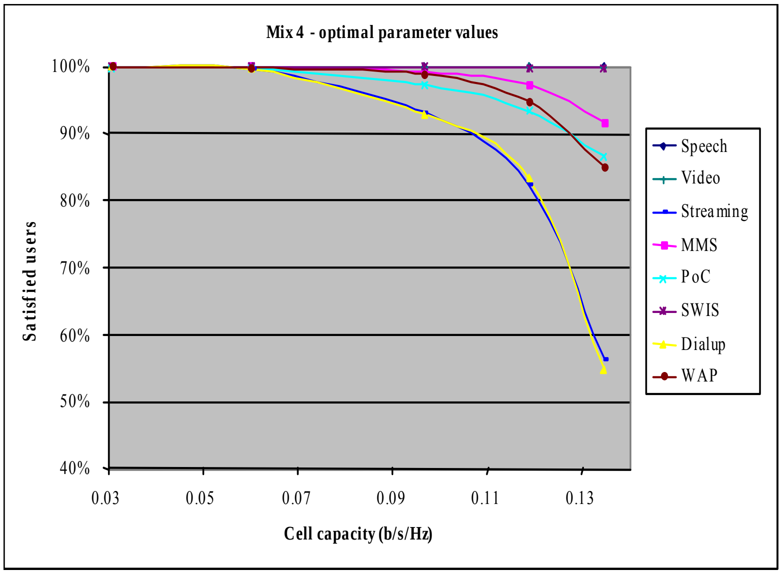


a)

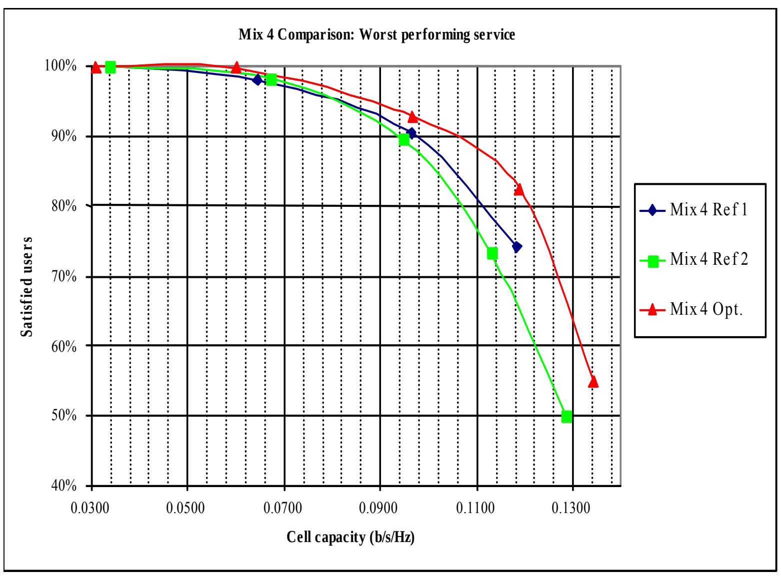


b)

Figure 7.17. Percentage of satisfied users for each of the analysed services as a function of the measured average cell throughput taken over all cells in the system: a) Mix 4, Ref 1 parameter values; b) Mix 4, Ref 2 parameter values.



a)

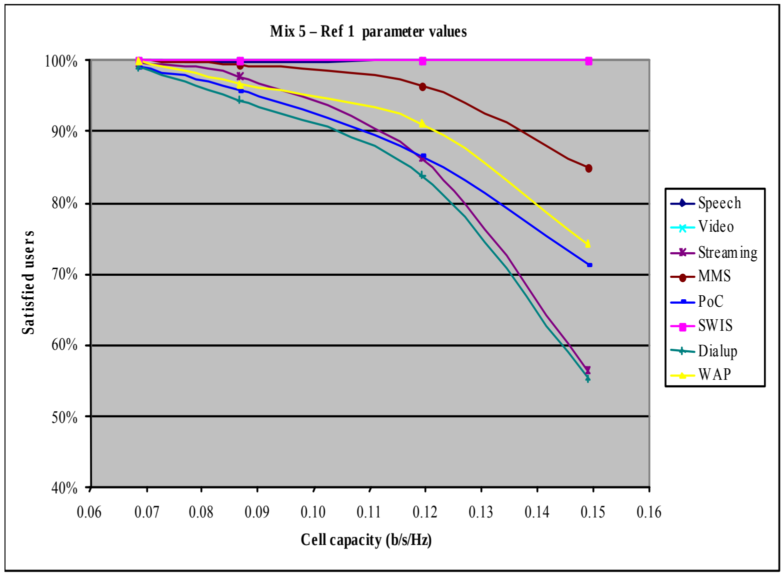


b)

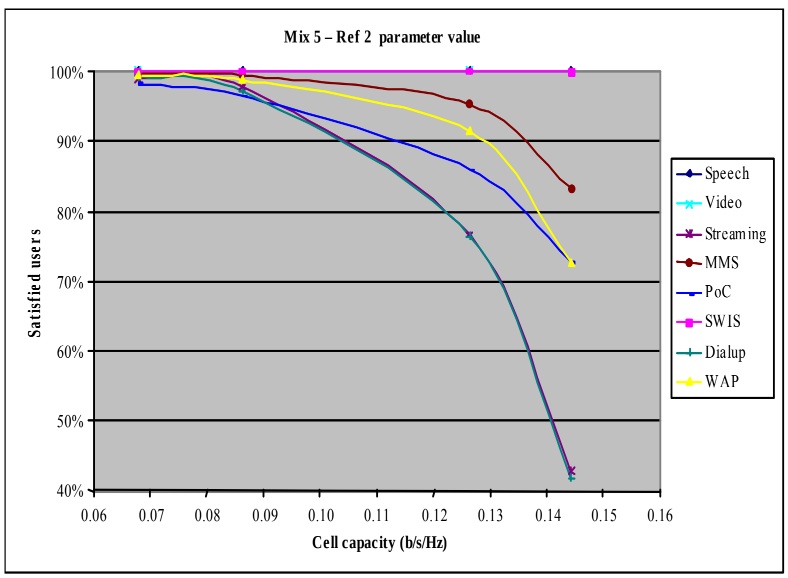
Figure 7.18. a) Mix 4 – "optimal" parameter values: Percentage of satisfied users for each of the analysed services as a function of the measured average cell throughput taken over all cells in the system; b) Mix 4 – Ref 1, Ref 2 and "optimal" parameter values: comparison in terms of the worst performing service for spectral efficiency gain computation.

Table 7.24. Mix 5: Percentage of satisfied users for each of the analysed services as a function of the measured average cell throughput taken over all cells in the system.

Parameters (Mix 5)	Av. Cell throughput (kb/s)	Speech	Video	Streaming	MMS	PoC	SWIS	Dialup	WAP
Ref 1	264	100%	100%	100%	100%	99%	100%	99%	100%
	333	100%	100%	98%	99%	96%	100%	94%	97%
	458	100%	100%	86%	96%	87%	100%	84%	91%
	572	100%	100%	56%	85%	71%	100%	55%	74%
Ref 2	260	100%	100%	99%	100%	98%	100%	99%	99%
	331	100%	100%	98%	99%	96%	100%	97%	99%
	485	100%	100%	77%	95%	86%	100%	76%	91%
	554	100%	100%	43%	83%	73%	100%	42%	73%
"Optimal"	257	100%	100%	100%	100%	100%	100%	100%	100%
	342	100%	100%	96%	100%	98%	100%	96%	99%
	478	100%	100%	85%	98%	94%	100%	83%	96%
	585	100%	100%	53%	90%	85%	100%	43%	79%

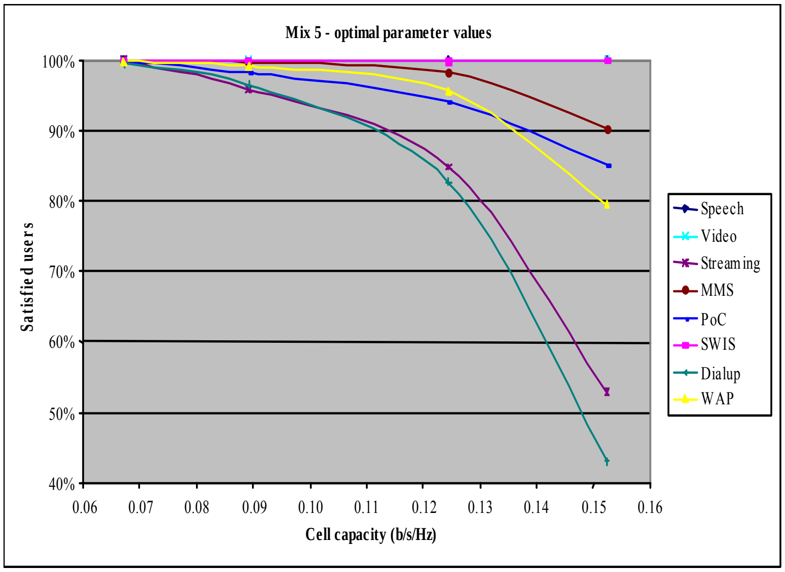


a)

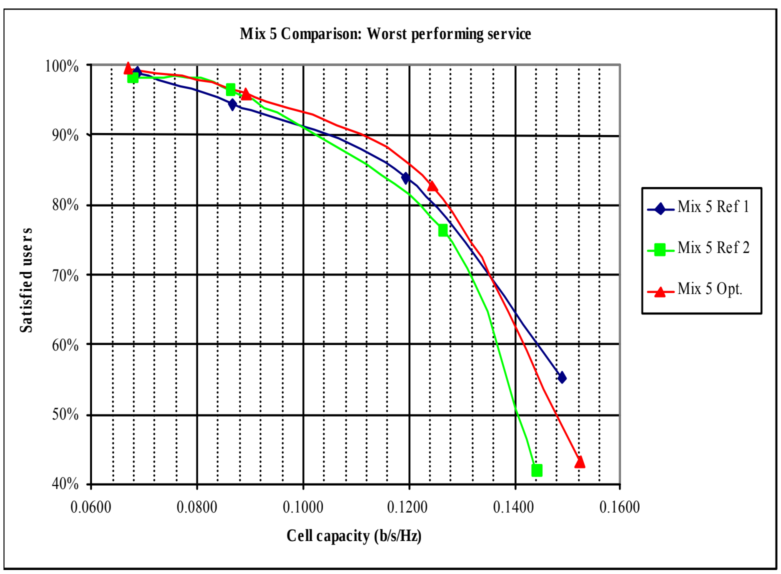


b)

Figure 7.19. Percentage of satisfied users for each of the analysed services as a function of the measured average cell throughput taken over all cells in the system: a) Mix 5, Ref 1 parameter values; b) Mix 5, Ref 2 parameter values.



a)



b)

Figure 7.20. a) Mix 5 – "optimal" parameter values: Percentage of satisfied users for each of the analysed services as a function of the measured average cell throughput taken over all cells in the system; b) Mix 5 – Ref 1, Ref 2 and "optimal" parameter values: comparison in terms of the worst performing service for spectral efficiency gain computation.

Table 7.25. Spectral efficiency gains for Mix 1-5: G_{21} = Gain provided by Ref 1 with respect to Ref 2; G_{32} = Gain provided by "optimal" parameter values with respect to Ref 1; G_{31} = Gain provided by "optimal" parameter values with respect to Ref 2. The minimal, maximal and average spectral efficiency gains taken over all traffic mixes.

Traffic	**Spectral efficiency (b/s/Hz)**			**Gains**		
	Ref 2 (1)	**Ref 1 (2)**	**"Optimal" (3)**	**G21 (%)**	**G32 (%)**	**G31 (%)**
Mix 1	0.0876	0.0922	0.0914	5.3	-0.9	4.3
Mix 2	0.0965	0.0945	0.109	-2.1	15.3	13.0
Mix 3	0.1012	0.1088	0.1098	7.5	0.9	8.5
Mix 4	0.0938	0.0975	0.10508	3.9	7.8	12.0
Mix 5	0.103	0.1046	0.1106	1.6	5.7	7.4
			Min	-2.1	-0.9	4.3
			Max	**7.5**	**15.3**	**13.0**
			Mean	3.2	5.8	9.0

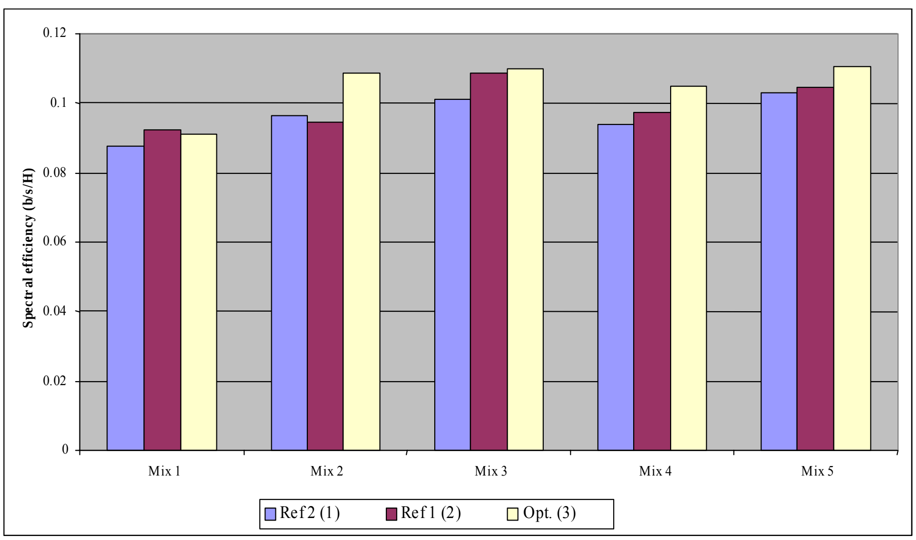


Figure 7.21. Spectral efficiency values for Mix 1-5 using Ref 1, Ref 2 and "optimal" parameter values.

