Radio Access Network Design for the Evolved UMTS Network

Xinzhi Yan

M.Eng

A thesis submitted in partial fulfilment
of the requirements for the degree of

Doctor of Philosophy

February, 2010

School of Electrical Engineering and Computer Science

The University of Newcastle
Callaghan, NSW 2308
Australia

I hereby certify that the work embodied in this thesis is the result of original research and has not been submitted for a higher degree to any other University or Institution.

Xinzhi Yan

Acknowledgements

First I would like to acknowledge Australian Research Council and Singtel OPTUS to provide me Australian Postgraduate Award Industry (APAI) to support my research in the University of Newcastle.

I would like to thank my research supervisor, Dr. Jamil Y. Khan, for his wise advice and encouragement during the completion of this Ph.D. works. I also would like to thank my research co-supervisor, A/Prof. Steve Weller, for his well organization of my research upgrade.

I also would like to give thanks to Dr. Brendan Jones, at Singtel OPTUS, for his technical direction to this work.

Foremost, I am profoundly grateful to my family, especially my wife, Wei Dai, who have always expressed me her unconditional support and love; my Cathy and Janet, their smile give me inspiration.

Abstract

The Radio Access Network (RAN) accounts for the major proportion of the UMTS system operating cost. Transmission from radio base station sites contributes a larger part of the RAN operating costs. Selection of suitable transport technologies and proper allocation of network resources are vital from an operator cost optimisation and the Quality of Experience (QoE) points of view. This thesis extensively investigated the performance of a RAN to support multimedia traffic on a HSDPA air interface. Transport network layer of a future RAN could be based on a number of transport protocols such as ATM, IP and Ethernet. With the increasing traffic volume and diversity the efficiencies of IP and Ethernet based RAN could increases significantly due to the use of larger payloads and simpler resource allocation techniques. Also, on IP and Ethernet based links relatively fewer overhead bits are transmitted compared to an ATM based link. Both the IP and Ethernet based links appear to perform better under heavy traffic load conditions. An IP based link could perform better than an Ethernet based link when an IP header compression technique is used. An Ethernet based link is an alternative transport technique for the UTRAN transport network due to its flexibility, economy and bandwidth efficiency.

The HSDPA (High Speed Downlink Packet Access) is considered to be one of the initial evolutionary steps to enhance the data rate, and QoS of downlink data and multimedia services for the evolved UMTS network. It can provide high data rate up to 28.8 Mbps on the downlink shared channel using the packet access technique. A HSDPA network can dynamically adjust a connection data rate to match radio conditions to ensure the highest possible data rate for different type of traffic. Inappropriate RAN capacity allocation could lead to low radio resource or RAN resource utilizations. In this thesis, a Markov chain based analytical model has been developed to study the interaction between the air interface and the RAN for a HSDPA network. The analytical model was used to study interactions of RAN transport protocols, flow control techniques and the air interface transmission conditions. Further a simulation model was developed to investigate the relationship between the HSDPA air interface and its RAN

parameters. Another important issue in the HSDPA network design is the scheduling algorithm used at the Node-B. A scheduling algorithm plays a key role in allocating a RAN's network resources. Impacts of scheduling algorithms are studied in this work using a simulation model. Based on the study of the HSDPA air interface and its RAN parameter interactions this work has developed an adaptive resource management algorithm, which uses the measured air interface information to allocate the corresponding connection data rate on the I_{ub} link. The developed algorithm reduces RAN resource requirements while increasing the air interface resource utilization and QoS of connections.

Abbreviations

3GPP	3rd Generation Partnership Project
AAL2	ATM Adaptation Layer type 2
AAL5	ATM Adaptation Layer type 5
ACK	Acknowledgement
ALCAP	Access Link Control Application Part
AM	Acknowledged Mode
AMD	Acknowledged Mode Data
AMR	Adaptive Multi-rate (speech codec)
ARQ	Automatic Repeat Request
ATM	Asynchronous Transfer Mode
BCCH	Broadcast Channel (logical channel)
BCH	Broadcast Channel (transport channel)
BCFE	Broadcast Control Functional Entity
BCH	Broadcast Channel (transport channel)
BER	Bit Error Rate
BLER	Block Error Rate
BMC	Broadcast/Multicast Control Protocol
BS	Base Station
BSS	Base Station Subsystem
BSC	Base Station Controller
CCCH	Common Control Channel (logical channel)
CCH	Common Transport Channel
CCH	Control Channel
CDF	Cumulative Distribution Function
CDMA	Code Division Multiple Access
CFN	Connection Frame Number
CIR	Carrier to Interference Ratio
CIP	Composite IP
CM	Connection Management
CN	Core Network

CPCH	Common Packet Channel
CPICH	Common Pilot Channel
CQI	Channel Quality Indicator
CRC	Cyclic Redundancy Check
CRNC	Controlling RNC
CS	Circuit Switched
CTCH	Common Traffic Channel
DCCH	Dedicated Control Channel (logical channel)
DCFE	Dedicated Control Functional Entity
DCH	Dedicated Channel (transport channel)
DL	Downlink
DPCCH	Dedicated Physical Control Channel
DPDCH	Dedicated Physical Data Channel
DRNC	Drift RNC
DRX	Discontinuous Reception
DSCH	Downlink Shared Channel
DTCH	Dedicated Traffic Channel
DTX	Discontinuous Transmission
E-UTAN	Evolved Universal Terrestrial Radio Access Network
FACH	Forward Access Channel
EDF	Earliest Deadline First
FIFO	First In First Out
FP	Frame Protocol
FTP	File Transfer Protocol
GFP	Generic Framing Procedure
GGSN	Gateway GPRS Support Node
GMSC	Gateway MSC
GPRS	General Packet Radio System
GSM	Global System for Mobile Communications
GTP-U	User Plane Part of GPRS Tunnelling Protocol
HARQ	Hybrid Automatic Repeat Request
HDTV	High-definition Television

HDLC	High-Level Data link Control
HLR	Home Location Register
HSDPA	High Speed Downlink Packet Access
HS-DPCCH	Uplink High Speed Dedicated Physical Control Channel
HS-DSCH	High Speed Downlink Shared Channel
HSS	Home Subscriber Server
HS-SCCH	High Speed Shared Control Channel
HSUPA	High Speed Uplink Packet Access
HSS	Home Subscriber Server
IETF	Internet Engineering Task Force
IMA	Inverse Multiplexing for ATM
IMS	IP Multimedia Sub-system
IMT-2000	International Mobile Telephony, 3rd Generation Networks are referred as IMT-2000 within ITU
IP	Internet Protocol
ITU	International Telecommunications Union
L1	Layer 1 (Physical Layer)
L2	Layer 2 (Data Link Layer)
L3	Layer 3 (Network layer)
LAN	Local Area Network
LAPS	Link Access Procedure for SDH
LIPE	Lightweight IP Encapsulation
MAC	Medium Access Control
MBMS	Multimedia Broadcast Multicast Service
ME	Mobile Equipment
MEF	Metro Ethernet Forum
MGCF	Media Gateway Control Function
MGW	Media Gateway
MMoIP	multimedia over IP
MPEG	Motion Picture Experts Group
MS	Mobile Station
MSC/VLR	Mobile Services Switching Centre/Visitor Location Register

MT	Mobile Termination
NBAP	Node B Application Part
NRT	Non-real Time
ODMA	Opportunity Driven Multiple Access
OFDMA	Orthogonal Frequency Division Multiple Access
PAD	Padding
PCCCH	Physical Common Control Channel
PCCH	Paging Channel (logical channel)
PCCPCH	Primary Common Control Physical Channel
PCH	Paging Channel (transport channel)
PCPCH	Physical Common Packet Channel
PDCP	Packet Data Converge Protocol
PDP	Packet Data Protocol
PDH	Plesiochronous Digital Hierarchy
PDSCH	Physical Downlink Shared Channel
PDU	Protocol Data Unit
PHY	Physical Layer
PICH	Paging Indicator Channel
PRACH	Physical Random Access Channel
PS	Packet Switched
PSCH	Physical Shared Channel
PSTN	Public Switched Telephone Network
QAM	Quadrature Amplitude Modulation
QoS	Quality of Service
QPSK	Quadrature Phase Shift Keying
RAB	Radio Access Bearer
RACH	Random Access Channel
RAN	Radio Access Network
RANAP	Radio Access Network Application Part
RB	Radio Bearer
RLC	Radio Link Control
RNC	Radio Network Controller

RNL	Radio Network Layer
RNS	Radio Network Sub-system
RNSAP	Radio Network Subsystem Application Part
RRC	Radio Resource Control
RRM	Radio Resource Management
RT	Real Time
RTP	Real Time Protocol
SAP	Service Access Point
SCCPCH	Secondary Common Control Physical Channel
SCH	Synchronisation Channel
SDH	Synchronous Digital Hierarchy
SDU	Service Data Unit
SF	Spreading Factor
SFD	Start of Frame Delimiter
SGSN	Serving GPRS Support Node
SHO	Soft Handover
SID	Silence Indicator
SINR	Signal-to-Noise Ratio where noise includes both thermal noise and interference
SIP	Session Initiation Protocol
SIR	Signal to Interference Ratio
SNR	Signal to Noise Ratio
SRB	Signalling Radio Bearer
SRNC	Serving RNC
SRNS	Serving RNS
TCH	Traffic Channel
TCP	Transport Control Protocol
TCRTP	Tunnelled multiplexed compressed RTP
TCTF	Target Channel Type Field
TE	Terminal Equipment
TF	Transport Format
TFCI	Transport Format Combination Indicator

TFCS	Transport Format Combination Set
TFI	Transport Format Indicator
TFRC	Transport Format and Resource Combination
TNL	Transport Network Layer
TPC	Transmission Power Control
TR	Transparent Mode
TTI	Transmission Time Interval
UDP	User Datagram Protocol
UE	User Equipment
UL	Uplink
UM	Unacknowledged Mode
UMTS	Universal Mobile Telecommunication Services
UNI	User Network Interface
USCH	Uplink Shared Channel
USIM	UMTS Subscriber Identity Module
UTRA	UMTS Terrestrial Radio Access
UTRAN	UMTS Terrestrial Radio Access Network
VoIP	Voice over IP
WCDMA	Wideband CDMA
WiMAX	Worldwide interoperability for Microwave Access
WLAN	Wireless Local Area Network
WRR	Weighted Round Robin

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

1.1 High Data Rate Mobile Communication Networks

Mobile Telecommunication systems have expanded at a tremendous pace over last 25 years. Within a short period of time the worldwide subscriber number for the UMTS based 3G networks (Universal Mobile Telecommunication System) have exceeded 280 millions served by over 180 networks in more than 100 countries [1]. The UMTS based 3G network is evolving as a more of a data centric network than predominantly voice only networks such as second generation mobile communication systems. Just a few years ago, the reality of 'mobile data' meant little more to real-word users other than sending text messages or struggling to browse share prices and weather reports via a slow mobile connection. However, mobile users have started to experience their first real taste of faster, user-friendly multimedia services on their mobiles while on the move [2]. Unlike 2nd generation mobile networks, which were designed mainly for providing voice services, the UMTS third generation networks have been designed for flexible delivery of multimedia services [3]. Currently the UMTS network supports two parallel standards namely the legacy WCDMA standard based on the 3GPP release 99 and new the data centric HSPA (High Speed Packet Access) standard announced in the 3GPP release 5 and 6. The UMTS network supports both Real Time (RT) services and Non-Real Time (NRT) services. Real Time services such as voice and real-time video requires high probability of guaranteed bandwidth and low transfer delays to maintain it's required QoS (Quality of Service). On the other hand NRT services are more delay tolerant as well as can operate under variable transmission data rate, but can't tolerate packet or information loss. With the increasing mobile networks transmission data rates and emergence of advanced packet transmission techniques operators are now able to support broadcast and multicast services. The broadcast services provided by 3G networks are becoming increasingly popular such as mobile television or video broadcast [4]. Real-time broadcast and multicast services can serve large number of users with a reasonable network capacity. However, these services require additional networking provisions to maintain the QoS of all services offered by mobile networks.