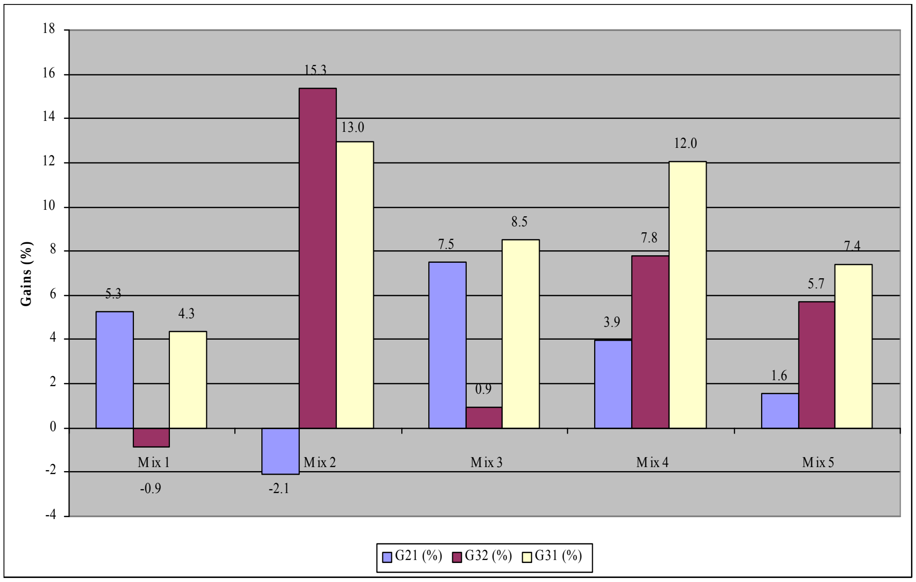


Figure 7.22. Spectral efficiency gains for Mix 1-5: G_{21} = Gain provided by Ref 1 with respect to Ref 2; G_{32} = Gain provided by "optimal" parameter values with respect to Ref 1; G_{31} = Gain provided by "optimal" parameter values with respect to Ref 2.

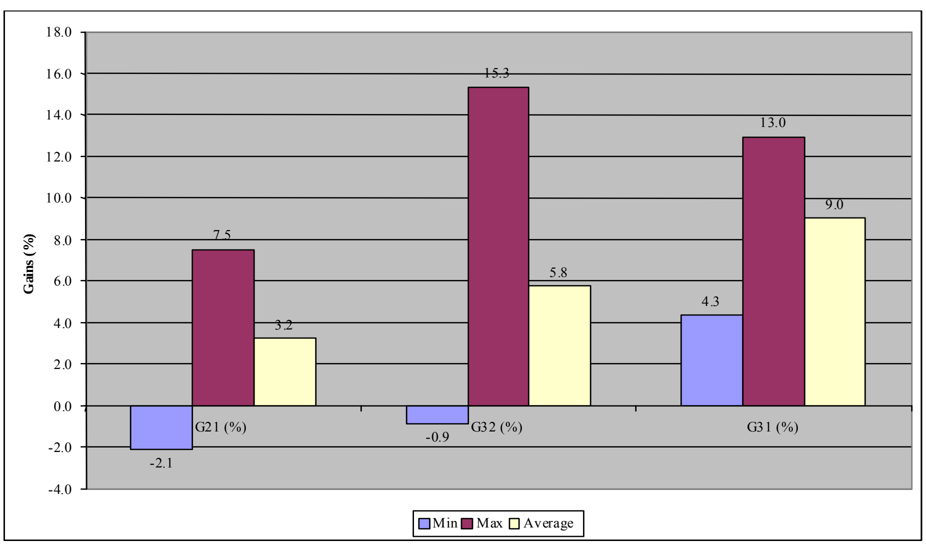


Figure 7.23. Minimal, maximal and average spectral efficiency gains taken over all traffic mixes: G_{21} = Gain provided by Ref 1 with respect to Ref 2; G_{32} = Gain provided by "optimal" parameter values with respect to Ref 1; G_{31} = Gain provided by "optimal" parameter values with respect to Ref 2.

7.6 Conclusions

Optimised service differentiation mostly benefitted intermediate traffic volumes, i.e. when around 20% of users of the worst performing service were unsatisfied. Simulation results showed clear needs for the network administrator to change parameter values according to input traffic mix, in order to maximise the services' performances. The genetic algorithm turned out to be an appropriate *research* method (solution space search algorithm) to tune simultaneously a number of parameters: 14 in this work. The "optimal" parameter settings were found after 20 to 30 iterations in all analysed traffic mix scenarios. Extensive simulation results showed that the spectral efficiency gain provided by "optimal" parameter values, with respect to the default differentiated settings, would be up to 15%; and the extra spectral efficiency gain provided by the "optimal" differentiated settings, with respect to undifferentiated QoS provisioning, would be up to 13%.

Although marginal in most of the investigated cases, the traffic volume played a role in the optimisation process. Higher numbers would be obtained if the multi-parameter tuning was performed at the same level of traffic as in the spectral efficiency computation and if the target thresholds were set differently taking into account the diverse QoS requirements. Simulation results showed business potential for service optimisation in UTRAN, whereas no adaptation gains are expected from undifferentiated parameter tunings.

As a remark, adoption of genetic algorithms may face difficulties in a network planning organisation due to the random character of optimisation. Experts may wish to understand the justification of parameter adjustments and to see good reasons why a specific action will not yield worse performance. Thus rule-based methods may be preferred to stochastic ones. The knowledge in the parameter control gained in this work (or, in general, using the proposed genetic solution) may be incorporated in the design of simple service optimisation methods that apply *fuzzy rules*. Basically, the expert may define specific rules (conditions and actions) as a response to specific performance measures. The fuzzy logic is a method for combining several performance measurements intuitively to give a signal for making a parameter adjustment.

About the research methodology adopted in this work, one may ask: What happens if a parameter setting is changed and/or if a new service is deployed in the network? How would that be parameterised? Would it affect other services' parameterisations? Would the results hold in other deployment scenarios? The proposed default parameter values (and mappings of services onto QoS profiles) form a good set of parameter settings for a service portfolio similar to the one simulated in this work. From the trends of the performance results (curves) attained in this work, and discussions thereof, it is possible to speculate upon the *qualitative* effects of changing the value of one of the crucial parameters or adding a novel service in the network. However, for a *quantitative* analysis, depending on the level of the desired accuracy, one of the tools proposed in this thesis may be used.

8

Conclusions and Further Research

8.1 Summary and Conclusions

In the present work, new bounds on the management aspects of QoS in UTRA FDD radio access networks were derived. The obtained results give more insight into the problems and limitations of handling differentiated services when applied to the WCDMA radio interface.

An overview of the UMTS architecture, interfaces, distributed functions, and QoS was given in Chapter 2. The content of this chapter was prevalently based on the R5 3GPP specifications. The gathered information constitutes essential background for the reader of this monograph.

Chapter 3 reports all models, main assumptions, and performance measures adopted in this work. The virtual time simulator, *Ares*, proposed for studying the provisioning of QoS in UTRAN, was also presented in this chapter. Besides this, plain methods and a simplified version of the above tool, *Max*, for radio interface dimensioning, were described. In this chapter, the performances obtained with Ares were compared to related results attained with the reference system level dynamic simulator, *Wallu*. The comparison results validated that Ares does quite well the task it was designed for: Absolute numbers resulted very close to each other, trends of the curves and performance results were consistent.

An effective QoS management framework with service differentiation was presented in Chapter 4. The proposed admission control, Fair Throughput (FT) and Fair Resources (FR) packet scheduling, and load control functions were analysed by varying the offered load and parameter settings using the models described in Chapter 3. Performance results were compared in terms of satisfied users and, at a given service quality, capacity gains were derived in terms of spectral efficiency. Simulation results showed the proposed algorithms with differentiated parameters to be a reliable solution to sustain the quality of real time services offered on best effort, i.e. without any resource reservation or bit rate guarantee. With respect to the non-prioritised scenario with undifferentiated parameter settings, the spectral efficiency gain due to the only service prioritisation was 30%, and the additional capacity gain when the services were also provisioned with differentiated parameters was 20%. The supplementary spectral efficiency gain provided by the FR scheduling depended heavily on the characteristics of existing bearer services and traffic mixes: The more traffic was carried with a wide range of bit rates, the more spectral efficient the FR algorithm was with respect to FT.

Chapter 5 described some use cases on how Max and Ares may be employed for service driven radio interface dimensioning and detailed radio network planning. Firstly, using Ares, the effectiveness of the supported QoS mechanisms was investigated. The utilisation of radio resources was analysed in terms of cell throughput and transmission powers. The QoE was determined separately for each of the offered services by tailored combinations of performance metrics. The simulation results showed the proposed simulator to be an appropriate tool for studying several aspects of QoS management, such as service planning and QoS provisioning. This tool was thus used to validate most of the content proposed in this work. As a part of this framework, the radio dimensioning issue that may arise from the introduction of novel services in WCDMA cellular networks was addressed in this chapter. The impact and entailed limitations of the deployment of new services on the performance of the offered ones were investigated using Max. For each of the accessible services, the corresponding user satisfaction criterion was defined, and performance results, at a given QoE, were presented as the maximum number of subscribers a cell can satisfactorily accommodate. Several case studies on the deployment of PoC (Push to talk Over Cellular) were analysed in this chapter. Simulation results showed that no extra capacity is required for PoC launch.

Chapter 6 aimed at providing some content for the operator to monitor and analyse the performance experienced by diverse packet flows across WCDMA radio access networks. This chapter presented theoretically simple formulas to estimate the most relevant integrity indicators for a reliable packed data communication, such as: Throughput, transfer delay, jitter, E_b/N_0, and data losses. The approach was based on an unambiguous mapping of service applications with different performance requirements onto distinct settings of the bearer attributes, e.g., bit rates, priorities and traffic classes. The proposed method makes it possible to formulate metrics to measure the performance of distinct service applications in the Access Stratum of UTRAN without any visibility of the content carried by upper layer protocols.

Chapter 7 presented a genetic approach to “optimal” QoS provisioning for UTRA FDD networks. Using a genetic algorithm, several parameters were changed simultaneously to maximise the percentage of satisfied users in different traffic scenarios. Gains provided by the “optimal” parameter settings with respect to the default network configurations were expressed in terms of spectral efficiency. Simulation results showed clear values for the network administrator to adopt different QoS provisioning for distinct input traffic mixes and the proposed genetic algorithm to be a feasible *research* method for multi-parameter optimisation. Practical aspects and generalisation of the results attained with the proposed research methodology were also discussed in this chapter.

Details on traffic models; cell based parameters; impact of transient time and statistical confidence on the simulation results; E_b/N_0, BLER and BER computation; as well as the pseudo code of the proposed genetic algorithm can be found in Appendix A, B, C and D, respectively.

8.2 Future Research

Possible directions for future research on the topic are now given. The results in this work were derived using approximate models for packet switched and circuit switched traffic and propagation channels, and conclusions were drawn by means of simulations. It could be interesting and useful for the operator, at least, to further explore the proposed methods for QoS monitoring and optimisation experimentally in the laboratory and by means of trials in WCDMA loaded networks. This would be useful (or better required) in order to take into account the actual implemented functions and compare bearer service performances to the particular characteristics and constraints of the real carried applications.

Almost all QoS and QoE aspects investigated in this work should be studied, formalised and validated for WCDMA Evolved Radio Access Networks, supporting the High Speed Downlink Packet Access (HSDPA) and/or High Speed Uplink Packet Access (HSUPA) concepts, within the 3GPP Release 5 and 6, and Release 7 frameworks, respectively. From the business perspective and time frame, higher priority should be given to HSDPA.

An overview of the HSDPA concept is given for example in [93]-[98]. To provide an evolution towards more sophisticated network and multimedia services, the main target of HSDPA is to increase user peak data rates, quality of service, and to generally improve the spectral efficiency for downlink asymmetrical and bursty packet data services. This is achieved by introducing a fast and complex channel control mechanism based on a short and fixed packet transmission time interval (TTI), adaptive modulation and coding (AMC), and fast physical layer (L1) hybrid ARQ. To facilitate fast scheduling with a per-TTI resolution in coherence with the instantaneous air interface load, the HSDPA-related MAC functionality is moved to the Node-B. The HSDPA concept facilitates peak data rates exceeding 2 Mb/s (theoretically up to and exceeding 10 Mb/s), and the cell throughput gain over previous UTRA-FDD releases has been evaluated to be in the order of 50-100% or even more, highly dependent on factors such as the radio environment and the service provision strategy of the network operator.

Hence, on these grounds, possible directions for further research on QoS management applied to HSDPA are:

- Evaluation and design of QoS based MAC-hs packet scheduling (with and without code multiplexing).
- Development of service and load aware dynamic transport channel type selection algorithms.
- Definition and validation of means and methods for HSDPA network deployment, i.e. service driven radio and transmission dimensioning and detailed network planning guides. What presented in [98] forms a good start.
- Identification of best parameter settings in MAC-hs flow control, handoffs, and “optimal” WCDMA radio interface resource utilisation.

For “optimal” mobility control, besides the genetic approach proposed in this work, would be interesting to take into account the role and importance of the economic aspects that are vital to the success of wireless services deployment and provider selection by users in a competitive environment. The econometric measures should meaningfully capture the users decisions and actions that should be utilised by the solution space search algorithms in changing parameter settings in the WCDMA evolved radio access networks. This could be accomplished by modelling the interactions by an operator and its customers as a non-cooperative game, where both users and the service provider may deploy a finite set of strategies. The “optimal” parameter settings should be the outcome of the game throughout a pure strategy Nash equilibrium. A similar approach was successfully adopted in [99]-[100] to achieve an integrated admission and rate control framework for CDMA data networks. The proposed solution significantly increased the provider’s revenues and also successfully offered differentiated QoS to the users.