All third generation mobile communications standards have been developed under the umbrella of the 3rd Generation Partnership Project (3GPP), which is a collaboration agreement that brings together a number of telecommunications standards bodies which are known as "Organizational Partners" [5]. The UMTS network architecture was developed with the aim to offer both circuit switch (CS) and packet switched (PS) services. The UMTS network consists of UTRAN (UMTS Terrestrial Radio Access Network), CN (Core Network) and UE (User Equipment). The 3GPP release 99 architecture introduced both connection oriented and connectionless data transmission features using dedicated and common transmission channels. These channels are shared between the real and non-real-time traffic sources. The legacy network was designed to support a maximum data rate of 2 Mbps/carrier. However, in practice typically maximum achievable average data rate of 384 kbps/carrier is achievable in mobile environments. With the introduction of the HSDPA standard the transmission data rate on the downlink increased considerably. Subsequently with the introduction of HSUPA standard the uplink capacity also increased considerably. Recently the HSPA standard has been announced which combines both HSDPA and HSUPA standards. Currently deployed HSPA networks can offer maximum end-users data rate up to 14.4 Mbps on the downlink and 5.76 Mbps on the uplink. The 3GPP release 7 standard which was published in 2007 is aiming at offering 28.8 Mbps on the downlink and 11.5 Mbps on the uplink. These enhancements make it possible for mobile operator to offer high bandwidth multimedia services such as high quality streaming video, high resolution images downloading, interactive gaming, broadband internet connections and other services.

The UMTS/HSPA is an evolutionary standard which will eventually migrate to the LTE (Long Term Evolution) standard. With the increase of the air interface data rates it is necessary to develop the UTRAN and the CN architecture. The UTRAN includes a Radio Network Controller (RNC) and Node-Bs. The 3GPP release 5 introduced the IP based transport protocol along with the ATM based protocol for the UTRAN. These UTRAN entities influence the end user data rate significantly. Many functionalities of the RNC were moved to the Node-B in the release 5 to improve the link latency and transmission data rate [6]. Figure 1.1 shows the improvements of UMTS network capacity and delay with the progression of UMTS network architecture.

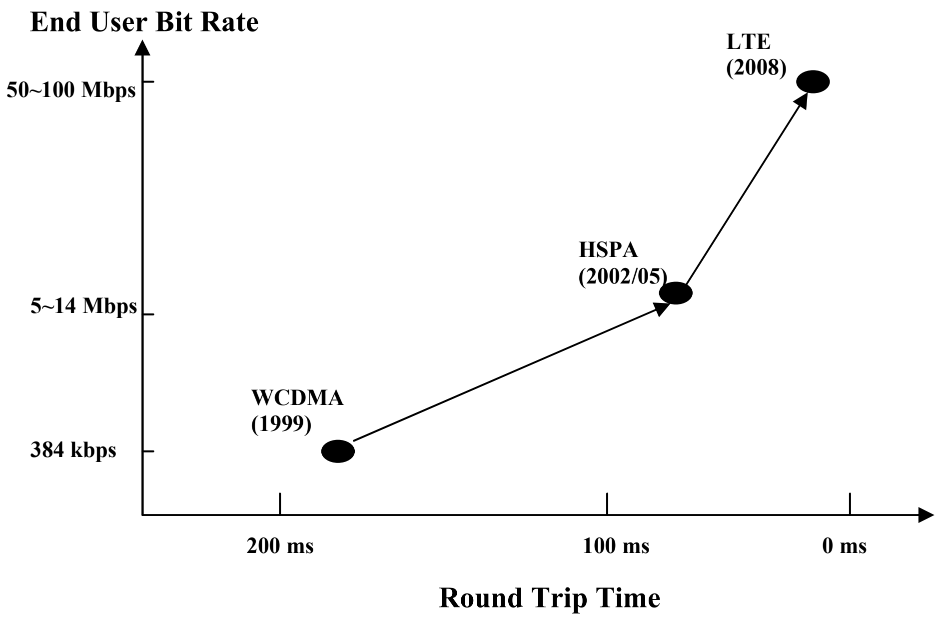


Figure 1.1 Mobile telecommunications network capacity evolution

Figure 1.1 shows that the legacy UMTS channels supports 384 kbps peak data transmission rate with latency value ranging between 100-200 ms. The figure shows that the introduction of the HSPA standard will significantly increase the data transmission rate as well as reducing the packet delay progressively. In order to further reduce cost, and improve service provisions to meet user expectation in a 10 year perspective and beyond, the 3GPP is evolving the current specifications. According to the road map of the LTE the downlink data rate will increase to 100 Mbps and uplink data rate will increase to 50 Mbps using a 20 MHz bandwidth [7].

1.2 Research Motivation

Advanced radio and network resource management algorithms are becoming increasingly important for the evolving UMTS based mobile telecommunication networks. With the increasing complexities and capabilities of the UMTS/UTRAN architecture it is important to jointly optimise radio and network resource algorithms to maintain the end-to-end QoS for different types of air interfaces and corresponding radio access network (RAN). Next generation mobile networks are going to support multiple air interfaces by using a single RAN. Hence, it becomes more important for RAN's to be optimised for different types of traffic services. In addition to optimising the performance of networks it is also important to consider the cost optimisation issue. Current mobile network operators work in a very competitive environment. Optimising expenditure by minimizing the CAPEX (capital expenditure) needed to create new services while keeping the OPEX (operational expenditure) under control is at the top of mobile network operators' agenda [7]. Recent cost analysis shows that the transmission resources of radio base station sites accounts for a significant part of a RAN operating cost [8]. Recent cost analysis shows that the transmission resources of radio base station sites accounts for a significant part of a RAN operating cost. A joint network and radio resource optimisation technique could benefit both application QoS and the operating cost of a network. Joint optimisation of network and radio resources is a fairly new but an important area of research. The joint network resource allocation techniques form the basis of this research work. This work examines the parametric relationship between the HSDPA air interface and its RAN for different transport technologies and services. With the introduction of packet services on air interfaces it is necessary to introduce dynamic network resource management algorithms for the RAN. The release 99 of the 3GPP standard introduced the ATM based UTRAN architecture mainly to cater for circuit switched (CS) and for some packet switched (PS) services [9]. The 3GPP release 5 introduced the first step towards the end-to-end IP services by introducing the IP (Internet Protocol) based transport technology for the UTRAN as an alternate approach to the ATM based RAN [10]. Recent introduction of the Carrier Ethernet standard introduces a possibility of Ethernet based RAN for future telecommunication networks. The focus of this research is to study the HSDPA based RAN optimisation for different types of traffic.

The HSDPA (High Speed Downlink Packet Access) is one of the initial evolutionary steps introduced by the 3GPP to enhance the downlink data and multimedia services by mapping the packet based architecture on the UMTS network. Currently the standard offers high data rate of 28.8 Mbps on a downlink shared channel. The HSDPA standard improves the spectral efficiency by introducing higher order modulation and coding schemes. An HSDPA connection adapts its transmission rate using the adaptive modulation and coding (AMC) technique which compensates for the variable transmission conditions. Unlike the legacy UMTS channels such as the DCH (Dedicated Channel) no power control mechanism is used for HSDPA connections. The standard employs a fast packet scheduling scheme at a Node-B. A HSDPA RAN requires dynamic traffic flow management scheme on the I_{ub} link to maintain a high throughput at the air interface level. The dynamic traffic management algorithms can be influenced by the air interface processes hence, it is necessary to develop RAN traffic management algorithms which are aware of the air interface conditions. An HSDPA connection between a RNC and an UE can be seen as a two hop network where the pull rate of the air interface will vary according to the radio channel conditions and backhaul transport network conditions. RAN resource allocation algorithms need to be dynamic and responsive enough to maintain high QoS for all type of services.

1.3 Summary of Contributions

The main contribution of this work is to study the HSDPA air interface and RAN dependencies, and to develop innovative algorithms to improve the resource utilisation and QoS of different types of services. The work initially studies the UTRAN performance for different transport protocols using multimedia traffic sources. The work then examines the performance of the RAN for different HSDPA air interface conditions. Based on these studies and performance analysis then the work developed an adaptive I_{ub} link resource allocation algorithm and investigate its performance for different traffic schedulers.

The impact of a RAN on the air interface depends on the network architecture and protocols used. In the HSDPA architecture two buffers are maintained between the RNC and UE on the downlink. The packet flow between the RNC and the Node-B is controlled by the I_{ub} link management algorithm as well as by the transport protocol architecture. This work developed a comprehensive simulation model to study the effect of three transport protocols on the multimedia traffic QoS. The simulation model studied the ATM, IP and the Ethernet based transport protocols to control the traffic flows between the RNC and the Node-B.

Quantifying the interdependence of RAN architecture and the HSDPA air interface performance is a new and an important area of research. This work developed a discrete Markov chain model to study the RAN performance. The developed model incorporates the operation of the I_{ub} link and the air interface in the same model. The developed model is a unique one which allows the RAN performance to be examined in terms of transmission channel condition and connection parameters. The model also allows examining the effect of the packet scheduler on the end-to-end QoS i.e. from the RNC to Node-B. The interdependency of two interfaces is also examined by using a simulation model to validate the analytical model. The developed model can be used to dimension the RAN of a HSDPA network.

An adaptive resource allocation algorithm has been developed to optimise the performance of a RAN for the multimedia traffic. The algorithm use the air interface QoS values to allocate appropriate resources for the I_{ub} link. The developed algorithm uses a feedback based approach where the Node-B measures the throughput of the air interface to dimension the I_{ub} link according to the current state of the air interface. Performance analysis shows that the developed algorithm can significantly optimise the performance of a HSDPA network.

1.4 Thesis Outline

Chapter 1 introduce the background of the research work and explains the key contributions of the work.

Chapter 2 review the UMTS and the 3GPP network architecture and protocols. The chapter describes the Medium Access Control (MAC) and the Radio Link Control (RLC) protocols in details. The QoS profile of different multimedia services is reviewed. The chapter also reviews the multimedia communication services using the legacy network architecture.

Chapter 3 reviews the HSPA standard with the main focus on the HSDPA protocol details. The Chapter initially reviews the HSDPA protocol details of both for the air interface and the RAN for packet services. The chapter examines the resource allocation and flow control algorithms for the RAN and the HSDPA air interface. The chapter also examines different HSDPA channel characteristics.

In the Chapter 4 the thesis start analysing the performance of the UTRAN for the HSDPA and DCH traffic channels supporting multimedia traffic. Comprehensive simulation models based on the OPNET have been developed in this chapter to examine the effects of different RAN transport protocols on the multimedia traffic QoS. This chapter also examines the effect priority transmissions in the RAN.

Chapter 5 analyzes the independency between the HSDPA air interface and RAN parameters. Analytical and simulation techniques are used to study the interactions between the HSDPA air interface and the RAN parameters. A

Markov chain based analytical model and a simulation model has been developed to evaluate the joint resource allocation algorithms.

At the beginning the Chapter 6 the impact of the I_{ub} link capacity on the HSDPA air interface for different packet scheduling techniques is investigated. Based on various performance analysis results the adaptive RAN resource allocation algorithm is proposed to improve the HSDPA air interface and the I_{ub} link utilizations. The algorithm adaptively allocate RAN resources by estimating the HSDPA air interface throughput to pro-actively prevent any congestion, and to optimize the both air interface and RAN performances to improve the overall QoS of HSDPA connections.

Chapter 7 draws main conclusion of this research and finishes with a summary of further research opportunities.

1.5 Prior Publications

The following articles have been published during the PhD work:

J. Y. Khan and X. Yan, “HSDPA Radio Access Network Design”, Chapter 9, HSDPA/HSUPA Handbook, CRC Press, USA, ISBN: 9781420078633, to be published June 2010. Accepted on 5 May 2009.

X. Yan, J. Y. Khan, and B. Jones, “Impact of a Radio Access Network Capacity on the HSDPA Link Performance”, in the Proceeding of the IEEE International Vehicular Technology Conference. (VTC) - Fall, Baltimore, USA, September 2007.