Appendix A: Traffic Models

The distributions of the traffic models presented in Section 3.2 are defined as follows.

Truncated Log-Normal Distribution

$$p(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{\frac{-(\ln(x)-\mu)^2}{2\sigma^2}}, \tag{A.1}$$

where μ, σ and the range (cut-off) of the x values (min, max) are defined in Table 3.1.

Truncated Pareto Distribution

$$p(x) = \frac{\alpha k^{\alpha}}{x^{\alpha+1}}, \qquad x \geq k \tag{A.2}$$

where k, α and the range (cut-off) of the x values (min, max) are defined in Table 3.1.

Truncated Inverse Gaussian Distribution

$$p(x) = \sqrt{\lambda / 2x^3\pi} e^{\frac{-\lambda(x-\mu)^2}{2x\mu^2}}, \tag{A.3}$$

where μ, λ and the range (cut-off) of the x values (min, max) are defined in Table 3.1.

Truncated Exponential Distribution

$$p(x) = \lambda e^{-\lambda x}, \tag{A.4}$$

where λ and the range (cut-off) of the x values (min, max) are defined in Table 3.1.

Truncated Geometric Distribution

Directly from Matlab, defined as:

$$p(x / p) = \sum_{i=0}^{floor(x)} pq^i \ , \qquad \text{(A.5)}$$

where q = 1- p. Since the mathematical distribution starts from zero to reach the aimed mean p must be set as: p = 1/(1 + *mean*). The mean and range (cut-off) of the x values (min, max) are defined in Table 3.1.

Poisson Process for Session/Call Arrivals

Fix a time interval [0 t_{max}]. Generate random events with intensity *lambda* and let n be the total number of events in the interval. This number is Poisson distributed with mean value *lambda**t_{max}. The successive times between arrivals is Exp(*lambda*).
Example using Matlab (pseudo code):

```
tmax=100;                     % Simulation interval (s)
lambda=1/MeanArrivalRate;     % Arrival intensity (Hz)
arrtime=-log(rand)/lambda;    % Poisson arrivals
i=1;
  while (min(arrtime(i,:))<=tmax)
    arrtime = [arrtime; arrtime(i, :)-log(rand)/lambda];
    i=i+1;
  end
n=length(arrtime);             % arrival times t_1,...t_n
```

Using the above process, on average, a user makes a connection every *MeanArrivalRate* seconds. In this work, the *MeanArrivalRate* values depend on the case studies presented in Chapter 3, 4, 5 and 7.

Appendix B: Helsinki Scenario Cell Data

The below table reports a part of the cell based parameters used in the simulation scenario adopted in this work. The other parameter settings depend on the case studies presented in Chapter 3, 4, 5 and 7.

Table B.1. Cell based parameters for the Helsinki scenario with 19 cells (see Figure 3.2).

XPos (m)	YPos (m)	Ground Height (m)	Antenna Height (m)	Cell Tx Max. Power (dBm)	Link Tx Max. Power (dBm)	Link Min. Power (dBm)	Antenna Direction (Deg)	Antenna Tilt (Deg)	Cable Losses (dB)	Window Addition (dB)
385150	6675980	12	18	43	40	-999	130	7	3	-5
385150	6675980	12	18	43	40	-999	270	0	3	-5
385060	6675440	19	19	43	40	-999	280	7	3	-5
385060	6675440	19	19	43	40	-999	140	0	3	-5
385060	6675440	19	19	43	40	-999	20	7	3	-5
385050	6674340	9	17	43	40	-999	270	0	3	-5
385050	6674340	9	17	43	40	-999	30	0	3	-5
384512.5	6674987.5	3	20	43	40	-999	150	0	3	-5
384512.5	6674987.5	3	20	43	40	-999	30	0	3	-5
385900	6675462.5	5	17	43	40	-999	20	5	3	-5
385900	6675462.5	5	17	43	40	-999	140	0	3	-5
385900	6675462.5	5	17	43	40	-999	260	0	3	-5
386312.5	6674987.5	14	17	43	40	-999	270	0	3	-5
385525	6674987.5	10	19	43	40	-999	150	0	3	-5
385525	6674987.5	10	19	43	40	-999	40	0	3	-5
385525	6674987.5	10	19	43	40	-999	245	5	3	-5
384425	6675987.5	6	20	43	40	-999	150	0	3	-5
385862.5	6674362.5	11	20	43	40	-999	30	5	3	-5
385862.5	6674362.5	11	20	43	40	-999	270	5	3	-5

Appendix C: Statistical Confidence on Simulation Results

In this appendix the statistical confidence on the spectral efficiency computation in Chapter 4 and 7 is analysed. Firstly, the effect of the transient time on the simulation results for different system load levels is discussed. Secondly, the statistical confidence on the spectral efficiency gains is evaluated.

Impact of the Transient Time

The transient time (or warm up time) depends on the traffic mix and traffic volume (load) offered in the simulated cells, and varies cell-by-cell depending on the traffic distribution. In order to compute the spectral efficiency gains, at a given traffic mix, several traffic volumes were considered in this work, i.e. the mean call/session arrival rate ranged from 1s to 0.25s. The typical session/call arrival intensity at which the 90% of the worst performing service was satisfied was about 2Hz. For lower traffic volumes the computed percentages of satisfied users have no influence on the spectral efficiency computation. Hence, several simulations were run to estimate the transient time with an arrival intensity of one and two hertz, and an average over the different traffic mix (see Table 7.5) and simulated cells (see Figure 3.2) was taken. The obtained results are illustrated in Figure C.1.

As shown in the figure, with a mean call/session arrival rate of one second, the warm up time is about 200s. The transient time decreases not linearly when the traffic volume increases. Conservatively, in the following calculations, a transient time of 300s is assumed for a mean arrival rate of one second, and for higher traffic volumes the transient time is reduced proportionally to the corresponding arrival intensities.

To estimate the impact of the transient time on the attained results the following considerations can be made.

For a particular service type, let S_a be the *actual* user satisfaction ratio and S_c the *computed* one without taking out the calls/sessions arrived during the warm up period. The two ratios are defined as:

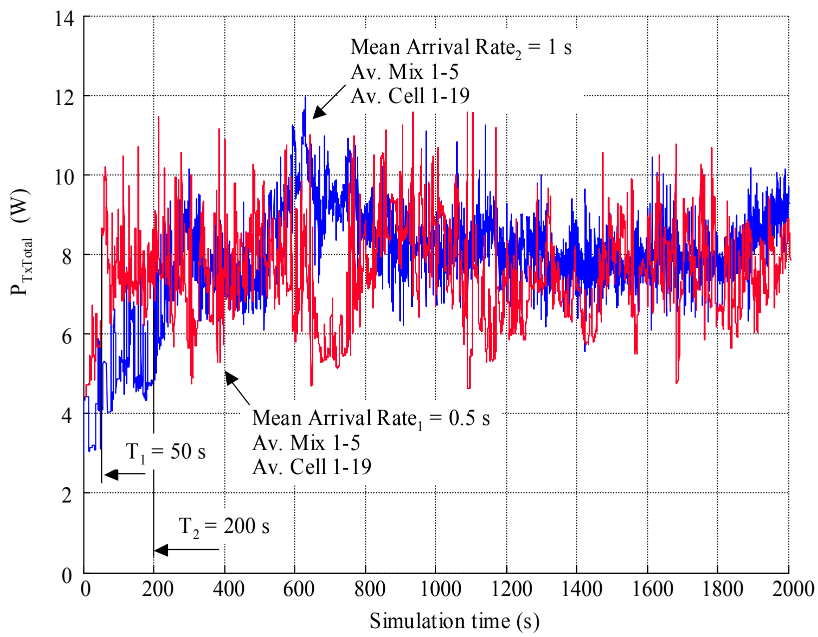


Figure C.1. Example of warm up time computation at two traffic volumes: One and two hertz call/session arrival intensity. The total base station transmission power ($P_{TxTotal}$ in the figure) is the mean value taken over all mixes and cells simulated in this work.

$$S_a = \frac{s_{tot} - s_{tr}}{u_{tot} - u_{tr}}, \tag{C.1}$$

$$S_c = \frac{s_{tot}}{u_{tot}}, \tag{C.2}$$

where s_{tot} are the total satisfied users during the simulated period, s_{tr} and u_{tr} are the satisfied users and arrivals during the transient time, respectively, and u_{tot} denotes the total number of arrivals in the system. Equation (C.1) can be rewritten as:

$$S_a = S_c \frac{1 - {s_{tr}}/{s_{tot}}}{1 - {u_{tr}}/{u_{tot}}} = S_c \frac{1 - {s_{tr}}/{0.9u_{tot}}}{1 - {s_{tr}}/{u_{tot}}}, \tag{C.3}$$

since $s_{tot} = 0.9u_{tot}$, by the definition of spectral efficiency adopted in this work (see Section 3.6), and $u_{tr} = s_{tr}$, bacause all users were satisfied during the transient times in this work (no outages due to coverage were registered). Equation (C.3) demonstrates that *all spectral efficiency gains attained in this work would be higher than the computed ones, if the statistics were collected removing the transient time*.

On average, s_{tr} and u_{tot} can be estimated from the simulated period (SP), transient time (TT), share of calls (SC) and mean arrival rates (MR), as follows:

$$u_{tot} = \frac{SP \cdot SC}{100 \cdot MR}, \tag{C.4}$$

$$s_{tr} = \frac{TT \cdot SC}{100 \cdot MR}, \tag{C.5}$$

which are valid for the *steady state*. Inserting (C.4) and (C.5) into (C.3) yields

$$S_t = S_c\left(1 + \frac{TT}{9(TT - SP)}\right), \tag{C.6}$$

which does not depend on the share of calls of that particular service type, but only on the overall traffic load. This means that, at a given traffic volume, for a particular traffic mix, the curves related to each service type are shifted proportionally downwards by the same quantity. The lower the traffic volume (the higher the mean arrival rate), the higher the shift of the curve towards lower values of the ordinate (y) axis. The offsets estimated using (C.6) for the traffic scenarios described in the two chapters are reported in Table C.1. In the worst case, the offset (shift of satisfied user ratios) is about 0.5%.

Table C.1. Estimated shifts (towards lower y values), at different traffic volumes, for the curves shown in the graphs of Chapter 4 and 7 and used for spectral efficiency computation.

		Measurement Period (s)		
		7200	5000	
Transient Time (s)	**Mean Arrival Rate (s)**	**Chapter 4: Shift (dB)**	**Chapter 7: Shift (dB)**	**Relative Variation (dB)**
300	1	2.103E-02	3.091E-02	9.880E-03
225	0.75	1.559E-02	2.280E-02	7.204E-03
150	0.5	1.028E-02	1.495E-02	4.671E-03
75	0.25	5.082E-03	7.355E-03	2.272E-03

Statistical Confidence on Spectral Efficiency Gains

Traditionally, people have expressed statistical analyses and simulation output in the form of confidence intervals (or confidence regions in the case of multivariate output). Compared to hypothesis tests, many people feel that confidence intervals are more informative and useful [101]. Both approaches are used in this work. The statistical hypothesis testing is introduced as data analysis stage of a comparative experiment in which two proportions of two populations are compared.

Confidence Interval

Suppose that we are sampling from a normal population with mean μ and variance σ^2. The assumption depends on the normal approximation to the binomial distribution discussed in Section 7.3.1.1. This requires the number of call/session arrivals in the system during the simulation period for each of the simulated services be greater than or equal to 50. This condition was always satisfied in this work. If $\bar{x}$ and s are the mean and standard deviation of a random sample from a normal distribution with unknown variance σ^2, then 100(1-α) percentage *confidence interval* on the population mean μ is given by [102]:

$$\bar{x} - t_{\alpha/2,n-1}\, s\big/\sqrt{n} \le \mu \le \bar{x} + t_{\alpha/2,n-1}\, s\big/\sqrt{n}\,, \tag{C.7}$$

where 1-α is the *confidence coefficient*, $t_{\alpha/2,\, n-1}$ is the upper $\alpha/2$ percentage point of the t distribution with n-1 degrees of freedom. The sample variance s and mean $\bar{x}$ are defined as:

$$s^2 = \sum_{i=1}^{n} \frac{\left(x_i - \bar{x}\right)^2}{n-1}\,, \tag{C.8}$$

$$\bar{x} = \sum_{i=1}^{n} \frac{x_i}{n}\,, \tag{C.9}$$

where x_i is the observation number i in a sample of size n.

Using α = 0.05, for a 95% confidence interval on the mean user satisfaction ratio, and n equal to 35 simulations with different seeds ($t_{0.025,\, 34}$ = 2.032), the above procedure, applied to a particular parameter settings and traffic scenario, leads to the following results.

The simulated network is illustrated in Figure 3.2, the traffic mix is reported in Table 4.1 and the parameters were set according to Case PDM1 described in Section 4.3. The traffic volume was regulated such that the percentage of satisfied users with the lowest QoE was close to 90% (for spectral efficiency computation). Conservatively, in order to generate in the system a relative low number of call attempts for that particular service type (Dialup in this case), traffic Mix 1 of Table 7.5 was selected and 5000s were simulated.

The attained satisfied users' ratios and corresponding average cell capacities (b/s/Hz) are reported in Table C.2. Scattered chart and discrete distributions thereof are depicted in Figure C.2, Figure C.3 and Figure C.4. Table C.3 collets the computed mean $\bar{x}$, standard deviation s, and two-sided 95% confidence intervals on the mean of satisfied users' ratios, using (C.7) and the performance results of Table C.2 as the sample of the population. From the analysed data we are 95% confident that, repeating the simulations, the deviations of the satisfied users' ratios, with respect to the ones displayed in Chapter 4 and 7, is less than 0.6%; and the deviation of the corresponding spectral efficiency values less than 0.066%.

Table C.2. Observations of satisfied users' ratios for each of the simulated services and corresponding cell capacities (b/s/Hz) in a sample of 35 simulations using different seeds. The observations x_1 and x_2 were used for the test of hypotheses of two proportions.

Sample	Speech	Streaming	MMS	PoC	SWIS	Dialup	WAP	Cell Capacity
x_1	0.999843	0.971300	0.991980	0.985391	0.998211	0.920000	0.987736	0.069650
x_2	0.999844	0.980400	0.990700	0.984218	1.000000	0.917200	0.979300	0.070322
x_3	0.999847	0.971496	0.983719	0.977403	0.998296	0.913100	0.979450	0.071514
x_4	0.999843	0.964010	0.993600	0.986903	1.000000	0.918919	0.984600	0.071336
x_5	1.000000	0.978049	0.993738	0.984538	1.000000	0.926871	0.987609	0.069960
x_6	0.999844	0.989390	0.989583	0.988364	0.998384	0.914614	0.979864	0.072276
x_7	0.999688	0.987593	0.988535	0.983435	1.000000	0.922500	0.977121	0.068780
x_8	0.999695	0.977500	0.990050	0.983678	1.000000	0.920400	0.978963	0.070480
x_9	0.999840	0.982800	0.993600	0.984409	1.000000	0.916900	0.978100	0.069777
x_{10}	1.000000	0.979000	0.992514	0.984698	1.000000	0.911864	0.981379	0.069900
x_{11}	0.999844	0.972800	0.996264	0.979809	1.000000	0.922500	0.991750	0.071027
x_{12}	1.000000	0.980200	0.986284	0.978898	1.000000	0.919008	0.972175	0.071938
x_{13}	0.999686	0.979900	0.994340	0.986541	1.000000	0.915700	0.984482	0.070733
x_{14}	0.999696	0.968059	0.985733	0.980241	1.000000	0.922414	0.978056	0.070856
x_{15}	1.000000	0.978100	0.990285	0.985926	1.000000	0.930700	0.986037	0.070246
x_{16}	0.999845	0.974900	0.987138	0.979271	1.000000	0.902609	0.977810	0.071062
x_{17}	0.999840	0.973333	0.984615	0.981468	1.000000	0.913876	0.968452	0.070290
x_{18}	0.999845	0.978700	0.989500	0.985100	1.000000	0.913559	0.978870	0.071281
x_{19}	0.999844	0.982323	0.985366	0.986732	0.998342	0.928814	0.975800	0.069370
x_{20}	0.999691	0.975200	0.993300	0.979278	1.000000	0.927817	0.973698	0.071630
x_{21}	0.999687	0.979600	0.995603	0.986549	1.000000	0.924800	0.985100	0.071158
x_{22}	0.999847	0.974900	0.988400	0.984900	1.000000	0.920700	0.979854	0.070920
x_{23}	0.999843	0.972200	0.989200	0.977969	1.000000	0.906700	0.971918	0.069966
x_{24}	1.000000	0.974300	0.998743	0.981900	1.000000	0.909370	0.990758	0.070570
x_{25}	1.000000	0.974000	0.988930	0.982649	0.998344	0.920400	0.987861	0.071575
x_{26}	1.000000	0.980000	0.992505	0.988408	1.000000	0.921000	0.987865	0.070644
x_{27}	1.000000	0.983700	0.997700	0.977037	0.998371	0.916300	0.994602	0.069890
x_{28}	0.999688	0.981500	0.989144	0.980720	1.000000	0.896714	0.980000	0.068976
x_{29}	0.999844	0.985000	0.996600	0.983400	1.000000	0.913700	0.984300	0.069172
x_{30}	0.999841	0.981432	0.992020	0.983281	0.998353	0.921000	0.984755	0.067690
x_{31}	0.999685	0.976247	0.995440	0.981873	0.998279	0.937400	0.983192	0.070096
x_{32}	1.000000	0.983200	0.991300	0.984300	1.000000	0.917800	0.981600	0.071159
x_{33}	0.999529	0.975962	0.992223	0.986197	0.998471	0.919275	0.981157	0.070800
x_{34}	1.000000	0.972400	0.995721	0.984859	1.000000	0.932874	0.985700	0.070432
x_{35}	0.999689	0.981500	0.995100	0.984144	1.000000	0.922200	0.984000	0.070192

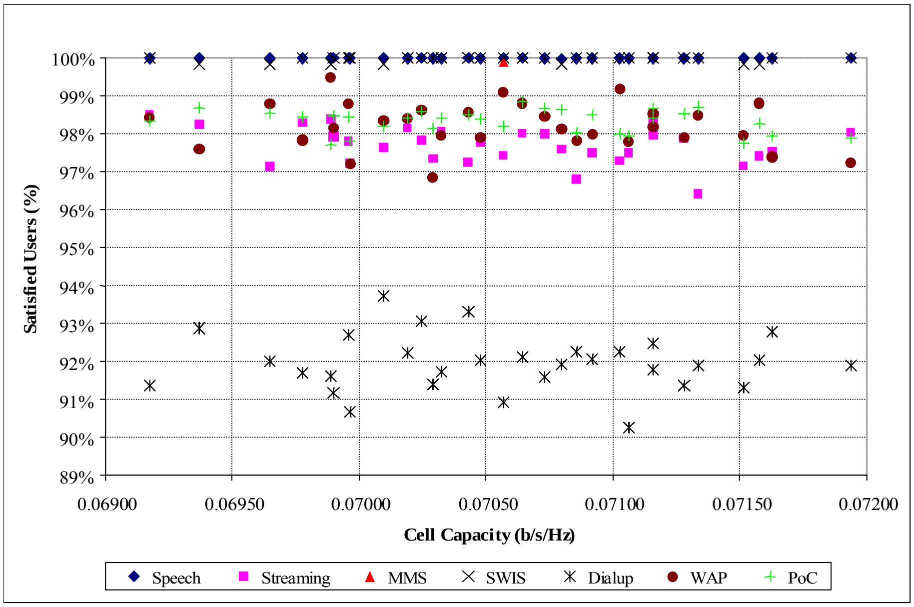


Figure C.2. Satisfied users ratios vs. normalised average cell throughput. The results of 35 simulations using different seeds are depicted. For each simulation, the average is taken over all cells of the simulated network. The normalisation factor is the chip rate (3.84 Mchip/s).

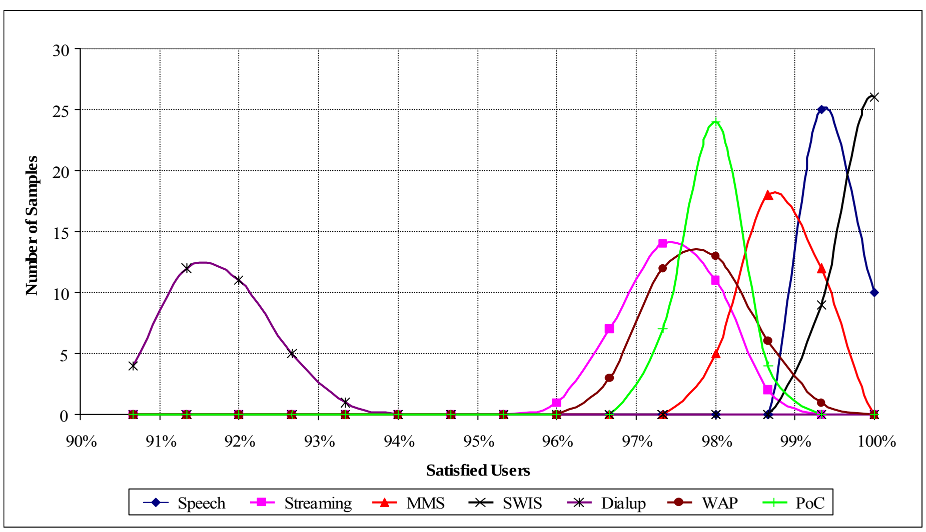


Figure C.3. Discrete distributions of satisfied users' ratios derived from the observations of 35 simulations using different seeds (see Table C.2).

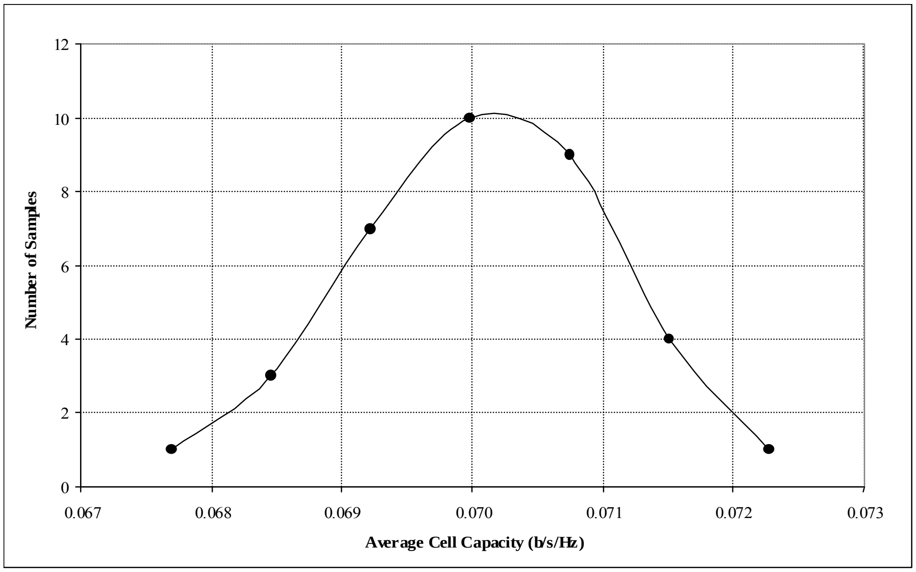


Figure C.4. Discrete distribution of the average cell capacity derived from the observations of 35 simulations using different seeds (see Table C.2).

Table C.3. Mean $\bar{x}$, standard deviation s, and 95% confidence intervals on the satisfied users' ratios attained with 35 simulations using different seeds, at a given traffic volume and parameter settings.

Measure	**Speech**	**Streaming**	**MMS**	**PoC**	**SWIS**	**Dialup**	**WAP**	**Cell Capacity**
Av. N. of Call Arrivals	6414	392	1604	2793	593	595	7593	-
$\bar{x}$	0.99984	0.97774	0.99141	0.98327	0.99957	0.91885	0.98183	0.07045
s (%)	0.01280%	0.53907%	0.38483%	0.30823%	0.07374%	0.80888%	0.57953%	0.09549%
95% Conf. Int.: t	**±0.004%**	**±0.185%**	**±0.132%**	**±0.106%**	**±0.025%**	**±0.278%**	**±0.199%**	**±0.033%**
95% Conf. Int.: $N(0,1)$	±0.004%	±0.179%	±0.127%	±0.102%	±0.024%	±0.268%	±0.192%	±0.032%

Test of Hypotheses on Two Proportions

A *statistical hypothesis* is a statement about the parameters of one or more populations. The assumption that a parameter θ of a population is equal to the corresponding parameter θ_0 taken from a sample of the population is denoted by H_0 and is called the *null hypothesis*. We wish to test the assumption $\theta = \theta_0$ against the assumption $\theta \neq \theta_0$. The assumption that

$\theta \neq \theta_0$ is denoted by H_1 and is called the *alternative hypothesis.* Because our decision is based on random variables, probabilities can be associated with two types of errors, denoted as:

- $\alpha = P(\text{type I error}) = P(\text{reject } H_0 \mid H_0 \text{ is true})$.
- $\beta = P(\text{type II error}) = P(\text{accept } H_0 \mid H_0 \text{ is false})$.

In the analysis, when the critical values are selected, it is usually easy to set the type I error probability at the desired value. Since the probability of wrongly reject H_0 can be directly controlled, we always think of rejection of the null hypothesis H_0 as a *strong conclusion*. The *power* of a statistical test is the probability of rejecting the null hypothesis H_0 when the alternative hypothesis is true. The power can be interpreted as the probability of correctly rejecting a false null hypothesis. The power is a very concise and descriptive measure of the sensitivity of a statistical test, where by sensitivity we mean the ability of the test to detect differences. More information on test of hypotheses can be found in [102] and [103].

To evaluate the statistical reliability of the simulation results, we can assume that there are two binomial parameter of interest, say, p_1 and p_2, and we wish to test that they are equal. That is, we wish to show that:

$$\begin{aligned} &H_0 : p_1 = p_2 \\ &H_1 : p_1 \neq p_2 \end{aligned}, \tag{C.10}$$

where H_0 and H_1 are respectively the null and alternative hypothesis.

Suppose that the two random samples of size n_1 and n_2 are taken from two populations, and let X_1 and X_2 represent the number of observations that belong to the class of interest in samples 1 and 2, respectively. Furthermore, suppose that the normal approximation of the population proportions $\hat{P}_1 = X_1/n_1$ and $\hat{P}_2 = X_2/n_2$, have approximated normal $N(0,1)$ distributions. Now, if H_0 is true, then using the fact that $p_1 = p_2 = p$, the random variable:

$$Z = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{p(1-p)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}, \tag{C.11}$$

is distributed approximately $N(0,1)$. An estimator of the common parameter p is

$$\hat{P} = \frac{X_1 + X_2}{n_1 + n_2}. \tag{C.12}$$

The *test statistic* for H_0: $p_1 = p_2$ is then

$$Z_0 = \frac{\hat{P}_1 - \hat{P}_2}{\sqrt{\hat{P}(1-\hat{P})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}. \tag{C.13}$$

Let z_0 be the computed value of the test statistic. Then if

$$z_0 > z_{\alpha/2} \quad \text{or} \quad z_0 < -z_{\alpha/2}, \tag{C.14}$$

the null hypothesis is rejected.