X. Yan, J. Y. Khan, and B. Jones, “Study of Interdependency between the HSDPA Air Interface and the Radio Access Network Design”, in the Proceeding of the IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), Athens, Greece, September 2007.

X. Yan, J. Y. Khan, and B. Jones, “An Adaptive Radio Access Network Resource Management Technique for a HSDPA Network”, in the Proceeding of the IEEE International Symposium on Personal Indoor and Mobile Radio Communications (PIMRC), Athens, Greece, September 2007.

X. Yan, J. Y. Khan, and B. Jones, "An Adaptive Resource Management Technique for a HSDPA Network", in Proceeding of the IEEE International Conference on Wireless and Optical Communications Networks (WOCN), Singapore, July 2007.

X. Yan, J. Y. Khan, and B. Jones, "Performance Evaluation of the UTRAN Transport Network for the High Speed Downlink Packet Access", in the Proceeding of the Australian Telecommunications Networks and Applications Conf. (ATNAC), Melbourne, Australia, Dec. 2006.

Chapter 2

UMTS Radio Access Network Architecture, Protocol and QoS

2.1 Introduction

With the rapid deployment of high-speed UMTS cellular networks, mobile telecommunications user are experiencing their real taste of faster, user-friendly multimedia services [2]. Unlike the second generation networks such as GSM which was initially designed for efficient delivery of voice services, the UMTS network is designed to provide flexible delivery of enhanced multimedia services without requiring network optimization for any particular service [3]. The UMTS network supports simultaneous voice and data capabilities with QoS differentiation for a high efficiency of service delivery. The UMTS network architecture should be able to support existing and future services, and adapt to technology changes and evolution of services. To cater for these very demanding requirements an advanced UMTS network architecture has been developed where access and transport technologies, connection control and user applications are separated from each other in different planes [11]. In this architecture user is allowed to negotiate a bearer characteristics that are most appropriate for supporting a particular type of information. To meet different

QoS requirements, the 3GPP standardized four classes of bearer services: conversation, streaming, interactive and background classes.

The chapter is organized as follows: section 2.2 overviews the UMTS system architecture and describes the key network components. Section 2.3 describes the UTRAN structure, outlines the RAN entities. Section 2.4 introduces relevant UTRAN protocols operating both in control and user planes. The UTRAN air interface protocols are described in section 2.5. Section 2.6 describes the UMTS QoS Classes. Section 2.7 summarizes this chapter.

2.2 UMTS System Architecture Overview

The main components of the UMTS system are the UTRAN (UMTS Terrestrial Radio Access Network), a CN (Core Network) and UEs (User Equipment). The system architecture is shown in Figure 2.1. The UTRAN handles all radio related functionalities such as access control, mobility management, radio resource management, broadcast and multicast services, and other necessary functionalities. The UE is the radio terminal used for communication over the Uu interface commonly known as the air interface. The UE holds the subscriber identity, performs and stores authentication information, and cipher keys and subscription related information needed in a terminal.

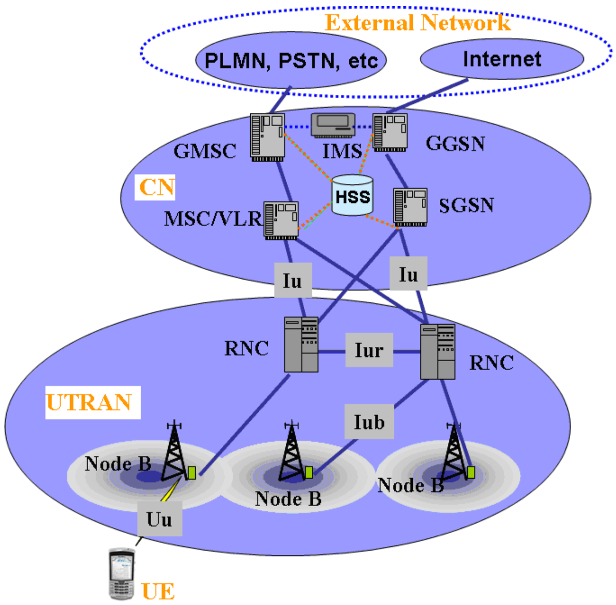


Figure 2.1 The UMTS System Architecture

A Core Network is responsible for switching and routing of calls and data connections to external networks such as the PSTN (Public Switched Telephone Network) and the Internet. A Core Network handles both circuit switch (CS) and packet switch (PS) services. It includes the physical equipments that support network features and provide telecommunications services. It is also responsible for users' mobility and location management at a higher level by using a number of databases. The Home Subscriber Server (HSS) is the master database where the subscription-related information is stored to support network entities to support calls and/or data sessions. The Visitor Location Register (VLR) is the database that contains information about all roaming mobile subscribers that visiting a particular MSC or a SGSN service area. The Mobile-service Switching Centre (MSC) performs all necessary functions to handle circuit switched services to and from the mobile stations. The Gateway MSC (GMSC) is the switch that connects to an external CS network such as the PSTN. The GMSC interrogates the HLR and then route the call from an external network to the MSC where the UE is located. The Serving GPRS (General Packet Radio Service) Support Node (SGSN) handles packet switched (PS) services to and from UEs which are located within the SGSN service area. The Gateway GPRS Support Node (GGSN) interfaces to an external IP packet network, and handles the UMTS session management and communication setup with the external network, etc [12]. The IP Multimedia Subsystem (IMS) provides a standardized way to deliver IP-based multimedia service such as VoIP, multimedia over IP (MMoIP), etc. [13].

The UMTS standards don't specify internal functionalities of network elements in detail. Instead, the interfaces between the logical network elements have been defined in order to make it as open interface [3]. The U_u interface is the WCDMA radio interface, through which the UEs access the fixed part network of the system. The I_u interface connects the UTRAN to a CN. The I_u interface can have two different physical instances, the I_u-CS and the I_u-PS. The I_u-CS connects the radio access network to a circuit-switched core network via an MSC. The I_u-PS connects the access network to a packet-switched core network via an SGSN. The I_{ur} interface support softer handover between RNCs. This interface can support the exchange of both signalling information and user data. Several Node-Bs can have active connections with the same UE at the same time in a WCDMA network. It is possible these Node-Bs could be controlled by different RNCs. Without an I_{ur} interface, it would be a very clumsy method to control this situation via the I_u interface (i.e, via the MSC). The I_{ub} interface is fully open interface that set up connection between a Node-B and a RNC. The I_{ub} separates a Node-B from a RNC so that none of their internal details are visible over the interface, as this could limit the future expandability of this technology. The RNC manages a Node-B over the I_{ub} interface, including the I_{ub} interface transport resource management, operation and maintenance of the Node-B, traffic management of dedicated channels, common channel and shared channels, system information management, and timing and synchronization management, etc. [14]

2.3 UTRAN Architecture

Like most of mobile networks, the network architectures of the UMTS is split between an access network and a core network. The access network is specific to the access technology being deployed, whereas the core network is shielded from the vagaries of the access technology and should ideally be able to handle multiple different access networks [15]. The access network of UMTS is known as UTRAN. The UTRAN consists of one or more Radio Network Sub-systems (RNS). An RNS includes one Radio Network Controller (RNC) and one or more Node-Bs.

2.3.1. Radio Network Controller

A RNC controls radio resources of the UTRAN. It manages the radio network resources, mobility and the user data transport using Radio Access Bearers between UEs and the CN. Radio Resource Control (RRC) protocol that defines the messages and procedures between UE and UTRAN is terminated at the RNC [3].

When an UE to UTRAN connection uses resources from more than one RNS, the involved RNC offers two separate logical roles. It acts as a serving RNC as well as a Drift RNC. The serving RNC of a mobile is the network entity that terminates both the I_{ub} link for user data transport and its corresponding Radio Access Network Application Part (RANAP) signalling to and from the Core Network. It also terminates the radio resource control signalling between the UE and the UTRAN. It performs the data link layer (L2) procedures to set up a bidirectional data flow on the radio interface. A serving RNC performs basic radio resource management operation, such as the mapping of Radio Access Bearer parameters on the air interface transport channel parameters, the handover decision, and the outer loop power control.

A Drift RNC does not perform the L2 processing of the user plane data rather it routes the data transparently between the I_{ub} and the I_{ur} interface, except when the UE is using a common or shared transport channel. One physical RNC normally contains all serving RNC and drift RNC functionalities.

2.3.2. Node-B

The Node-B performs the air interface physical layer (L1) processing such as channel coding and interleaving, rate adaptation, spreading, etc. The Node-B also performs some basic radio resource management operation such as the inner loop power control, mapping logical resources onto hardware resources, and transmitting system information message according to scheduling parameters given by the RNC, etc. In addition, the air interface physical layer is terminated at the Node-B, functions related to air interface are also processed at the Node-B, such as transport channel error detection, encoding and decoding, multiplexing and demultiplexing, rate matching, and RF processing etc. In 3GPP release 5, the Node-B functionalities have been enhanced to support High Speed Downlink Packet Access (HSDPA) traffic. Enhancements such as fast scheduling and fast physical layer hybrid automatic repeat request (H-ARQ) techniques are used at the Node-B to improve the throughput and delay. The relevant functionalities of these entities are described in Chapter 3.

2.4 UTRAN Protocols

The general protocol structure of the UTRAN interfaces is depicted in Figure 2.2. The structure is based on the principle that the layers and planes are logically independent of each other. This makes it easy to change protocol stacks and planes to fit future requirements of different services [3], [16],[17].

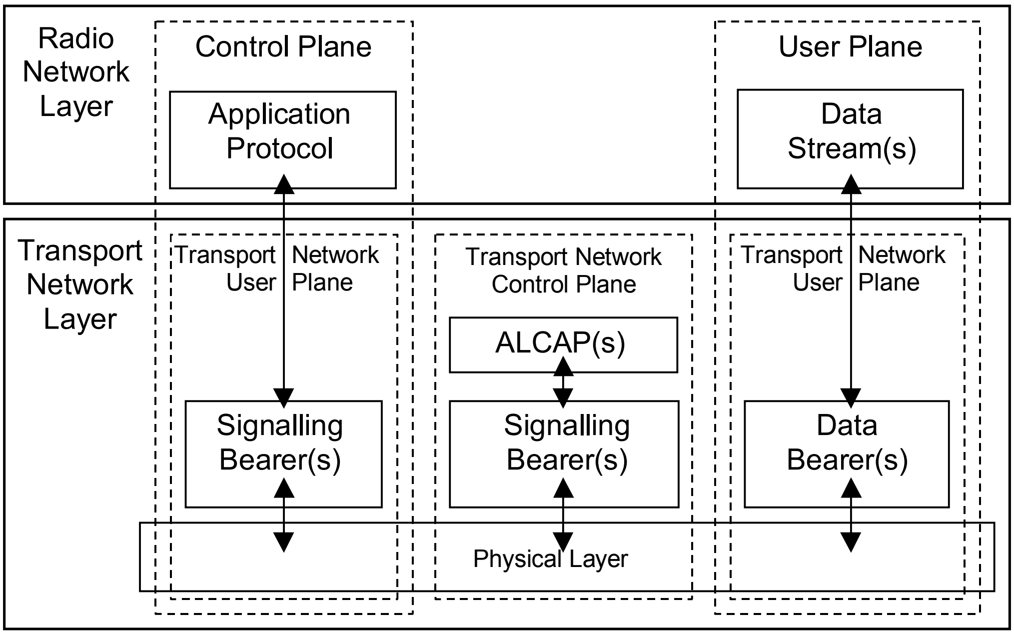


Figure 2.2 General protocol structure for UTRAN interfaces [17]

From the horizontal layer point of view, the general protocol structure consists of two main layers which are the Radio Network Layer (RNL) and the Transport Network Layer (TNL). All UTRAN related issues are located in the Radio Network Layer. The Transport Network Layer only deploys standard transport technologies such as ATM, IP etc, which could be selected for the UTRAN. There shouldn't be any specific modifications for these standard transport technologies to support the operation of the UTRAN. The TNL only offers a transport means for above the RNL. From the TNL point of view, all packets from the RNL are only the payload of the TNL.

From vertical layer point of view, the protocol structure includes a Control Plane, a User Plane, a Transport Network Control Plane and a Transport Network User Plane. The control plane is used for all UMTS-specific control signalling. It includes the Application Protocol, such as the Radio Access Network Application Part (RANAP), the Radio Network Subsystem Application Part (RNSAP) or the Node-B Application Part (NBAP), and the Signalling Bearer for transporting the Application Protocol message. The Application Protocol is used for a number of tasks including setting up bearers for the UE (i.e. the Radio Access Bearer on the I_{ub} interface and subsequently the Radio link on I_{ur} and I_{ub} interfaces).

The user Plane is used to transport all user information such as the coded speech blocks in a voice call or the packets of an Internet connection. The User Plane includes the Data Stream(s), and the Data Bears(s) for various connections. The Data Stream is characterized by one or more frame protocol specified for the interface, for example, the dedicate channel (DCH) frame protocol delivers transport block set (TBs) and outer loop power control messages, across the I_{ub} interface between the RNC and the Node-B.

The Transport Network Control Plane is only used for all control signalling within the Transport Layer. It doesn't include any Radio Network Layer information because the TNL adopts standard transport technologies, provides transport services for the RNL. It employs the ALCAP protocol to set up the transport bearers (Data Bearer) for the User Plane. It also includes the Signalling Bearer need for the ALCAP. The ALCAP only be used for Data Bearers, which need control signalling to set up. Transport Network Control Plane is not needed at all when an IP based link is used between two UTRAN nodes or between a UTRAN node and a CN node since there is no ALCAP signalling transaction for the IP transport.

The Transport Network User Plane provides Data Bear(s) in the User Plane, and the signalling Bearer(s) for the Application Protocol. As described in the previous sections, the Data Bearers in a Transport Network User Plane are directly controlled by the Transport Network Control Plane during real time operations, but the control actions required for setting up the Signalling Bearer(s) for Application Protocol are considered for only O&M actions.

Figure 2.3 shows a simplified view of the UTRAN protocol architecture [17], [18], [19], [20]. This figure only illustrates the User Plane in a packet switched domain, circuit switched domain isn't included. In the figure the G_n interface between GGSN and SGSN has been omitted; both GGSN and SGSN protocols in CN are combined for the figure clarity. When an internet node routes an IP datagram to the UMTS network, the GGSN process these datagram, and establish a Packet Data Protocol (PDP) context between an UE and the GGSN for the data transfer. The GTP-U protocol provides connectionless data transfer services for upper layers and allows multi-protocol user data packets to be tunnelled across the I_u-PS, G_n and G_p interfaces. Encapsulation and tunnelling mechanisms make it is possible to transfer user data packets, although different routing protocols are used in the packet domain of an IP based backbone network. The GTP-U multiplexes packets received from one interface (e.g., G_i at the GGSN) and addressed to different UEs [11]. The tunnel of User Plane Part of GPRS Tunnelling Protocol (GTP-U) is built up between the GGSN and the SGSN, the SGSN and the RNC separately. Packets are tunnelled from the GGSN to the RNC. Data Convergence Protocol (PDCP) performs a compression function on TCP/IP and on RTP/UDP/IP headers. The Radio Link Control (RLC) protocol provides segmentation and retransmission services for user data [21]. The Medium Access Control (MAC) layer offers services to upper layers in the form of data transfer. More details of RLC and MAC layer functionalities are described in the following section. From the user plane point of view, the I_{ub} interface can be seen as an air interface prolongation [18]. The Frame Protocol (FP) and its underneath transport network are responsible for MAC protocol data units (PDUs) transport between the RNC and the Node-B over the I_{ub} interface. These packets are subsequently sent over the air interface from the Node-B to UEs. In an UE the PDCP decompresses packets headers and send packets to upper application layer.

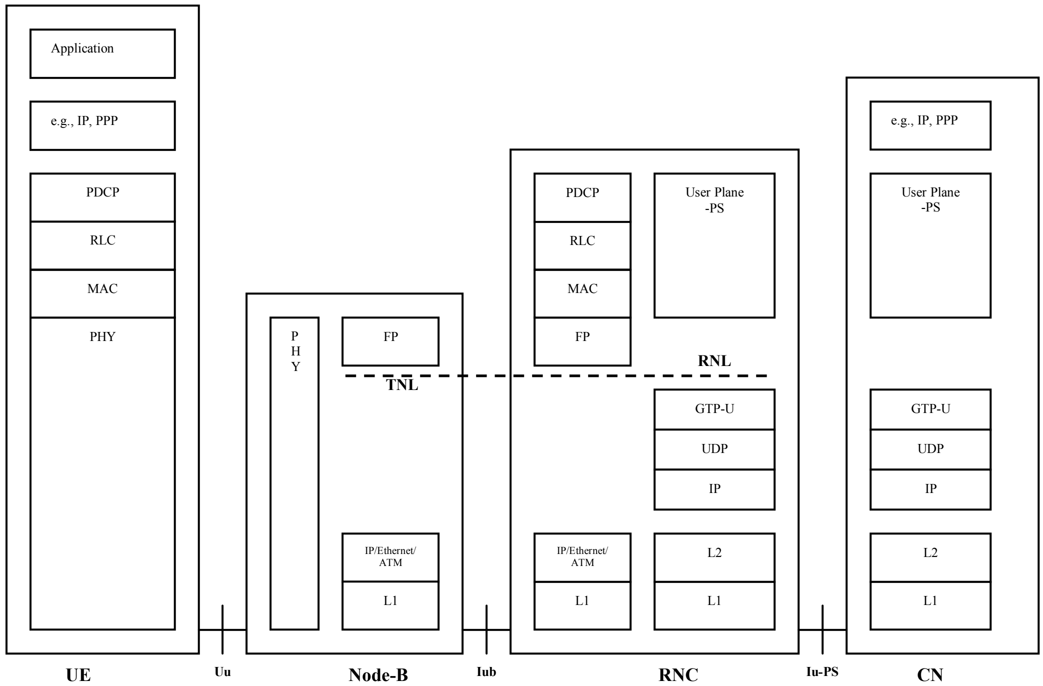


Figure 2.3 The UTRAN protocol architecture (User Plane – PS)

2.5 Radio Interface Protocol

The radio interface protocols are deployed to set up, reconfigure and release the radio bearer service. The overall radio interface protocol architecture is shown in Figure 2.4. The radio interface is layered into three protocol layers: the physical layer (L1), the data link layer (L2) and network layer (L3) [22]. The layer 2 is split into several sub-layers which are MAC, RLC, PDCP and Broadcast/Multicast Control (BMC). The layer 3 consists of the Radio Resource Control (RRC) functionalities which are located in the control plane. The physical layer offer services to the MAC layer via transport channels. In the following section RLC and MAC protocols are described in detail.

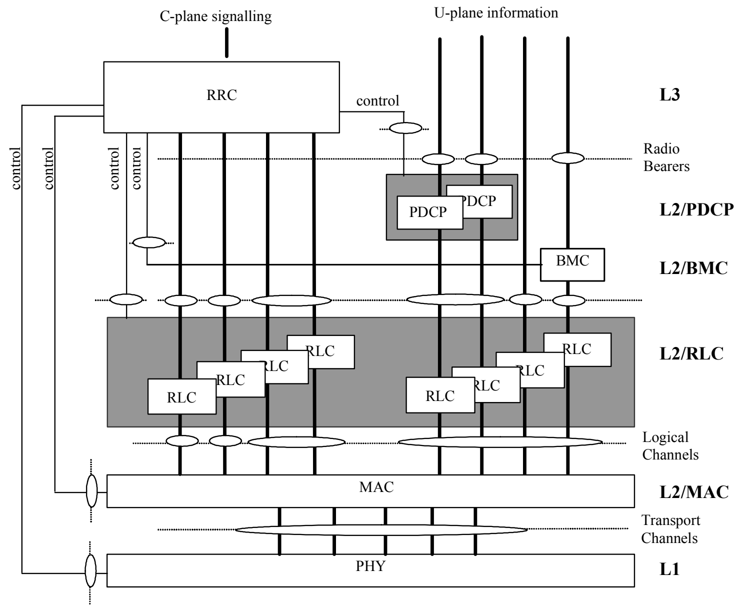


Figure 2.4 Radio interface protocol architecture [22]

2.5.1. RLC Protocol

The RLC protocol provides segmentation and retransmission services for both user and control data [21]. In the control plane, the RLC services are used by the RRC layer for signalling transport. The service is known as the Signalling Radio Bearer (SRB) service. In the user plane, the RLC services are used either by the service-specific protocol layers such as PDCP or BMC, or by other higher layer user plane functions, e.g. a speech codec. The PDCP is only used for the PS domain service, performs a headers compression function for TCP/IP and RTP/UDP/IP packets. The BMC is used to convey messages over the radio interface originating from the Cell Broadcast Centre [3]. In the user plane, the Radio Bearer (RB) service is provided by the PDCP or the BMC. The RLC layer provides the RB service when PDCP and BMC aren't used, such as voice packets in the CS domain.

In the RLC layer, the data from upper layer are formed into a RLC Service Data Unit (SDU). The RLC performs the segmentation/concatenation of the received data. The RLC header is added with the SDU to generate the RLC PDU (not in the Transparent Mode). The remainder of the data field is filled with padding bits when concatenation is not applicable and the remaining data to be transmitted does not fill an entire RLC PDU. The MAC layer adds a MAC header and sends to the physical layer below. If the MAC SDU is transferred over the DCH channel then no logical channel multiplexing is performed. In this case no MAC header is added because data from a single source is transmitted. In the receiver side, the RLC will disassemble the RLC PDU into a RLC SDU and pass on to the higher layer. Any duplicated received RLC PDU is detected by the receiver and ensures that a RLC PDU is received only once. A flow control function allows a RLC receiver to control the rate at which the peer RLC transmitting entity may send information. Protocol error detection and recovery functions detect and recover from errors to maintain the operation of the RLC protocol.

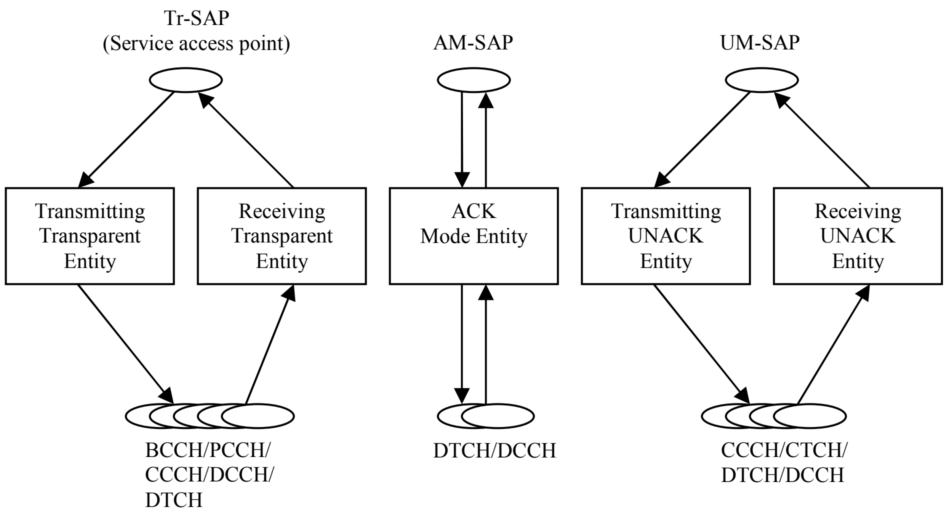


Figure 2.5 The RLC layer architecture [3]

The RLC protocol can be configured to work in three modes. These modes are the Transparent Mode (TM), the Unacknowledged Mode (UM), and the Acknowledged Mode (AM) [21]. Figure 2.5 shows the RLC layer architecture. Ciphering is performed in the RLC layer for AM and UM modes to keep the confidentiality of communication. For all RLC modes, the CRC error detection is performed at the physical layer, and the result of the CRC check is delivered to the RLC layer along with the actual data.