(Note: In this case, the test hypotheses on the mean of a normal distribution where the variance is unknown cannot be applied, since the mean value of the population of satisfied users ratios, at the different traffic volumes, is not known a priori. Hence, the test of the two-sided alternative hypothesis (H_0: $\mu = \mu_0$; H_1: $\mu \neq \mu_0$) cannot be formulated for all analysed services.)

On these grounds, two different simulations (using different seeds) are being evaluated. The performance results for each of the deployed services are shown in Table C.2. For the service i, let $x_1(i)$ and $x_2(i)$, first two observations, be the approximations of the population proportions, p_1 and p_2. Is there any reason to believe that the two simulation results differ?

Using $\alpha = 0.01$ ($z_{0.005} = 2.58$), the above hypotheses procedure leads to the following results. For each service i, the parameters of interest are the two satisfaction ratios, $x_1(i)$ and $x_2(i)$. We wish to show that the null hypothesis and the alternative hypothesis are as defined in (C.10).

Table C.4 reports the values of the test statistics obtained using (C.13) and x_1, x_2, from Table C.2. Using (C.14), as shown in Table C.4, there is a strong evidence to support the claim that the simulations, if repeated, would lead to the same spectral efficiency gains, since for six out of seven services the null hypothesis H_0 is accepted.

Table C.4. Test of hypotheses on two proportions using the simulation results of Table C.2, and x_1 and x_2 as approximations of the population proportions, p_1 and p_2.

Test of Hypotheses (Type I error = α)	**Speech**	**Streaming**	**MMS**	**PoC**	**SWIS**	**Dialup**	**WAP**
$p_1 = x_1$	99.9843%	97.1300%	99.1980%	98.5391%	99.8211%	92.0000%	98.7736%
$p_2 = x_2$	99.9844%	98.0400%	99.0700%	98.4218%	100.0000%	91.7200%	97.9300%
$q_1 = 1 - p_1$	0.0157%	2.8700%	0.8020%	1.4609%	0.1789%	8.0000%	1.2264%
$q_2 = 1 - p_2$	0.0156%	1.9600%	0.9300%	1.5782%	0.0000%	8.2800%	2.0700%
n_1	6387	385	1621	2875	571	584	7665
n_2	6405	375	1572	2788	572	587	7612
s = Std($p_1 - p_2$) [100*sqrt($p_1q_1/n_1+p_2q_2/n_2$)]	0.02%	1.11%	0.33%	0.33%	0.18%	1.60%	0.21%
err (%) [100*(p_1-p_2)/p_1]	0.000	-0.937	0.129	0.119	-0.179	0.304	0.854
Test Statistic (z_0)	0.00	-2.48	1.45	2.37	-0.72	1.21	45.95
Conclusions [H_0: $p_1 = p_2$]	**Accept**	**Accept**	**Accept**	**Accept**	**Accept**	**Accept**	Reject

Appendix D: UL BLER, BER and E_b/N_0 Derivation

This appendix presents a plain method for determining performance metrics, such as E_b/N_0, BLER (Block Error Rate) and BER (Bit Error Rate), associated to a selected active uplink transport channel (TrCH) multiplexed with at least one other transport channel to a dedicated physical channel (DPCH) for an effective usage of available resources.

During a radio resource control (RRC) connection several TrCHs may be multiplexed in the uplink by a terminal onto one code composite transport channel (CCTrCH), which is subsequently mapped to the data part of a dedicated physical channel (DPCH) [71]. The DPCH onto which the CCTrCH is mapped consists of a dedicated physical control channel (DPCCH) and one or more dedicated physical data channels (DPDCH). The DPDCHs comprise the user data of the TrCHs. For the different DPDCHs different spreading codes are used, which enables a simultaneous WCDMA transmission. The DPCCH comprises predefined pilot bits to support channel estimation for coherent detection, transmit power control (TPC) commands and a transport format combination indicator (TFCI). DPCCH and DPDCHs are I/Q multiplexed within each radio frame with complex scrambling.

When the connection is set up, the terminal is given the minimum allowed SF of the uplink channelisation code for the data part and the variable rate is handled in a dynamic rate matching by changing the DPDCH bit rate (SF) frame by frame or by repeating or puncturing the code bits to achieve the total physical channel bit rate. In the uplink direction repetition is preferred. Puncturing is used to avoid multicode transmission or when facing the limitations of the user equipment (UE) transmitter or Node B receiver. The relative power level is set such that for higher bit rates the power of the DPCCH is higher, thus enabling a more accurate channel estimation, and the overhead, i.e. DPDCH vs. DPCCH power, of the DPCCH is still lower.

Since several TrCHs may be multiplexed onto one CCTrCH, and since the quality, and in particular the (bit energy per noise interference) requirements can be different in each transport channel, a rate-matching attribute is introduced for each TrCH, which is used in balancing the E_b/N_0 values of the radio connection for a static or service specific rate matching. According to [71], higher layers assign a rate-matching attribute for each transport channel. The rate-matching attribute is used to calculate a rate matching value when multiplexing several TrCHs for the same frame. The attribute is semi-static and can only be changed through higher layer signalling. By adjusting the rate-matching attribute, an admission control (AC) is thus able to fine-tune the quality of different bearer services in order to reach an equal or near equal symbol power level requirement. Success of the static rate matching depends on, e.g., how correct the E_b/N_0 value is in respect to the quality - e.g. BLER - target. With the aid of the rate matching attribute and the TFCI of the DPCCH, the receiver can calculate backwards the rate matching parameters used and perform the inverse operation. Static and dynamic rate matching are done simultaneously.

Neither the E_b/N_0 nor the quality of any TrCH set up during the same RRC connection is available from the conventional implementations. In order to overcome the effect of the possible incomplete static rate matching in the measured quantities, it would be an advantage, if uplink connection based measurements could be carried out for each bearer service or TrCH multiplexed onto one CCTrCH. E_b/N_0 is also in general a key figure for the receiver performance and can be employed for a variety of functions, as well in controlling as in evaluation. Moreover, other performance indicators associated to specific TrCHs, in particular quality parameters like BER and/or BLER, can be of interest, which parameters are only available for entire DPCHs in the current specifications.

This appendix enables the determination of cell-based values for the E_b/N_0, BLER and BER associated to a selected TrCH multiplexed with more TrCHs to a DPCH based on the available target SIR in the RNC, which is influenced by signals transmitted on all active TrCHs multiplexed to the DPCH. This is possible by computing the metric in question only during predetermined periods of time when any other than the selected TrCH was inactive. The proposed method proceeds from the idea that even if several TrCHs are multiplexed to a DPCH, only the respective active ones are able to influence the target SIR determined for the entire DPCH. Inactive transport channels are in particular transport channels currently used for downlink transmissions. Therefore, it is proposed how to control the activity of all TrCHs multiplexed to the considered DPCH. Thereby, for respective periods of time, it is known if only the selected TrCH influences the target SIR associated to the DPCH, and only these target SIR values are then accepted as basis for determining the desired transport channel specific performance values. This means, in case other TrCHs than the selected one were active during a period of time, the E_b/N_0 and quality values are either not determined in the first place for this period, or determined but then discarded.

This method makes use of the uplink outer loop power control (OLPC) controller of the serving RNC in UTRAN, which is presented in the following. The activity of each TrCH mapped onto a DPCH is controlled in a dedicated uplink OLPC entity of the SRNC. The uplink OLPC entities can then provide (for each of the periods of time) an indication to the OLPC controller whether the respective TrCH was active during this period of time.

The E_b/N_0 and quality values of the selected active uplink transport channels are determined for each management class, cell by cell for a selected group of cells, and in particular for cells involved in soft handover.

Uplink outer loop power control

Figure D.1 shows the elements in UTRAN involved in determining the metrics to assess the performance of a selected active TrCH and the conventional logical architecture of an uplink outer loop PC algorithm in the RNC.

In the RNC, a macro diversity combiner (MDC) is provided for each bearer service within a single RRC connection. The MDCs enable the connection of the Node B with the RNC via an I_{ub} interface. Each MDC is moreover connected within the RNC with a dedicated uplink outer loop power control (OLPC) entity provided for the same data channel as the respective MDC. The OLPC entities are further bi-directionally connected to an uplink (OLPC) controller provided for one RRC connection. This OLPC controller receives in addition input from a load control (LC) function and from an admission control (AC) function. The AC has in addition a direct access to the Node B.

At the setup of the radio access bearer RAB, the AC computes an initial target SIR after rate matching when multiplexing several TrCHs for the same frame. The AC provides the Node B directly with this computed initial SIR target value for reception from the respective terminal. Additionally, the AC provides the uplink OLPC controller with the initial target SIR and with other configuration parameters. A part of these parameters are then forwarded to the OLPC entities. The MDCs receive uplink quality information from the Node B and combine the incoming data from different SHO (soft handover) branches in a selection and combining procedure. Each uplink OLPC entity then receives the processed uplink quality information from the associated MDC. This quality information includes, depending on the type of the radio bearer, either a BLER estimate computed in the MDCs based on the CRC-bits of the selected frame, and/or a BER estimate calculated in the WCDMA Node B. If the CRC is not correct (NOK), the respective MDC selects the best one of the BER estimates. Each OLPC entity calculates for the respective channel a required change in the target SIR according to the received uplink quality information and provides the calculated required change to the OLPC controller up to once each TTI. Equally provided to the OLPC controller is an activity report for each reporting periods. The duration of a reporting period is provided as one of the parameters from the AC via the OLPC controller. Based on the current target SIR and the different change requests, the OLPC controller computes a new target SIR. The new target SIR is reported together with PC parameters to the OLPC entities. One of the entities, in particular an entity assigned to a signalling link or to a dedicated control channel DCCH, is selected to transmit the new target SIR via the respective MDC to the Node B. For interactions between the RNC and the Node B, the DCH frame protocol (FP) is used.

Figure D.2 illustrates in more detail an algorithm that can be used by the OLPC controller and the OLPC entities for determining a new target SIR. The lower box indicates the calculations carried out by the OLPC entities, and the box above indicates the calculations carried out by the OLPC controller.

Each active or semi active OLPC entity may contribute with a change request to the new SIR target computation in any TTI, provided that the requested change is greater than 0.1 dB. The respective desired change Δ is calculated either based on a BLER or a BER estimate provided by the respective MDC. When based on a BLER estimate, the desired change is calculated as the difference between the received BLER estimate and a target BLER, multiplied by a step size. The BLER estimate is calculated as the quotient of the number of CRCs which are not OK (CRC NOK) and the sum of the CRCs which are OK (CRC OK) and of the CRCs which are not OK, which sum corresponds to the number of transport blocks on a transport block set (TBS). When based on a BER estimate, the desired change is calculated as the difference between a received BER estimate of a first itineration and a target BER, multiplied by a step size. The step size is set by radio network planning

(RNP) and corrected by each entity considering the interleaving depth of the respective transport channel. The OLPC controller receives the changes requested by each OLPC entity, and determines which requests are allowed to contribute to the new target SIR. The new target SIR is then calculated by the OLPC controller as the old SIR target plus the sum over all valid requested changes. In the corresponding equation included in the lower box in Figure D.2, k is the number of OLPC entities contributing to the new SIR target computation, i.e. the number of all active and semi-active entities allowed to send a request for a change of the SIR to the OLPC controller. The first old SIR target after a RAB setup is the initial SIR target received from the AC. The new SIR target is then forwarded to BTS (base transceiver station) or Node B via one of the OLPC entities as mentioned with reference to Figure D.1.

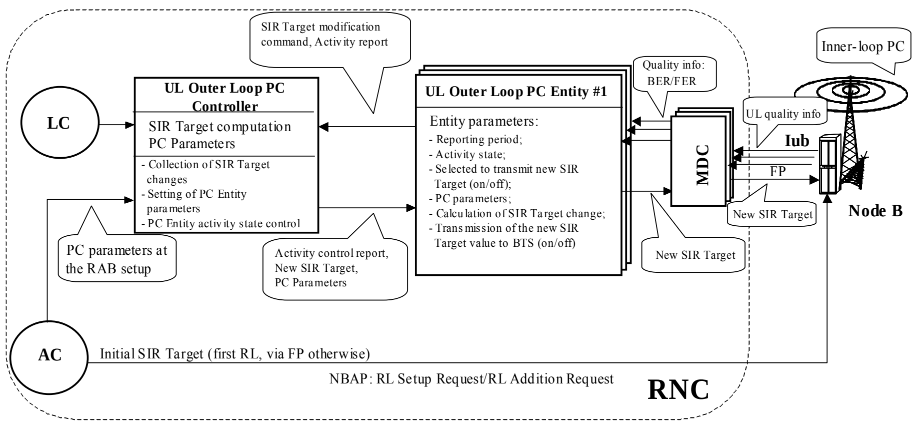


Figure D.1. Logical architecture of an uplink outer loop PC algorithm employed in this work.

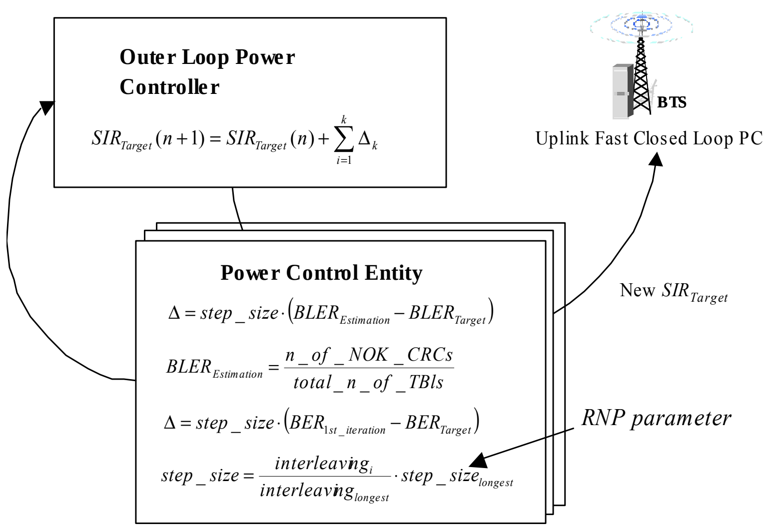


Figure D.2. Conventional uplink power control algorithm.