For the Transparent Mode transmission, no RLC protocol overhead is added with the higher layer data. At the receiver erroneous PDUs are detected using the CRC coding at the physical layer, and these PDUs can be discarded or marked erroneous. These features allow the TM mode to be used for voice and circuit switched data where the delay should be as low as possible.

In the Unacknowledged Mode no retransmission protocol is used and data delivery is not guaranteed. Received erroneous data is either marked or discarded depending on the configuration. On the transmitter side, RLC SDUs that are not transmitted within a specified time are simply removed from the transmission buffer and dropped. In this mode sequence numbers is included in the PDU structure in order to observe the integrity of higher layer PDUs. An RLC entity in the Unacknowledged Mode is configured in a unidirectional mode. There are no association between uplink and downlink parameters. The Unacknowledged Mode is deployed in cell broadcast services and voice over IP (VoIP) etc.[3]

In the RLC Acknowledge Mode, error correction and in-sequence delivery of higher layer PDUs function is used in the AM mode to maintain integrity of a connection. An Automatic Retransmission Request (ARQ) mechanism is used as one of the error correction measure. The performance (link quality vs. delay) of the RLC layer can be controlled by RRC algorithms through configuration of the number of retransmissions provided by the RLC layer. The AM is the normal mode for packet-type services such as Internet browsing and email transfer. It is also used for signalling, when it is important that the signalling packet is received correctly but without any delay restrictions. The 3GPP has proposed the use of Multiple Rejects (MR) ARQ as one of the RLC layer procedure. The MR mechanism uses a poll-status mechanism with a timer control to reduce the number of status packets [23] . The Acknowledged Mode Data (AMD) PDUs are placed in the retransmission buffer. AMD PDUs are buffered in the retransmission queue are either deleted or retransmitted based on the status report from the receiver. The transmitter deletes all positively acknowledged AMD PDUs and retransmits the negatively acknowledged AMD PDUs only.

2.5.2. MAC Protocol

A standard MAC layer offers services to upper layers in the form of data transfer. The UMTS MAC layer provides data transfer services using logical channels which maps on transport channels provided by the physical layer [24]. The data transfer service provides an unacknowledged mode of transfer of MAC SDUs between peer entities. This service does not provide any data segmentation which is done at the RLC layer. The MAC layer multiplexes upper layer PDUs into transport blocks and delivered to the physical layer using common transport and dedicated transport channels. Similarly the MAC layer demultiplexes PDUs for higher layer using received information. On the request of the RRC (Radio Resource Control) procedures the MAC layer reallocates radio resources and changes MAC parameters such as channel type, data flow handling priority, etc. The MAC layer also performs local measurements such as traffic volume. The MAC layer compares the amount of data corresponding to a transport channel with the thresholds set by the RRC layer, and report to the appropriate RRC entity. The RRC layer uses these reports for triggering reconfiguration of Radio Bearers and/or Transport Channels. [3]

Each RLC PDU is mapped onto a Transport Block. In an UE for the uplink, all MAC PDUs are delivered to the physical layer within one TTI are defined as the Transport Block Set (TBS). The TBS consists of one or several Transport Blocks, each containing one MAC PDU. The transport Blocks shall be transmitted in the same order as it was delivered from the RLC layer. When multiplexing of RLC PDUs from different logical channels is performed at the MAC layer, the MAC layer should maintain the transmission sequence which will be same as the RLC delivery sequence. The order of transmission of the different logical channels in a TBS is set by the MAC protocol.

A MAC PDU consists of an optional header and a MAC SDU supplied by the RLC layer. Both the header and the MAC SDU are of variable size. The content and the size of the MAC header depend on the type of logical channel used. Some cases for example; when DCCH and DTCH channels are mapped on to DCHs channel without multiplexing, none of the parameters in the MAC header are needed. The size of the MAC-SDU depends on the size of the RLC-PDU, which is defined during a call set up.

2.5.3. Channels Mapping

The UMTS standard deploys a number of logical and transport channels to deliver user and control traffic. Channel mappings are essential to optimise transmission resources and to maintain QoS of various service requirements. Packets are passed from the RLC layer to the MAC layer in the form of logical channels. The logical channels are mapped to transport channels, which in turn are mapped to physical channels.

In the UMTS Logical channels are generally classified into two groups. One set of channels are grouped under Control Channels and the other set of channels are grouped as Traffic Channels. Control Channels include Broadcast Control Channel (BCCH), Paging Control Channel (PCCH), Dedicated Control Channel (DCCH), and Common Control Channel (CCCH). Traffic Channels include Dedicated Traffic Channel (DTCH) and Common Traffic Channel (CTCH). A DTCH is a point-to-point traffic channel, which is dedicated to one UE to transfer user information on both the uplink and the downlink. A CTCH is a point-to-multipoint downlink traffic channel, which is used to transfer dedicated user information for all, or a group of specified users.

In UTRAN the data generated at higher layers is transmitted by transport channels after mapping logical channels on different physical channels in the physical layer which are carried over the radio link. The interface between the physical layer and the MAC layer is comprised of a number of transport channels. Packets from the RLC layer are passed to the MAC layer in the form of logical channels, and then these logical channels are mapped on transport channels.

Transport channels are grouped in to dedicated channels and common channels groups. For a dedicated channel transmission resources such as code or timeslot or frequency are allocated to each UE. Whereas common channel resources are shared among number of users using MAC channel allocation technique.

The only dedicated transport channel is the Dedicated Channel (DCH). The DCH carriers all the information intended for the given user coming from layers above the physical layer, including data for the actual service and higher layer control information.

There are six different common transport channel types [15]:

Broadcast Channel (BCH), which is used in the downlink to transmit system information over the entire coverage area of a cell.

Forward Access Channel (FACH), which is used to send downlink control information to one or more users in a cell.

Paging Channel (PCH), which used in the downlink to page a given UE when the network wants to initiate communication with the user.

Random Access Channel (RACH), which is used in the uplink when a user wants to gain access to the network.

Uplink Common Packet Channel (CPCH), which is similar to the RACH but can last for several frames.

Downlink Shared Channel (DSCH), which is used to carry dedicated user data or control signalling to one or more users in a cell.

There are numerous options for mapping between logical channels and transport channels. The mapping depends on a range of criteria such as the types of information to be sent, whether it is to be sent to multiple UEs (in the downlink), and whether the UE has already an established connection with the network, etc. [15]

Figure 2.6 shows the mapping between logical channels and transport channels. The PCCH is connected to PCH. BCCH is connected to BCH and may also be connected to FACH. DCCH and DTCH can be connected to RACH, CPCH and DCH in the uplink. These two logical channels are connected to FACH, DSCH and DCH in the downlink. CCCH is connected to RACH and FACH in the uplink and the downlink respectively. CTCH is connected to FACH.

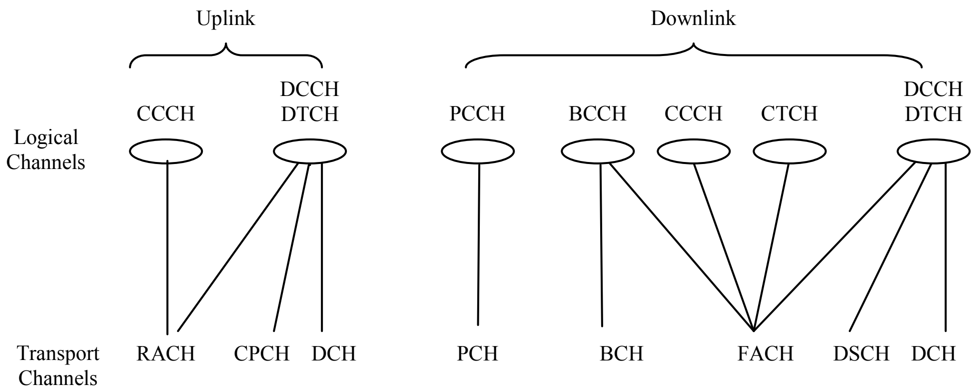


Figure 2.6 Mapping between logical channels and transport channels [3]

As mentioned, packets from upper layers are delivered to the physical layer through a number of transport channels. These transport channels are mapped on to a number of physical channels on the air interface. The physical layer is required to support variable bit rate transport channels to offer flexible bandwidth on demand, and to multiplex several services on to one connection [3]. In general, a physical channel is identified by a specific frequency, scrambling code, channelization code, duration, and, in the uplink, phase. In addition to those physical channels that are mapped to or from transport channels, a number of physical channels exist only for the correct operation of the physical layer. Such channels are not visible to higher layers. At the physical layer, the UE communicate with UTRAN via a number of physical channels. Many of these physical channels are used to carry packets that are passed to the physical layer from higher layers.

Transport Channel	**Physical Channel**
BCH	Primary Common Control Physical Channel (PCCPCH)
FACH	Secondary Common Control Physical Channel (SCCPCH)
PCH	
RAC	Physical Random Access Channel (PRACH)
DCH	Dedicated Physical Data Channel (DPDCH)
	Dedicated Physical Control Channel (DPCCH)
DSCH	Physical Downlink Shared Channel (PDSCH)
CPCH	Physical Common Packet Channel (PCPCH)
	Synchronization Channel (SCH)
	Common Pilot Channel (CPICH)
	Acquisition Indication Channel (AICH)
	Paging Indication Channel (PICH)
	CPCH Status Indication Channel (CSICH)
	Collision Detection/Channel Assignment Indicator Channel (CD/CA-ICH)

Figure 2.7 Mapping between transport channels and physical channels [25]

There are several physical channels to carry only information relevant to physical layer procedures. The Synchronisation Channel (SCH), the Common Pilot Channel (CPICH) and the Acquisition Indication Channel (AICH) are transmitted from every base station. These Channels are not directly visible to higher layers and are mandatory from the system function point of view. [3]

A DCH is mapped into two physical channels: DPDCH and DPCCH. The DPDCH carries higher layer information. The DPCCH carries the necessary physical layer control information

The transport channel to physical channel mapping process is illustrated in Figure 2.7.

The DCH exists in the uplink and the downlink is mapped on to the physical channels DPDCH and DPCCH. In the uplink, the combination of frequency, the scrambling code, the channelization code, and the phase is used to indicate a particular DPDCH or DPCCH at the physical layer. In the downlink, the DCH is mapped to a DPCH, which is identified in the downlink by a particular channelization code. The downlink DPDCH and DPCCH are time multiplexed onto the downlink DPCH. The data rate on a DCH can vary on a frame-by-frame basis. [15]

2.6 Quality of Service

In the context of UMTS networks, Quality of Service (QoS) refers to the collective effect of service performance that determines the degree of satisfaction of the end-user of the service [26]. The QoS provision is crucial for successful delivery of a variety of packet data services in a mobile communication network due to the bandwidth-constraint and error-prone transmission environment. The UMTS QoS architecture covers all traffic types, including Non-Real Time and Real Time services [27]. Non-real time services include different type of data transfer services including background services, email transfer, file transfer services, etc. On the other end real time services include audio and video transmissions including the streaming services. It is important for the real time services to maintain the temporal relationship between data samples. QoS requirements of these traffic classes can be defined in terms of delay, packet losses, delay jitters, etc. According to the service QoS requirement, the 3GPP classify these services into four different QoS classes (also refer to as traffic classes): conversation, streaming, interactive and background classes [28].

Among a number of parameters the delay sensitivity is one of the main distinguishing QoS feature for different class of traffic.

2.6.1. QoS Architecture

A bearer service with clearly defined characteristics and functionality is needed to be set up from a source to a destination with a certain network QoS for an end-to-end service. Figure 2.8 shows a UMTS bearer service layered architecture [29]. The figure shows that the end-to-end bearer service has three main components. Those components are the local bearer service to support the TE/MT link, the UMTS bearer service which supports radio and core network services, and the last component is the external bearer services to interface with external network. A bearer service includes all aspects to enable the provision of a contracted QoS. These aspects include the control signalling, user plane transport and QoS management functionalities. By combining services provided by the layers below, each bearer service on a specific layer offers its individual services.