Uplink E_b/N_0, BLER and BER derivation

The desired E_b/N_0 for each transport channel can be put into relation to the determined target SIR according to the following considerations.

The Node B detects physical channel pilot symbols and estimates the SIR per symbol on the DPCCH in correspondence with [88] as:

$$SIR_{DPCCH} = SF_{DPCCH} \frac{RSCP_{DPCCH}}{ISCP}, \tag{D.1}$$

with SF_{DPCCH} = W/R_{DPCCH}, where R_{DPCCH} is the bit rate of the DPCCH, W the chip rate, $RSCP_{DPCCH}$ is the received signal code power on the DPCCH, and with $ISCP$ being the uplink interference signal code power. SF_{DPCCH} is always 256 in the uplink direction. Based on this observation, and the fact that,

$$ISCP = WN_0, \tag{D.2}$$

where N_0 is the interfering noise spectral density, the SIR estimate by the node B can be reformulated as follows:

$$SIR_{DPCCH} = SF_{DPCCH} \frac{RSCP_{DPCCH}}{RSCP_{DPDCH}} \frac{RSCP_{DPDCH}}{WN_0}. \tag{D.3}$$

The uplink DPCCH overhead in terms of power may be expressed as a function of the 3GPP specified *gain factors* (amplitude offsets) β_c and β_d, that is:

$$\frac{RSCP_{DPCCH}}{RSCP_{DPDCH}} = \left(\frac{\beta_{c,TFC_{Max}}}{\beta_{d,TFC_{Max}}} \right)^2 = r^2. \tag{D.4}$$

Substituting (D.4) into (D.3) leads to:

$$SIR_{DPCCH} = SF_{DPCCH} \frac{RSCP_{DPDCH}}{WN_0} r^2. \tag{D.5}$$

The energy per bit for a specific dedicated channel can be expressed as:

$$E_{b,DCH} = \frac{E_S \left(N + r^2 \right)}{N R_{DCH}^{C,M}}. \tag{D.6}$$

Where E_s is the energy of the DPDCH component of a WCDMA modulation symbol, N the number of the employed DPDCHs, r^2 the overhead of the DPCCH, and $R_{DCH}^{C,M}$ is the layer 1 encoding rate after coding and rate matching. $R_{DCH}^{C,M}$ may be quantified as follows:

$$R_{DCH}^{C,M} = \frac{N_{DCH}^{U}}{N_{DCH}^{C,RM}}, \tag{D.7}$$

where N_{DCH}^{U} is the number of bits per radio frame prior to CRC attachment, i.e. at the top of layer 1, and $N_{DCH}^{C,RM}$ is the number of bits per radio frame after layer 1 rate matching. The rate-matching algorithm defined in [71] imposes that:

$$N_{DCH}^{C,RM} = N \frac{RM_{DCH} N_{DCH}^{C}}{\sum\limits_{DCH \in RL} RM_{DCH} N_{DCH}^{C}} N_{SF}, \tag{D.8}$$

where RM_{DCH} is the rate matching attribute for a specific dedicated channel, N_{DCH}^{C} is the number of bits per radio frame subsequent to radio frame equalisation, i.e. prior to rate matching, and N_{SF} is the total number of DPDCH bits available per radio frame for a specific spreading factor. Re-arranging (D.6) and substituting (D.8) leads to:

$$E_S = \frac{E_{b,DCH} N_{DCH}^{U}}{\left(N + r^2\right) \frac{RM_{DCH} N_{DCH}^{C}}{\sum\limits_{DCH \in RL} RM_{DCH} N_{DCH}^{C}} N_{SF}}. \tag{D.9}$$

Now considering the fact that the received signal code power of a DPDCH may be expressed as:

$$RSCP_{DPDCH} = E_S R_{DPDCH}, \tag{D.10}$$

where E_S is the energy of the DPDCH component of a WCDMA modulation symbol and R_{DPDCH} is the corresponding symbol rate, substituting (D.9) into (D.10) yields:

$$RSCP_{DPDCH} = \frac{E_{b,DCH} N_{DCH}^{U} R_{DPDCH}}{\left(N + r^2\right) \frac{RM_{DCH} N_{DCH}^{C}}{\sum\limits_{DCH \in RL} RM_{DCH} N_{DCH}^{C}} N_{SF}}. \tag{D.11}$$

Hence, (D.5) can be re-arranged as:

$$SIR_{DPCCH} = SF_{DPCCH} \frac{E_{b,DCH}}{N_0} \frac{r^2}{N + r^2} \frac{N_{DCH}^{U} R_{DPDCH}}{W N_{SF}} \frac{\sum\limits_{DCH \in RL} RM_{DCH} N_{DCH}^{C}}{RM_{DCH} N_{DCH}^{C}}, \tag{D.12}$$

which can be simplified by considering the fact that:

$$\frac{N_{DCH}^{U}}{N_{SF}} = \frac{R_{DCH}}{R_{DPDCH}}, \tag{D.13}$$

where N_{DCH}^{U} is the number of bits per radio frame prior to CRC attachment, R_{DCH} is the bit rate prior to CRC attachment, N_{SF} is the total number of DPDCH bits available per radio frame for one code channel, R_{DPDCH} is the DPDCH symbol rate, and N is the number of codes employed in the uplink transmission. Re-arranging (D.13) and substituting within (D.12) leads to:

$$SIR_{DPCCH} = SF_{DPCCH} \frac{E_{b,DCH}}{N_0} \frac{r^2}{N + r^2} \frac{R_{DCH}}{W} \frac{\sum_{DCH \in RL} RM_{DCH} N_{DCH}^C}{RM_{DCH} N_{DCH}^C} . \tag{D.14}$$

Re-arranging for E_b/N_0, and changing to logarithmic units leads:

$$\frac{E_{b,DCH}}{N_0} = SIR_{DPCCH} - 20\log(r) - 20\log\left(\frac{R_{DCH}}{R_{DPDCH}(N + r^2)}\right) - 10\log\left(\frac{\sum_{DCH \in RL} RM_{DCH} N_{DCH}^C}{RM_{DCH} N_{DCH}^C}\right) - 10\log\left(\frac{SF_{DPCCH}}{SF_{DPDCH}}\right) . \tag{D.15}$$

Equation (D.15) defines the calculation of the E_b/N_0 requirement for a specific transport channel. The composite E_b/N_0 figure for all transport channels belonging to a physical channel can be obtained by summing the individual E_b/N_0 values.

This leads to the conclusion that when during the current RRC connection either only one DCH is mapped onto the DPDCH or only a specific bearer service is active, i.e. all other bearer services are not in uplink transmission, the actual E_b/N_0 target reduces to:

$$\frac{E_{b,DCH}}{N_0} = SIR_{DPCCH} - 20\log(r) - 20\log\left(\frac{R_{DCH}}{R_{DPDCH}(N + r^2)}\right) - 10\log\left(\frac{SF_{DPCCH}}{SF_{DPDCH}}\right) . \tag{D.16}$$

In the case of HSDPA, (D.16) needs to take into account the overhead introduced by the UL HS-DPCCH. Let Δ be the power offset between the uplink associated DPCCH and the HS-DPCCH [89], the energy per bit for a specific dedicated channel can be expressed as:

$$E_{b,HSDPA} = \frac{E_S (N + r^2)\left(1 + \frac{r^2}{(N + r^2)}\Delta\right)}{NR_{DCH}^{C,M}} = E_{b,DCH}\left(1 + \frac{r^2}{(N + r^2)}\Delta\right) . \tag{D.17}$$

Hence, (D.16) for HSDPA becomes:

$$\frac{E_{b,HSDPA}}{N_0} = SIR_{DPCCH} - 20\log(r) - 20\log\left(\frac{R_{DCH}}{R_{DPDCH}(N + r^2)}\right) - 10\log\left(\frac{SF_{DPCCH}}{SF_{DPDCH}}\right) + 10\log\left(1 + \frac{r^2}{(N + r^2)}\Delta\right) . \tag{D.18}$$

The target SIR, however, can be adversely affected by the semi-active OLPC entities and static rate matching attributes when the associated inactive DCH is mapped onto the same uplink DPDCH as the selected active DCH. This is illustrated in Figure D.3 for one active OLPC entity and one semi-active OLPC entity. In the figure, a first, upper time line is associated to a selected active DCH and the corresponding active OLPC entity. This entity has TTIs of e.g. 80 ms. A second, lower time line is associated to a semi active entity associated to an inactive DCH. The second entity has TTIs of e.g. 10 ms. Both time lines have the length of one complete reporting period of e.g. 480 ms. The reporting period is a radio network planning (RNP) parameter. At the end of the respective time line, both entities send an activity report to the OLPC controller.

As mentioned above, each active or semi-active entity is allowed to send each TTI a request for a change of the current SIR to the OLPC controller. In the presented example, the active entity sends a change request after the first and the second TTI, and the semi active entity sends a change request after the 5th and the 7th TTI, each request being indicated in the figure as ΔSIR. Thus the target SIR, which is computed by the OLPC controller from all received valid requests, can be affected by the semi active entity and static rate matching attributes when the inactive DCH is mapped onto the same uplink DPDCH as the active DCH. Equally, the quality estimates BER and/or BLER can be affected by such semi active entities. Thus, in order to be able to calculate transport channel specific parameter values from determined target SIR or from determined quality estimates, it has to be known whether the respective SIR or the respective quality estimates are influenced by the presence of other transport channels. Hence, the uplink OLPC controller is informed whether a transport channel has experienced any traffic or not in a reporting period by a corresponding indication in the activity report, which is sent to the OLPC controller by the OLPC entity associated to the transport channel each reporting period.

Since the static rate matching does not affect the computation when only the wanted bearer service is active, all other bearer services being in DTX, the E_b/N_0 of the transport channel can be estimated for such times from (D.16) (or (D.18) in the case of HSDPA).

For solving this equation, it is assumed that β_c, β_d, R_{DPDCH} and SF_{DPDCH} are produced by the AC considering all other bearer services than the currently selected bearer service not to be in uplink transmission (DTX mode) and the maximum bit rate of the transport channel to be used, which is assumed to be $MAX_{DCH}\{TFS_{DCH}\}$, TFS being the transport format set.

In order to be able to determine for each transport channel an E_b/N_0 value and quality parameter values, a set of parameters provided by the AC to the OLPC controller and a set of parameters provided by the OLPC entities to the OLPC controller are defined.

At the RAB setup, after static rate matching, i.e. E_b/N_0 balancing, the uplink OLPC controller receives from the AC on the one hand radio link specific parameters and on the other hand DCH specific parameters.

The radio link specific parameters comprise as one parameter a counters update period. This parameter is needed for online and/or trace and must be set as an integer number of activity reporting periods of the uplink OLPC entities. It can be set in a range of 1 to 1000 reporting periods, e.g. to 100 reporting periods. A further parameter is the size of an E_b/N_0 Sliding Window, e.g. 20 reporting periods. This parameter is used for averaged E_b/N_0 computations and should lie equally in a range of 1 to 1000 reporting periods. The last radio link specific parameter is an online and/or trace parameter, which can be set to “Y” or “N”.

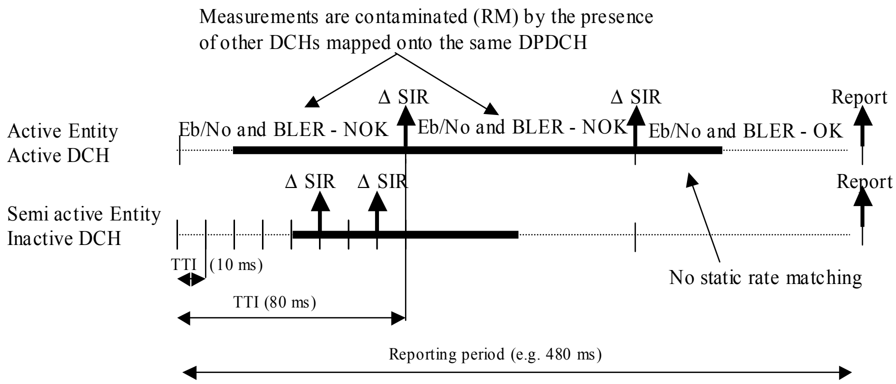


Figure D.3. Influence of other DCHs on the E_b/N_0 measurements for one DCH.

DCH specific parameters comprise an enable measurements parameter, which can be used to enable the management class based measurements. Further comprised are a coding gain in dB, a gain factors gain in dB and an SF gain in dB, each gain for the case that only the currently regarded transport channel is active, i.e. all others are in DTX, and that the maximum bit rate, i.e. $MAX_{DCH}\{TFSDCH\}$, is used (see (D.16), or (D.18) in the case of HSDPA). Another parameter is the management class, which is defined based on the RAB attributes, as presented in Section 6.3.1. Further a quality target for BLER and/or BER is given. Equally, the size of a BLER sliding window and/or the size of a BER sliding window are provided, the BER sliding window however only if turbo coding is used. Both can be set e.g. to 20 in terms of reporting periods. Both can be selected again from a range of 1 to 1000 reporting periods.

During the RRC connection, the measurements are started and stopped by a Mobile Connection Control (MCC) unit using a separate message, which allows cell-based measurements.

Based on the parameters the OLPC entity receives from the uplink OLPC controller when it is set up, it sends an activity report to the controller once every reporting period. As already mentioned, the reporting period is one of several RNP parameters that each OLPC entity receives during setup by the uplink OLPC controller. The reporting period ranges from 80 ms up to 2400 ms in steps of 80 ms. The default value is 480 ms, as in Figure D.3. The activity report comprises the sum of the SIR target changes during the DTX-off periods. It further comprises the DTX state of the controlled bearer, which can be 'on' or 'off'. Moreover, a DTX indicator is provided, which is used to inform the OLPC controller whether the DCH associated to the respective OLPC entity has been active or not during the terminated reporting period. In addition, the number of CRC OK and the number of CRC NOK after selection and combining during the reporting period are provided in the activity report. The sum of the QEs, after selection and combining, divided by the number of DCH-FP (Frame Protocol) instances, received during the reporting period, are equally included in the activity report. Finally, a connection frame number CFN is used as a time stamp for measurement synchronisation.

With the information of the AC and of the OLPC entities, the uplink OLPC controller is then able to calculate values for the E_b/N_0 of a specific transport channel. When DCH

Active is the transport channel associated with the currently active OLPC entity, the uplink OLPC controller can determine an average E_b/N_0 for this transport channel as follows:

$$\overline{\left(\frac{E_b}{N_0}\right)}_{DCH,Active} = \frac{\sum_{SW_1}\left(\frac{E_b}{N_0}\right)_{DCH,Active}}{SW_1}, \tag{D.19}$$

where

$$\left(\frac{E_b}{N_0}\right)_{DCH,Active} = \frac{\sum_{RP} 10^{\frac{SIR_{DPCCH} - 20\log(r) - 20\log\left(\frac{R_{DCH}}{R_{DPDCH}\left(N+r^2\right)}\right) - 10\log\left(\frac{SF_{DPCCH}}{SF_{DPDCH}}\right)}{10}}}{RP}, \tag{D.20}$$

and SW_1 denotes the sliding window size for the average computation and RP the reporting period.

Further a BLER can be determined for the active transport channel DCH Active by:

$$\overline{BLER}_{DCH,Active} = \frac{\sum_{SW_2} BLER_{DCH,Active}}{SW_2}, \tag{D.21}$$

where SW_2 denotes the BLER sliding window size, and

$$BLER_{DCH,Active} = \frac{\sum_{RP} CRC\_NOK}{\sum_{RP}(CRC\_OK + CRC\_NOK)}. \tag{D.22}$$

In addition or alternatively, also a BER can be determined for the active transport channel DCH Active by:

$$BER_{DCH,Active} = \frac{\sum_{SW_3} BER_{DCH,Active}}{SW_3}, \tag{D.23}$$

where SW_3 denotes the BER sliding window size, and

$$BER_{DCH,Active} = \frac{\sum_{RP} QE}{RP}. \tag{D.24}$$

The BER computation is only possible, however, when turbo coding is used.

Since each activity report includes an information on whether the respective entity has been active during the last reporting period, the OLPC controller is able to reset the quality and E_b/N_0 computations carried out for one entity any time one of the respective other entities has experienced traffic within the preceding reporting period, i.e. if it has not been in DTX mode over the entire reporting period.

Furthermore, the sliding window content for quality computations needs to be reset when the target SIR is changed and thus sent to the WCDMA node B.

Based on MCC start/stop measurement commands, it is possible to update the RRC connection based counters cell by cell and management class based, since the MCC is aware of the cell participating in diversity handover (DHO). Furthermore, for each management class, within the same RRC connection, the uplink OLPC controller updates each reporting period the following counters, when receiving a Start Measurements Command from the MCC: the determined average E_b/N_0 in linear units, the determined BLER, the determined BER, the total number of CRC OK, and the total number of CRC NOK.

If the MCC needs OLPC data from a specific radio connection for an online and/or trace, i.e. if the online and/or trace parameter has been set to 'Y', the counters are delivered to the MCC together with the actual CFN every counters update period.

Upon receiving the stop measurements command, the uplink OLPC controller sends the collected counters to the MCC and resets the sliding window contents. The same actions are taken by the OLPC controller when it receives the RRC connection release command from the MCC, when the relocation of the serving RNC (SRNC) is executed, or when a handover to another system or frequency is triggered off. Also if a DCH is released and a RAB reconfiguration is carried out, within the current RRC connection, the counters corresponding to this bearer service are sent to the MCC.

Appendix E: Genetic Algorithm Pseudo Code

This appendix describes the pseudo code of the genetic algorithm proposed in this work for QoS optimisation in UTRAN. See Chapter 7 for more details.

```
% Number of iterations
nIterations = 150;

% Number of initial individuals
populationSize = 10;

% Probability of mutation
mutationProb = 0.1;

% ============= Parameters to be optimised ==============

QoS(1).minBitRate = [8, 16]; % PoC
QoS(1).maxBitRate = [8, 16];
QoS(1).inactTimer = [1, 2, 5, 10, 20, 30];

QoS(2).minBitRate = [64, 128]; % Streaming
QoS(2).maxBitRate = [64, 128];
QoS(2).inactTimer = [1, 2, 5, 10, 20, 30];

QoS(3).minBitRate = [32, 64, 128, 144, 256, 384]; % WAP/MMS
QoS(3).maxBitRate = [32, 64, 128, 144, 256, 384];
QoS(3).inactTimer = [1, 2, 5, 10, 20, 30];

QoS(4).minBitRate = [64, 128, 144, 256, 384]; % Dialup
QoS(4).maxBitRate = [64, 128, 144, 256, 384];
QoS(4).inactTimer = [1, 2, 5, 10, 20, 30];

minAllocTime = [1, 2, 5, 10, 15, 20];

maxCRQueuingTime = [1, 2, 5, 10, 15, 20];


% Targets for users satisfaction for each of the offered services

TargetSatisfaction(1) = .90; % Signalling
TargetSatisfaction(2) = .90; % Emergency call
TargetSatisfaction(3) = .90; % Conversational speech
TargetSatisfaction(4) = .90; % Conversational data
TargetSatisfaction(5) = .90; % Streaming
TargetSatisfaction(6) = .90; % Multimedia Messages (MMS)
TargetSatisfaction(7) = .90; % Push to Talk (PoC)
```

```
TargetSatisfaction(8) = .90; % See What I See
TargetSatisfaction(9)  = .90; % Dialup connection
TargetSatisfaction(10) = .90; % WAP browsing

parentProb = zeros(populationSize,1);
parentDiv = (populationSize+1)*populationSize/2;
for i = 1:populationSize
    parentProb(i) = (populationSize-i+1)/parentDiv;
    if i > 1
        parentProb(i) = parentProb(i) + parentProb(i-1);
    end
end

for step = length(population)+1:nIterations
    % Generate the tested parameter values
    notgood = 1;
    mutProb = 1; % for generation of initial population
    while notgood == 1
        notgood = 0;
        if step > populationSize
            % select parents with weighted probability
            cut = rand(2,1);
            parents = zeros(2,1);
            for p = 1:2,
                i = 1;
                while cut(p) > parentProb(i)
                    i = i + 1;
                end
                parents(p) = i;
            end
            parentList = [parentList; parents];
            for i = 1:14,
                population(step,i) = population(parents(ceil(rand*2)),i);
            end
            mutProb = mutationProb; % After in. pop. decrease prob. to 0.1
        end
        % mutation
        for i=1:4
            if rand < mutProb
                population(step,i) =
                 QoS(i).minBitRate(ceil(rand*length(QOS(i).minBitRate)));
            end
            if rand < mutProb
                population(step,i+4) =
                 QoS(i).maxBitRate(ceil(rand*length(QOS(i).maxBitRate)));
            end
            if population(step, i+4) < population(step, i)
                notgood = 1;
            end
            if rand < mutProb
                population(step,i+8) =
                 QoS(i).inactTimer(ceil(rand*length(QOS(i).inactTimer)));
            end
        end
        if rand < mutProb
            population(step,13) =
                 minAllocTime(ceil(rand*length(minAllocTime)));
        end
        if rand < mutProb
            population(step,14) =
                 maxCRQueuingTime(ceil(rand*length(maxCRQueuingTime)));
```

```
        end
        % check for duplicates
        if step > 1
            for j=1:step-1
                if sum(population(step,1:14)-population(j,1:14))==0
                    notgood = 1;
                end
            end
        end
    end

    % ================== Provision new parameters ===================
    ProvisionNewParameters;

    % =================== Process calls ============================
    ProcessAllCalls;

    % ==================== Compute fitness ==========================

    % Compute number of calls/sessions for each service type
    ServiceType(i);


    % Compute amount of satisfied users for each service type
    SatisfiedUsers(i);

    for i=1:NofApplicationServices
        RatioService(i)= SatisfiedUsers(i)/max(ServiceType(i),1);
        SatisfactionService(i)=(RatioService(i)-
               TargetSatisfaction(i))/sqrt((TargetSatisfaction(i)*(1-
               TargetSatisfaction(i)))/max(ServiceType(i),1));
    end

    fitness = -min(SatisfactionService(find(RatioService>0)));

    % Collect population and corresponding fitness values
    population(step, 15) = fitness;
    population(step, 16) = step;
    population = sortrows(population, 15)
    fitnessList = [fitnessList; population(1,15)];
    satisfiedUserTable(step,:) = RatioService(find(RatioService>0));
```

end % Iterations

References

[1] 3GPP, R5, TS 23.107, "QoS concept and architecture."

[2] 3GPP, R5, Technical Specification 32.101, "3G Telecom Management: Principles and high-level requirements."

[3] Tele Management Forum, "SLA Management Handbook," v.1.5, June 2001.

[4] V.K. Garg, O.T.W. Yu, "Integrated QoS support in 3G UMTS networks," IEEE, Wireless Communications and Networking Conference, Sept. 2000, pp. 1187-1192, vol.3.

[5] F. Ruijun, S. Junde, "Some QoS issues in 3G wireless networks," IEEE Conference on Computers, Communications, Control and Power Engineering, Oct. 2002, pp. 724-727, vol.2.

[6] E. Villier, L. Lopes, S. Lambotharan, "Approaches to modelling the physical layer performance in a UMTS radio system simulator," IEEE, Third International Conference on 3G Mobile Communication Technologies, May 2002, pp. 560-564.

[7] A. Viterbi, CDMA principles of Spread Spectrum Communication, Addison-Wesley, 1995.

[8] S. Cordier, S. Ortega, "On WCDMA downlink multiservice coverage and capacity," IEEE, VTC Fall 2001, VTS 54th, pp. 2754 – 2758, vol. 4.

[9] D. Adiego, C. Cordier, "Multi-service radio dimensioning for UMTS packet-switched services," IEEE, PIMRC, 2002, pp. 2409 – 2413, vol. 5.

[10] A. Wacker, J. Laiho, K. Sipilä, M. Jasberg, "Static simulator for studying WCDMA radio network planning issues," VTC Spring 1999, pp. 2436-2440, vol.3

[11] J. Laiho, A. Wacker, and T. Novosad, (Editors), Radio Network Planning and Optimisation for UMTS, John Wiley & Sons, April 2002, 484p.

[12] E.L. Andrade, A.A. Shinoda, M.E. Pellenz, M.D. Yacoub, "A comprehensive 3G link level simulator," IEEE, 35th Annual Simulation Symposium, April 2002, pp. 381-388.

[13] R. Hoppe, H. Buddendick, G. Wolfle, F.M. Landstorfer, "Dynamic simulator for studying WCDMA radio network performance," IEEE VTC 2001 Spring, pp. 2771-2775, vol.4.

[14] S.A. Ghorashi, E. Homayounvala, F. Said, A.H. Aghvami, "Dynamic simulator for studying WCDMA based hierarchical cell structures," IEEE, Personal, Indoor and Mobile Radio Communications, 30 Sept.-3 Oct. 2001, pp. D-32 - D-37, vol.1.

[15] T. Klingenbrunn, P. Mogensen, "Modelling radio link performance in UMTS W-CDMA network simulations," IEEE VTC Spring, Tokyo, 2000, pp. 1011-1015, vol.2.

[16] S. Hämäläinen, H. Holma, K. Sipilä, "Advanced WCDMA radio network simulator," Proceedings of PIMRC, Osaka, Japan, September 1999, pp. 951-955.

[17] H. Holma and A. Toskala, (Editors), WCDMA for UMTS, John Wiley & Sons, April 2000, 344 p.

[18] S.A. Malik, S. Akhtar and D. Zeghlache, "Performance of prioritised resource control for mixed services in UMTS W-CDMA networks," VTC Fall 2001, pp. 1000 –1004, vol.2.

[19] S. Akhtar, S.A. Malik, D. Zeghlache, "Prioritised admission control for mixed services in UMTS WCDMA networks," 12th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, 2001, pp. B-133 - B-137, vol.1.

[20] S.A. Malik and D. Zeghlache, "Downlink capacity and performance issues in mixed services UMTS WCDMA networks," VTC Spring 2002, pp. 1824–1828, vol. 4.

[21] S.A. Malik and D. Zeghlache, "Resource allocation for multimedia services on the UMTS downlink", ICC 2002, pp. 3076 –3080, vol.5.