Within the UMTS network the UMTS bearer service consists of the Radio Access Bearer (RAB) service and the Core Network (CN) Bearer service. The UTRAN are responsible for the Radio Bearer service and the I_u bearer service. The SGSN, GGSN and the possible IP-based network between those network elements provide the CN Bearer service.[30]

In the PS domain the UMTS bearer service is realised by a packet data protocol (PDP) context. Every PDP context has a set of QoS attributes associated with it. The set of QoS attributes is referred to as a QoS Profile, which includes all essential parameters to describe the QoS need of its media streams for an application, such as traffic classes, target transfer delay, reliability, guaranteed bit rate, priority, etc. The QoS profile is negotiated during the PDP context activation procedure. It describes the quality of the UMTS bearer service offered to the user. To enable a multimedia session, in which there may be variable QoS requirement for each individual media streams such as voice and video traffic. The 3GPP supports multiple PDP contexts of the same IP address with different QoS requirements. Each PDP context corresponds to an individual UMTS bearer.

Based on the request by application, UE maps the QoS requirements to relevant PDP context QoS parameters. UE requests a PDP-context activation from the SGSN. The SGSN first checks whether the subscriber is entitled to use such a QoS profile. If so, and there are adequate resources available for the SGSN then the SGSN will signal the RAN for a Radio Access Bearer (RAB) establishment, based on the call QoS request. A RAN carries out internal admission control and resource reservation procedures. Once a call is successfully set up, a RAB will be set up. The SGSN makes a PDP-context activation request to the GGSN. If there are adequate resources for GGSN to offer the service with requested QoS, a transport connection (e.g. a tunnel) is set up between the SGSN and GGSN, and a PDP context with the requested QoS capabilities is established between the UE and the GGSN. [30]

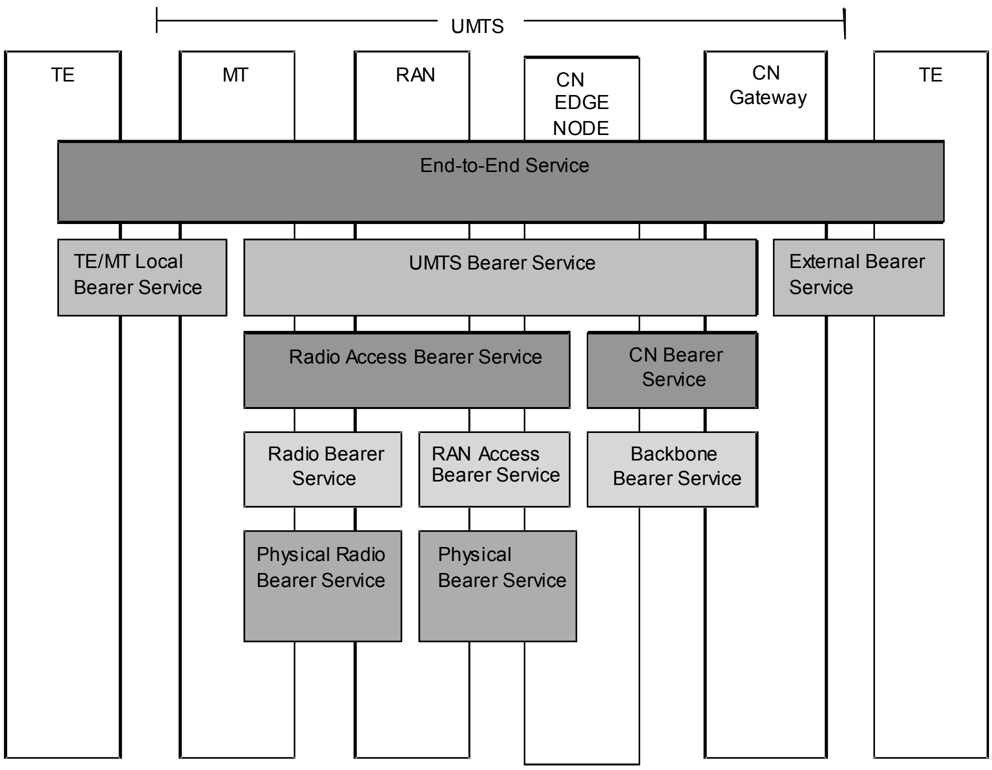


Figure 2.8 End to end QoS architecture [29]

2.6.2. UMTS QoS Classes

The restrictions and limitation of the air interface have to be taken into account when defining the UMTS QoS classes. The QoS mechanisms provided in the cellular network have to be robust and capable of providing reasonable QoS resolution. There are four different QoS classes defined by the UMTS service categories [31].

- Conversational Class;
- Streaming Class;
- Interactive Class;
- Background Class.

UMTS QoS classes are illustrated in Table 2.1.

Table 2.1 Traffic classes in UMTS

Traffic Class	Fundamental characteristics	Application Examples
Conversational class Real time conversation	Preserve time relation (variation) between information entities of the stream Conversational pattern (stringent and low delay)	Voice
Streaming class Real time stream	Preserve time relation (variation) between information entities of the stream	Streaming video
Interactive class Interactive best effort	Request response pattern Preserve payload content	Web browsing
Background class Background best effort	Destination is not expecting the data within a certain time Preserve payload content	Background download emails

Conversation and streaming classes are mainly intended to be used to carry real time traffic flows. The main difference between them is the delay sensitivity. The maximum transfer delay for the real time conversation is controlled by the human perception of video and audio conversation. Failure to provide low and bounded transfer delay will results in an unacceptable link quality. A real time streaming service requires a one way transport channel. The delay variation of the end-to-end stream flow shall be limited. The highest acceptable delay variation over transmission media is given by the capability of the time alignment function of the application.

Interactive and background classes are mainly used by Internet application such as WWW, Email, and FTP etc. Due to looser delay requirement, both provide better error rate by means of channel coding and retransmission. The main difference between interactive and background class is that the interactive class is mainly used by interactive application, e.g. interactive web browsing, while background class is used for traffic such as background download of Emails. Traffic in the interactive class has a higher scheduling priority on the back ground class. Background application use transmission resources only when interactive applications don't need them.

Following QoS attributes of the UMTS bearer service are extensively used for the performance analysis in the thesis [31]:

Transfer Delay: indicates the maximum delay for 95th percentile of the distribution of delay for all delivered SDUs during the lifetime of a bearer service, where delay for a SDU is defined as the time from a request to transfer an SDU at one Service Access Point (SAP) to its destination SAP.

Traffic handing priority: specifies the relative importance for handling of all SDUs belonging to a UMTS bearer compared to the SDUs of other bearers.

SDU error ratio: indicates the fraction of SDUs lost or detected as erroneous. SDU error ratio is defined only for conforming traffic.

Maximum SDU size (octets): the maximum SDU size for which the network shall satisfy the negotiated QoS.

Maximum bit rate (kbps): maximum number of bits delivered by UMTS and to UMTS at a SAP within a period of time, divided by the duration of the period. The maximum bit rate is upper limit a user or application can accept or provide.

Guaranteed bit rate (kbps): describes the bit rate the UMTS bearer service shall guarantee to the user and applications. Guaranteed bit rate may be used to facilitate admission control based on available resources.

2.7 Summary

The present chapter provides a general overview of the UMTS system architecture, UTRAN structure and protocols, and QoS demand for different class of traffic, which are necessary to achieve complete understanding of the analysis and design in following chapters.

Access technology, transport technology, connection control and user application are separated from each other in the UMTS system in order to carrying existing and future services, adapt technology changes and evolution.

The UTRAN protocol structure is based on the principle that the layers and planes are logically independent of each other. This makes it easy to change protocol stacks and planes to fit future requirements.

The radio interface protocols are used to set up, reconfigure and release the radio bearer service. The MAC layer offers services to upper layers in the form of data transfer. The MAC layer provides data transfer services on logical channels, maps between logical channels and transport channels. The RLC protocol provides segmentation and retransmission services for both user and control data. The RLC can be configured to work in Transparent Mode, Unacknowledged Mode, and Acknowledged Mode.

The QoS mechanisms of in the cellular networks are robust and provide reasonable QoS resolution for all type of traffic. To meet service QoS requirements, the 3GPP classify these services into four different QoS classes. Delay sensitivity is one of the main distinguishing factors between these QoS classes.

Chapter 3

High Speed Downlink Packet Access

3.1 Introduction

The WCDMA based third generation radio access technology is now being widely deployed all over the world. Number of mobile users is continuing to grow with an accelerated pace in next decades. The WCDMA technology is continuing to evolve to meet the demand of larger system capacity, higher data rate and improved QoS. As one of the initial evolutionary steps to develop a high data rate network, the 3GPP introduced the HSDPA (High Speed Downlink Packet Access) standard to transmit data and multimedia traffic on the downlink in the release 5 of the 3GPP specification. The HSDPA shared channel uses a short 2 ms transmission time interval (TTI), and uses variable coding and modulation techniques to achieve high data rates. In addition to above techniques the Node-B based fast scheduling and fast physical layer hybrid automatic repeat request (H-ARQ) techniques are also used. Currently the HSDPA networks can offer peak transmission rate of 28.8 Mbps. According to the HSDPA roadmap the air interface data rates could reach up to 40 Mbps within the next few years.

This chapter will provide a brief overview of the HSDPA standard that will be required to comprehensively study interdependence between the HSDPA air interface and its radio access network in following chapters. This chapter is organised as follows. Section 3.2 briefly introduces key features of the HSDPA standard. Section 3.3 presents the overview of the HSDPA protocol architecture. Section 3.4 reviews the HSDPA channel structures. Section 3.5 summarizes various HSDPA terminal capabilities which could be used to support multimedia services. The flow control and congestion control algorithms between the RNC and the Node-B are introduced in section 3.6 separately. Section 3.7 concludes the chapter.

3.2 HSDPA Principle

The HSDPA standard has been developed based on the shared channel concept. A fraction of total downlink resource available within a cell, such as channelization codes and transmission power is seen as a common resource which are dynamically shared among users, primarily in the time domain [32]. An HSPDA connection uses the High Speed Downlink Shared Channel (HS-DSCH) to transmit data from the RNC to a UE. In the Release 99 the DCH channel uses a faster power control to stabilize the received signal E_b/N_o and maintain a constant data rate link. This is a suitable approach for services requiring a constant data rate e.g. the circuit switch voice service. In this connection the transmission power is increased to compensate the fades of the received signal. This causes peaks in the transmission power and subsequently increases the total transmission power, which reduces the total network capacity due to the lower SNR values. The HS-DSCH uses a link adaptation technique to adapt the data rate to the continuously varying channel conditions while keeping the transmission power constant.

A link adaptation algorithm adapts channel code, transmission rate and modulation to match with the instantaneous channel conditions rather than adjusting the transmission power only. An HSDPA connection selects either the 16-QAM or the QPSK modulator based on the channel condition. The 16-QAM modulator is used in favourable channel conditions. In this case a maximum of

15 parallel channels with a constant spreading factor of 16 is used for the HS-DSCH connection. These channels may all be assigned to one user, or may be split among several users.

The radio channel conditions of users could vary independently. The channel dependent scheduling technique allocates channel to a user with favourable transmission conditions. A packet scheduler in close cooperation with a link adaptation mechanism can significantly increase a cell capacity. The gain is known as multi-user diversity gain. The Link adaptation and the channel dependent scheduling algorithms could exploit the rapid channel condition variations. The TTI is shortened to 2 ms to improve the tracking of channel variations and reduce the roundtrip time. The Hybrid ARQ with a soft combining technique is adopted to rapidly retransmit erroneously received transport blocks and combines the soft information with the original transmission with any subsequent retransmission before the decoding the received information. Additional redundant information is incrementally transmitted in subsequent retransmissions i.e. using the Incremental Redundancy technique [33].