[22] S. Kim, H. Kim, J. A. Copeland, “Dynamic radio resource allocation considering QoS in UMTS network,” 4th International Workshop on Mobile and Wireless Communications Network, 9-11 Sept. 2002, pp. 636 – 640.
[23] F.Y. Li, N. Stol, “A priority-oriented call admission control paradigm with QoS renegotiation for multimedia services in UMTS,” IEEE, VTC Spring, May 2001, pp. 2021 – 2025, vol. 3.
[24] L. Vignali, “Adaptive admission control for granted speech user QoS in a WCDMA mixed services scenario,” IEE, 3rd International Conference on 3G Mobile Communication Technologies, May 2002, pp. 239 – 243.
[25] Eric Hsiao-kuang Wu, Hao-Wei Chang, K.C. Hsu, “A QoS-based hybrid multiple access transmission strategy in WCDMA downlink,” IEEE, Wireless Communications and Networking, March 2003, pp. 1685 – 1690, vol. 3.
[26] R. Vaanithamby, E.S. Sousa, “Resource allocation and scheduling schemes for WCDMA downlinks,” IEEE, International Conference on Communications, June 2001, pp. 1406 – 1410, vol. 5.
[27] S. Heier, C. Ellerbrock, M. Malkowski, M., “UMTS medium access control quality of service scheduling,” IEEE, 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, Sept. 2002, pp. 1893 – 1897, vol.4.
[28] S. Heier, M. Malkowski, “Quality of service scheduling for UMTS,” IEE, 5th European Personal Mobile Communications Conference, April 2003, pp. 529 – 533.
[29] S. Chemiakina, L. D'Antonio, F. Forti, F., R. Lalli, J. Petersson, A. Terzani, A., “QoS enhancement for adaptive streaming services over WCDMA,” IEEE Journal on Selected Areas in Communications, Dec. 2003, pp. 1575 – 1584, vol. 21.
[30] 3GPP, R5, TS 32.403, “Telecommunication management; Performance Management (PM); Performance measurements - UMTS and combined UMTS/GSM.”
[31] ETSI TS 102 250-1, “Speech Processing, Transmission and Quality Aspects (STQ); QoS aspects for popular services in GSM and 3G networks; Part 1: Identification of Quality of Service aspects.”
[32] ETSI TS 102 250-2, “Speech Processing, Transmission and Quality Aspects (STQ); QoS aspects for popular services in GSM and 3G networks; Part 2: Definition of Quality of Service parameters and their computation.”
[33] ETSI TS 102 250-3, “Speech Processing, Transmission and Quality Aspects (STQ); QoS aspects for popular services in GSM and 3G networks; Part 3: Typical procedures for Quality of Service measurement equipment.”
[34] ETSI TS 102 250-4, “Speech Processing, Transmission and Quality Aspects (STQ); QoS aspects for popular services in GSM and 3G networks; Part 4: Requirements for Quality of Service measurement equipment.”
[35] ETSI TS 102 250-5, “Speech Processing, Transmission and Quality Aspects (STQ); QoS aspects for popular services in GSM and 3G networks; Part 5: Definition of typical measurement profiles.”
[36] ETSI TS 102 250-6, “Speech Processing, Transmission and Quality Aspects (STQ); QoS aspects for popular services in GSM and 3G networks; Part 6: Post processing and statistical methods.”
[37] ETSI TS 102 250-7, “Speech Processing, Transmission and Quality Aspects (STQ); QoS aspects for popular services in GSM and 3G networks; Part 7: Sampling methodology.”
[38] Y. Jiang, C. Tham and C. Ko, “Providing quality of service monitoring: Challenges and approaches,” Network Operation and Management Symposium, April 2000, IEEE/IFIP, pp. 115-128.

[39] R. Vukovic and J. Rådemar, “Approach to E2E service assurance on the mobile internet,” Ericsson Review No. 3, 2001.

[40] A. Gurtov, M. Passoja, O. Aalto and M. Raitola, “Multi-layer protocol tracing in a GPRS network,” IEEE VTC fall 2002, pp. 1612-1616, vol. 3.

[41] H. Olofsson, S. Magnusson, M. Almgren, “A concept for dynamic neighbour cell list planning in a cellular system,” IEEE, PIMRC, Oct. 1996, pp. 38 – 142, vol.1.

[42] S. Magnusson, H. Olofsson, “Dynamic neighbour cell list planning in a microcellular network,” IEEE, Universal Personal Communications Record, Oct. 1997, pp.223 – 227, vol. 1.

[43] R. Guerzoni, D. Soldani, I. Ore, K. Valkealahti, “Automatic neighbour cell list optimisation for 3G networks: theoretical approach and experimental validation,” International Wireless Summit, WPMC05, Sept. 2005, Aalborg, Denmark, in press.

[44] P. Calegari, F. Guidec, P. Kuonen, D. Wagner, “Genetic approach to radio network optimisation for mobile systems,” VTC, 4-7 May 1997, pp. 755 – 759, vol. 2.

[45] K. Lieska, E. Laitinen, J. Lahteenmaki, “Radio coverage optimisation with genetic algorithms,” Personal, Indoor and Mobile Radio Communications, 1998, pp. 318 – 322, vol.1.

[46] D. Soldani, M. Abramowski, “An improved method for assessing packet data transfer performance across an UMTS Network,” The 5th International Symposium on Wireless Personal Multimedia Communications, WPMC, Oct. 2002, pp. 534 – 538, vol.2.

[47] Kim Yun Sik, Jung Hyun-Meen, “Efficient radio network optimisation,” VTC Spring 2003, April 22-25, 2003, pp. 1546 – 1549, vol. 3.

[48] A. Höglund, K. Valkealahti, “Quality-based tuning of cell downlink load target and link power maxima in WCDMA,” VTC Fall 2002, pp. 2248 – 2252, vol.4.

[49] K. Valkealahti, A. Höglund, “UMTS Radio network multiparameter control,” IEEE, International Symposium on Personal, Indoor and Mobile Radio Communication Proceedings, Sept. 7-10, 2003, pp. 616-621, vol. 1.

[50] A. Höglund and K. Valkealahti, “Automated Optimisation of Key WCDMA Parameters”, Wireless Communications and Mobile Computing (in Press).

[51] D. Soldani, R. Cuny, “On the Deployment of Multimedia Services in Wireless Networks: Radio Dimensioning Aspects for UTRAN FDD,” IEEE, 1st International Conference on Quality of Service in Heterogeneous Wired/Wireless Networks, QShine 2004, Dallas, USA, pp. 189-196.

[52] D. Soldani and J. Laiho, “A Virtual Time Simulator for Studying QoS Management Functions in UTRAN,” IEEE, VTC Fall 2003, pp. 3453 – 3457, vol.5.

[53] D. Soldani, A. Wacker and K. Sipilä, “An enhanced virtual time simulator for studying QoS provisioning of multimedia services in UTRAN,” Management of Multimedia Networks and Services, Lecture Notes in Computers Science 3271, Springer, pp. 241-254.

[54] D. Soldani and J. Laiho, “An enhanced radio resource management with service and user differentiation for UMTS networks,” IEEE, VTC Fall, Orlando, USA, Oct. 2003, pp. 3473 – 3477, vol. 5.

[55] D. Soldani and J. Laiho, “User perceived performance of interactive and background data in WCDMA networks with QoS differentiation,” WPMC, October 2003, Yokosuka, Japan, pp. 303-307, vol. 2.

[56] J. Laiho and D. Soldani, “A policy-based quality of service management system for UMTS radio access networks,” WPMC, October 2003, Yokosuka, Japan, pp. 298-302, vol. 2.

[57] D. Soldani, K. Sipilä and A. Wacker, “Provisioning radio access networks for effective QoS management: capacity gains of service differentiation in UTRAN”, IEEE International

Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM2005), June, 2005, Italy.

[58] D.Soldani, P.T. Kohonen, "Determination of parameter values of an uplink transport channel," International application published under the Patent Cooperation Treaty (PCT), WO 03/021976 A1.

[59] D. Soldani, N. Lokuge, A. Kuurne, "Service performance monitoring for GPRS/EDGE network based on treatment classes," IEEE, 12th International Workshop on QoS, iwQoS04, June, 2004, pp. 121-128.

[60] D. Soldani, K. Valkealahti, "Genetic approach to QoS optimisation for WCDMA mobile networks", IEEE VTC spring, June, 2005, Stockholm, Sweden.

[61] K. Valkealahti, D. Soldani, "QoS sensitivity to selected packet scheduling parameters in UTRAN," IEEE VTC Spring, June, 2005, Stockholm, Sweden.

[62] 3GPP, R5, TS 25.401, "UTRAN Overall Description."

[63] 3GPP, R5, TS 25.301, "Radio Interface Protocol Architecture."

[64] H. Kaaranen, A. Ahtiainen, L. Laitinen, S. Naghian, V. Niemi, UMTS Networks – Architecture, Mobility and Services, John Wiley & Sons, second edition, 2005, 406 p.

[65] 3GPP, R5, TS 25.321, "MAC Protocol specification."

[66] 3GPP, R5, TS 25.322, "RLC Protocol specification."

[67] 3GPP, R5, TS 25.323, "PDCP Protocol specification."

[68] 3GPP, R5, TS 25.324, "BMC Protocol specification."

[69] 3GPP, R5, TS 25.331, "RRC Protocol specification."

[70] 3GPP, R5, TS 25.302, "Services provided by the physical layer."

[71] 3GPP, R5, TS 25.212, "Multiplexing and channel coding (FDD)."

[72] 3GPP, R5, TS 25.211, "Physical channels and mapping of transport channels onto physical channels (FDD)."

[73] 3GPP, R5, TS 23.060, "General Packet Radio Service (GPRS); service description; stage 2."

[74] ETSI, TR 101 112 v.3.2.0, "Selection procedures for the choice of radio transmission technologies of the UMTS," UMTS 30.03 v.3.2.0.

[75] N. Shankaranarayanan, Z. Jiang, and P. Mishra, "User-perceived performance of Web-browsing and Interactive data in HFC cable access networks," ICC 2001, pp. 1264 –1268, vol. 4.

[76] P. Barford and M. Crovella, "A performance evaluation hypertext transfer protocols", Boston University Technical Report BU-TR-98-016, Oct 27, 1998. (http://www.cs.bu.edu/techreports)

[77] D. Staehle, et al., "Source traffic modelling of wireless applications", University of Würzburg, Report N.261, June 2000.

[78] A. Klemm, C. Lindemann and Marco Lohmann, "Traffic models for characterisation of UMTS networks," GLOBECOM'01, IEEE, pp. 1741 – 1746, vol. 3.

[79] T. Halonen, J. Romero and J. Melero, (Editors), GSM, GPRS and EDGE Performance, John Wiley & Sons, second edition, April 2003, 615 p.

[80] COST 231, TD(91)73, Urban transmission loss models for mobile radio in the 900 and 1800 MHz bands.

[81] ITU-R M. 1225, "Guidelines for Evaluation of Radio Transmission Technologies for IMT-2000," Recommendation, 1997.

[82] Stephen V. Hanly, Information Capacity of Radio Networks, Ph.D. Dissertation, Kings College, University of Cambridge, August 1993, 225 p.

[83] J. Laiho, A. Wacker, T. Novosad, A. Hamalainen, “Verification of WCDMA radio network planning prediction methods with fully dynamic network simulator,” IEEE, VTC 2001 Fall, pp. 526 – 530, vol.1.

[84] 3GPP, TS 25.922, “Radio resource management strategies.”

[85] Ajay R. Mishra, Fundamentals of Cellular Network Planning and Optimisation: 2G/2.5G/3G... Evolution to 4G, John Wiley & Sons, May 2004, 286 p.

[86] H. Holma, D. Soldani, K. Sipilä, “Simulated and measured WCDMA uplink performance,” IEEE, Proc. of the VTC 2001 Fall, IEEE VTS 54th, Oct. 2001, pp. 1148 – 1152, vol.2.

[87] 3GPP, R5, TS 25.433, “UTRAN Iub Interface NBAP Signalling.”

[88] 3GPP, R5, TS 25.215, “Physical layer - Measurements (FDD).”

[89] 3GPP, R5, TS 25.214, “Physical layer procedures (FDD).”

[90] J. F. Kenney, and E. S. Keeping, “Confidence Limits for the Binomial Parameter” and “Confidence Interval Charts.” §11.4 and 11.5 in Mathematics of Statistics, Pt. 1, 3rd ed. Princeton, NJ: Van Nostrand, pp. 167-169, 1962.

[91] R. R. Johnson and P. J. Kuby, Elementary Statistics, Duxbury Press, 9th edition, 2003.

[92] 3GPP, R5, 3GPP TS 25.413 “UTRAN Iu interface RANAP signalling.”

[93] 3GPP, R5, TS 25.308, “UTRA High Speed Downlink Packet Access (HSDPA); Overall description; Stage 2.”

[94] 3GPP, R5, TR 25.877, “High Speed Downlink Packet Access (HSDPA) - Iub/Iur Protocol Aspects.”

[95] 3GPP, R5 TR 25.890, “High Speed Downlink Packet Access: UE Radio Transmission and Reception.”

[96] T. Kolding, K. Pedersen, J. Wigard, F. Frederiksen, P. Mogensen “High Speed Downlink Packet Access: WCDMA Evolution,” IEEE Vehicular Technology Society (VTS), News, n. 1, 2003, pp. 4 – 10, vol. 50.

[97] 3GPP, R6, TR 25.899, “HSDPA Enhancements”.

[98] P. Zanier, D. Soldani, “A simple approach to HSDPA dimensioning,” IEEE, PIMRC, 2005, Berlin, Germany, in press.

[99] Haitao Lin, Mainak Chatterjee, Sajal K. Das and Kalyan Basu, “ARC: An Integrated Admission and Rate Control Framework for CDMA Data Networks Based on Non-Cooperative Games,” International Conference on Mobile Computing and Networking (ACM MobiCom), San Diego, CA, Sept, 2003, pp. 326 – 338.

[100] Sajal K. Das, Mainak Chatterjee and Haitao Lin, “An Econometric Model for Resource Management in Competitive Wireless Data Networks,” to appear in IEEE Network Magazine, Nov, 2004.

[101] W. D. Kelton, “Statistical analysis of simulation output,” IEEE, Proceedings of the 29th Winter Simulation Conference, Atlanta, Georgia, United States, 1997, pp. 23 – 30.

[102] D. C. Montgomery, G. C. Runger, Applied Statistics and Probability for Engineers, John Wiley & Sons, Inc. 1994, 950 pp.

[103] Papoulis, Probability, Random Variables, and Stochastic Processes, McGraw Hill, 1991, 3rd edition, 666 pp.