The release 6 standard introduces a number of changes compared to the release 99. One of the main changes is to reduce end-to-end delay of a HSDPA connection and rapidly allocate transmission resources. To implement such change the MAC functionalities related to the HS-DSCH channel allocation is moved from the RNC to the Node-B. The HS-DSCH channel exploits the diversity gains through scheduling process instead of a downlink soft handover used by the release 99 based DCH channel. The inter-Node-B soft handover is not possible in absence of any power control. Hence, the HSDPA standard only supports the hard handover between the RNCs.

3.3 HSDPA Protocol Architecture

Similar to Release 99 architecture, HSDPA protocol architecture consists of a control plane part and a user plane part. The radio resource control (RRC) layer in control plane part handles all the signalling such as channel configuration, mobility management and so on. User plane part handles user data transmission. The HSDPA user plane protocol architecture is shown in Figure 3.1 [34]. The

Radio link control (RLC) layer handles the segmentation and retransmission of all user and control data. The RLC layer can only operate either in acknowledged mode (AM) or in an unacknowledged mode (UM) mode. The transparent mode (TM) isn't supported due to ciphering only done in the RLC layer for a HSDPA connection. Even if physical layer retransmission is introduced for a HSDPA connection, the RLC layer will still handle the retransmission caused by the physical layer operation failure or the change of serving HS-DSCH cell. The MAC-d (dedicated medium access control) layer is retained in the RNC to handle transport channel switching. The MAC-hs (hs for high speed) layer is located in the Node-B and responsible for scheduling, priority handling, H-ARQ and selection of an appropriate format and resources. The MAC layer functionalities of HS-DSCH channel are enhanced to support the flow control between the RNC and the Node-B. The HS-DSCH is the transport channel that carries the actual user HSDPA data. The HS-DSCH FP (Frame Protocol) handles all data transport between a RNC and a Node-B. For a HSDPA connection data received from a core network is handled by RLC and MAC-d layers. These layers forms the MAC-d PDUs (Protocol Data Unit) and packs the PDU into HS-DSCH FP frames and transmitted on the I_{ub} link using the ATM, the IP or Carrier Ethernet transport protocols.

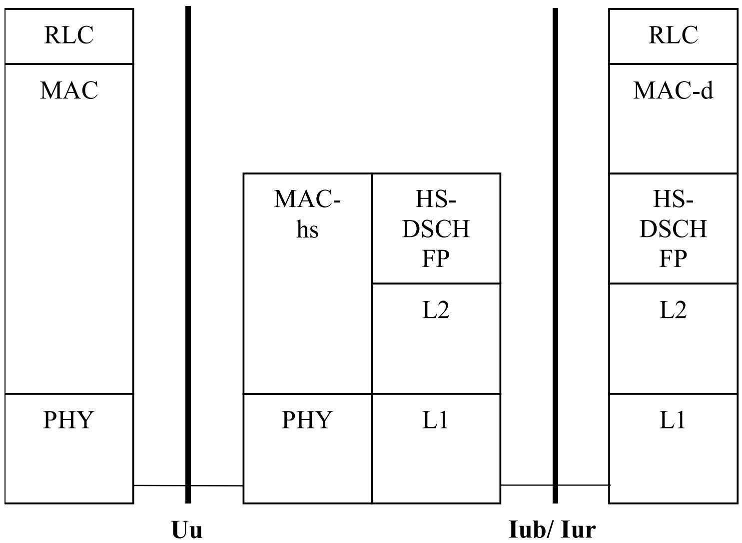


Figure 3.1 HSDPA protocol architecture

3.4 HSDPA Channel Structure

In release 99, the downlink shared channel (DSCH) was introduced to transport downlink bursty packet traffic. The DSCH supports several users using shared common code resources in a time multiplexed manner. An HSDPA connection is implemented using the HS-DSCH channel. The introduction of HS-DSCH channel can be seen as an evolution of the DSCH channel. The DSCH channel has been removed from the Release 5 and onwards. All user data is carried using the HS-DSCH and its corresponding physical channel known as the high-speed physical downlink shared channel (HS-DSCH). A HS-PDSCH channel corresponds to one channelization code of fixed spreading factor SF=16 from the set of channelization codes reserved for the HS-DSCH transmission. A multi-code transmission is supported, which assign a user multiple channelization codes in the same TTI. For the multi-code transmission up to a maximum of 15 codes can be used. Furthermore, multiplexing of multiple users in the code domain within a HS-DSCH TTI is allowed. Beside the HS-DSCH, the high-speed shared control channel (HS-SCCH) on the downlink and the high-speed dedicated physical control channel (HS-DPCCH) on the uplink direction are used to transport associated signalling for HSDPA [6] connections. The HSDPA channel structure is shown in Figure 3.2.

The figure shows that the HS-SCCH channel carries time-critical signalling information that allows the terminal to demodulate the correct code. The HS-SCCH channel has two slots offset compared with the HS-DSCH channel, which allows the information to be available before the commencement of the HS-DSCH channel. The first part of the HS-SCCH data block carries the information needed to de-spread correct codes and modulation information. The second part contains the Hybrid ARQ related information that indicates whether the transmission is a current or a retransmission, etc.

The HS-DPCCH carries the uplink feedback information from the terminal to the Node-B in order to enable the link adaptation and physical layer retransmission. The HS-DPCCH has a three-slot/2ms structure. The fist slot contains the Hybrid ARQ information. The last two slots are used for the CQI use.

Once the Node-B determines to serve a terminal in a particular TTI, the necessary HS-DSCH parameters are identified by the Node-B. The Node-B transmits signalling information in the HS-SCCH two slots before the corresponding HS-DSCH TTI. Once the terminal decodes the first part of an HS-SCCH data block intended for it, it starts to decode remaining parts of the HS-SCCH data block and buffer necessary codes from the HS-DSCH channel. Upon decoding the potentially combined data, the terminal sends an ACK/NACK indicator via the HS-DPCCH channel in the uplink direction.

The Release 5 based HSDPA connection is always operating with the DCH channel in parallel. The uplink user data always is transported using the DCH channel. The signalling radio bearer (SRB) is carried on the DCH channel when only packet data is severed. In Release 6, the SRB can be carried on the HS-DSCH channel. The fraction DCH (F-DCH) is introduced to reduce the code space utilization. The Enhanced DCH (E-DCH) with high-speed uplink packet access (HSUPA) is as an alternative to uplink DCH mainly designed to carry the uplink traffic[6].

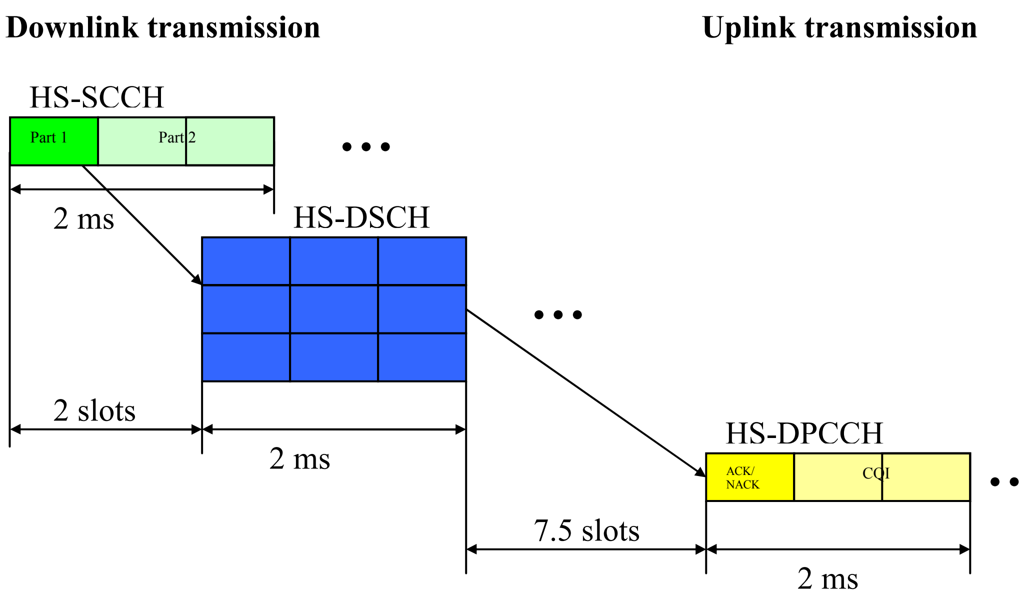


Figure 3.2 HSDPA channel structure

3.5 HSDPA Terminal Capability

In the release 6 of the 3GPP standard HSDPA terminals have been divided into 12 different categories supporting maximum downlink data rates from 0.9 Mbps to 14.4 Mbps. HSDPA terminals must be compatible with Release 99 standard to support any operator's update strategies where operators may gradually update their infrastructure to HSDPA/HSUPA based Node-B by upgrading the legacy base stations. Release 6 specifies that it is mandatory for High-speed uplink packet access (HSUPA) terminals to support HSDPA traffic [6].

Table 3.1 presents HSDPA terminal categories [6]. HSDPA terminal capabilities will depend on a number of factors which includes the modulation technique, maximum number of allocated parallel codes and the minimum inter-TTI interval. The minimum inter-TTI interval value indicates the sustained peak data rate over multiple TTIs. Terminal with inter-TTI interval value 1 TTI can sustain the peak rate over multiple TTIs, while terminal with a value 2 or 3 TTIs must wait for 2 or 4 ms (1 or 2 TTIs) after each received TTI.

Table 3.1 HSDPA terminal capability categories

Category	Max. number of parallel codes per HS-DSCH	Min. inter-TTI interval	Modulation	Max. peak rate
1	5	3	QPSK & 16-QAM	1.2 Mbps
2	5	3	QPSK & 16-QAM	1.2 Mbps
3	5	2	QPSK & 16-QAM	1.8 Mbps
4	5	2	QPSK & 16-QAM	1.8 Mbps
5	5	1	QPSK & 16-QAM	3.6 Mbps
6	5	1	QPSK & 16-QAM	3.6 Mbps
7	10	1	QPSK & 16-QAM	7.3 Mbps
8	10	1	QPSK & 16-QAM	7.3 Mbps
9	15	1	QPSK & 16-QAM	10.2 Mbps
10	15	1	QPSK & 16-QAM	14.4 Mbps
11	5	2	QPSK only	900 Kbps
12	5	1	QPSK only	1.8 Mbps

Table 3.2 RLC layer data rate for different UE categories

Category	RLC blocks per 2ms TTI	Transport	Max. RLC data rate
12	10 X 320	3440	1.6 Mbps
5/6	21 X 320	7168	3.36 Mbps
7/8	42 X 320	14155	6.72 Mbps
9	60 X 320	20251	9.6 Mbps
10	83 X 320	27952	13.3 Mbps

Table 3.2 shows the data rate at the RLC layer for different UE categories when PDU size is fixed to 320 bits and RLC headers are taken into account.

3.6 Flow Control between RNC and Node-B

The HSDPA architecture shows that the MAC layer is split between the RNC and the Node-B. For a downlink connection MAC-d PDUs are generated by the RNC, which are aggregated into HS-DSCH Data Frames, then sent to the Node-B over the I_{ub} link. The HSDPA architecture incorporates a downlink buffer and scheduler in the Node-B to support a high data rate transmission. By employing different radio resource management (RRM) algorithms, the scheduler can be configured to provide users with different peak to average ratios of data rates over the air interface. The PDUs are buffered in Node-B until they are scheduled and successfully transmitted to UE over the air interface. A flow controller is generally employed on the I_{ub} link between a radio network controller (RNC) and a Node-B to avoid buffer overflow or buffer starvation in the Node-B.

To maintain the data flow on the I_{ub} link the RNC sends a capacity request Control Frame to the Node B to request HS-DSCH link capacity allocation, which indicate the user buffer size in the RNC for a given priority level,. The Node-B allocates capacity to the RNC via a capacity allocation Control Frame, which indicates the maximum MAC-d PDU length, number, HS-DSCH interval and the repetition rate. This mechanism is used to control the user data flow on the I_{ub} link [35], [36] . The HS-DSCH data frame is used to transfer user data. According to the allocated capacity by the Node-B via the HS-DSCH CAPACITY ALLOCATION Control Frame, the HS-DSCH data mapped onto the HS-DSCH Data Frame and transmitted immediately from the RNC to the Node-B using the I_{ub} interface.

Users experiencing good radio channel conditions get higher allocated capacity to prevent draining the Node-B buffer because data is transmitted at a higher rate on the downlink. On the other hand the flow control algorithm will slow down the data stream for the user experiencing poor radio conditions to avoid the buffer overflow at the Node-B[6]. Figure 3.3 shows the principle of flow control principle used on the I_{ub} interface. The Node-B requires the RNC to reduce the data rate of the UE1 on the I_{ub} link to prevent Node-B buffer overflow because the UE1 is experiencing poor radio channel conditions, thus reducing the transmission rate from the Node-B to the UE1. The Node-B requests the RNC to increase the data rate for UE2 to cope higher transmission rate on the downlink because the UE2 is experiencing a good radio channel condition.

There are conflicting goals for flow control strategies [37]. To minimize latency, data should not be held at the RNC if there is a possibility of it being scheduled by the Node-B. On the contrary, due to the absence of any soft handover for HSDPA connections, all MAC-d PDUs are held in the previous Node-B are lost when the UE makes a hard handover to a new Node-B. This causes the RLC layer retransmission to recover lost data (in the Acknowledge Mode). Due to the handover and retransmission processing latency is increased. So, the number of data packets in the Node-B buffer should be optimised. That Node-B buffer could have limited resources which is a case for using an advanced flow control algorithm. Delivering the flow control signalling consumes significant I_{ub} link bandwidth, hence, the volume of flow control signalling should be minimized to maintain high data throughput on the I_{ub} link.

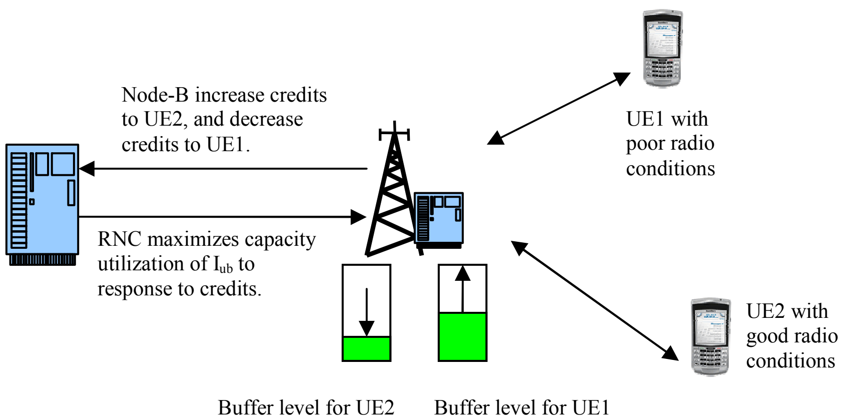


Figure 3.3 HSDPA flow control on the I_{ub} interface

A HSDPA flow control algorithm regulates traffic volume over a sampling time by measuring the Node-B buffer size. The available resource on the I_{ub} link isn't considered, which increases the risk of congestion when HSDPA data rate is high. The 3GPP release 7 introduces I_{ub}/I_{ur} congestion control features, which combines the flow control algorithm with congestion control functionalities to improve the I_{ub} link congestion handling performance [38]. A congestion control algorithm consists of congestion detection and congestion reduction processes.

The Node-B can detect delay build-ups by noting the arrival time of subsequent Delay-Reference-Time (DRTs) in the FP frame and comparing them. A delay build-up indicates that frames are being queued due to overload in the I_{ub} link. A frame lose is detected with the sequence number in a FP frame. A frame loss indicates that the I_{ub} link overloading could lead to packets losses. When the Node-B has detected that there is a congestion situation in the I_{ub} link, the RNC is needed to reduce the bit rate of the MAC-d flow. But the purpose of the congestion control introduced by [38] is not act as a flow control rather as an "emergency break" in order to keep the system at a stable state.

3.7 Summary

This chapter has presented a brief overview of technologies adopted by the HSDPA standard to significantly enhance the UMTS network performance. Discussion showed that network capacity and QoS can be increased by using the channel dependent fast scheduling technique in close cooperation with a link adaptation mechanism to exploit favourable radio channel conditions. The use of short TTI could improve the tracking of channel condition variations and reduce the roundtrip time. The Hybrid ARQ with soft combining can rapidly retransmits erroneously received transport blocks and combines the soft information from original transmission with any subsequent retransmission before the decoding process. Discussion also showed that the flow control algorithm plays a key role in maintaining the QoS of a UTRAN. An efficient flow control algorithm can enhance the radio resource usage as well as improve the HSDPA connection usages. Improvement of the HSDPA throughput and efficient resource utilisation is one of the main focuses of this thesis which will be further discussed in following chapters.

Chapter 4

Performance Analysis of the UTRAN for Multimedia Services

4.1 Introduction

The UMTS terrestrial radio access network (UTRAN) handles all radio related functionalities such as access control, mobility management and radio resource control, broadcast and multicast services and other minor functionalities. The UTRAN consists of one or more RNCs, a number of Node-Bs and a number of interfaces such as I_{ub}, I_{ur} and I_{up} [17]. Among the UTRAN interfaces the I_{ub} link plays a vital role in the overall operation of the network which connects RNC to the Node-B regulating end-to-end user and control traffic. To maintain the end-to-end QoS it is very important to appropriately design the link by analysing the performance of an I_{ub} link for multimedia traffic. The I_{ub} link performance is influenced by transport and link layer protocols along with the offered traffic load. The 3GPP standard doesn't specify any specific transport layer requirements which open up the possibility of using any transport protocol as long as these protocols satisfy the QoS requirements of the UTRAN. The selection of a transport layer protocol for the I_{ub} link is crucial for a radio access network design. The I_{ub} link can affect the QoS of users' connections on the air interface. In the release 99 of the 3GPP standard the AAL2/ATM transport technology was chosen to support multimedia traffic on the I_{ub} link due to its ubiquitous nature that offers QoS guarantee to support heterogeneous traffic. Also, the widespread deployment of the ATM technology in the existing telecommunications networks was one of the another main reasons to select the AAL2/ATM based transport protocol [39]. With the advancement of the IP based network architecture the 3GPP release 5 proposed an IP-based transport technology for the next generation system, particularly for the packet based services. It is noticeable that the 3GPP release 5 standard has started focussing more on packet switched domain. In the release 5 three user plane transport protocol solutions are proposed, those include the CIP (Composite IP), LIPE (Lightweight IP Encapsulation) and PPP-MUX (Point-to-Point Protocol Multiplexing) based I_{ub} link design. In addition to above protocols recently the Carrier Ethernet has been proposed as a high speed transport network which has attracted a lot of attention of network researchers for the wide area network deployment because of relatively lower cost, simplicity and flexibilities[40]. This solution gives an impetus to the IP/Ethernet transport technologies in the UMTS network from the convergence benefit point of view. Currently the fixed network architecture migrates towards to a unified IP/Ethernet infrastructure. There is a number of other wireless access networks have emerged which could potentially be connected to a UTRAN to support multimedia traffic along with the UMTS network. Some of the notable standards which can work along with the UMTS networks are the IEEE802.11 wireless local area network (WLAN) and the IEEE802.16 based WiMAX (Worldwide interoperability for Microwave Access) network. The IEEE 802.11 WLAN standard is now widely used in many enterprise networks and also in public places as hotspots to provide Internet access and VoIP (Voice over IP) services. The IEEE 802.16 WiMAX standard has the potential to become an important BWA (broadband wireless access) standard for fixed and mobile services which can operate either cooperatively or competitively with the UMTS network [41]. In future it is very likely the both the WLAN and WiMAX standards will be deployed for public access where these access points or base stations will be connected to an Ethernet based backhaul network or a radio access network. An Ethernet based UTRAN could makes it easy for operators to offer inter-networking functions as well as to integrate UMTS, WLAN and WiMax based networks. A backhaul network which includes a radio access network and a core network, contributes a significant proportion of the construction and operational cost of a deployed wireless or a mobile network. A unified backhaul network could significantly reduce the operating cost of a future heterogeneous wireless networks. This chapter evaluates the performance of the UTRAN for the multimedia traffic. In this chapter the UTRAN performance will be evaluated for a number of different transport protocols: AAL2/ATM, IP and Ethernet.

In this section we review some of the earlier works carried by other researchers to evaluate the performance of the UTRAN from different points of interest. Makke et. al. [9] studied some performance issues related to the AAL2 protocol in the context of the UTRAN design the paper studied the impact of the Timer-CU (Timer combined use, which is used for the AAL2 multiplexing to assure that packed packets don't wait longer than the value of Timer-CU) on the I_{ub} link performance, the scheduling mechanism of the AAL2 multiplexer and the AAL2 switching. The author presented an optimal Timer-CU value for a mono-service VC (Virtual Circuit). Comparison between different scheduling mechanisms including weighted round robin (WRR), earliest deadline first (EDF), and first in first out (FIFO) techniques were studied. A comparison between the AAL2 switching and ATM switching was presented as well. Isnard, et. al. [42] investigated two approaches to handle UMTS traffic classes at AAL2 and ATM layers over the I_{ub} and I_{ur} interfaces in the UTRAN. The first approach was used to improve the AAL2 CPS scheduler by introducing the notion of traffic priority. The second approach was used to multiplex different UMTS traffic class connections onto separate ATM connections with specific QoS requirements. Garcia et. al. [18] studied the relationship between the voice frame loss and the minimum ATM virtual circuit capacity to meet a frame loss ratio requirements (voice and web data traffic) of the UTRAN. Chong et.al. [43] introduced an enhanced dynamic weighted round robin (DWRR) scheduling scheme, which dynamically changes the weight of different traffic classes based on the UTRAN traffic load situation. This scheme increased the overall throughput of the UTRAN by guaranteeing the delay requirements of each service class. The MWIF (Mobile Wireless Internet Forum) [44] has compared the performance of ATM and IP based UTRAN transport networks. The work examined the effect of some important IP related technical issues such as QoS, link efficiency, size of IP packets, routing, and security on the UTRAN. This work describes several transport protocol architecture which includes PPP multiplexed frame transmission using HDLC, ATM/AAL5, and L2TP tunnel protocols, composite IP (CIP) and the lightweight IP encapsulation (LIPE). This work presented analytical and simulation results describing the bandwidth efficiency and delay properties of different proposed transport protocol stacks used in an IP-based RAN. Results showed that most protocols performed better than the reference protocol stack (AAL2/ATM) when used with PPP/HDLC link layer protocol. One exception was that TCRTP (Tunnelled multiplexed compressed RTP) protocol performance degraded due to its tunnelling overhead. The CIP protocol with its low overhead and segmentation capability showed a very good performance. Jung et. al. evaluated the performance of ATM and IP based UTRAN transmission schemes with different types of offered traffic in terms of link efficiency and average link transmission delay [45]. The result showed the AAL2/ATM transmission scheme performs better than IP based schemes at a low load state whereas IP based schemes offered better performance at a high load state.

Review of above related works showed that UTRAN performance was evaluated for a number of transport protocols which include ATM and IP based transport protocol options. This chapter extends the above studies by analysing the performance of the UTRAN using the IP/Ethernet based transport layer and comparing these results with the ATM and IP based UTRAN solutions. This chapter first introduces multimedia traffic models used to analyse the performance of the I_{ub} link. A priority mechanism is also used to improve the

UTRAN performance by offering priority to the retransmitted traffic for different radio channel conditions.

Rest of the chapter is organised as below. Section 4.2 briefly describes the UTRAN I_{ub} link protocol structure. Section 4.3 overviews different transport layer protocols proposed for the UTRAN. This section also introduces different frame mapping and multiplexing structures for the I_{ub} link. Section 4.4 describes traffic models and frame transmission strategies in the UTRAN. Section 4.5 analyse different simulation results and compare performance of the UTRAN for various transport protocols. A short summary is presented in section 4.7.

4.2 The I_{ub} link Protocol Structure

As described in chapter 2, the RNC and the Node-B are connected via the I_{ub} link. The RNC control radio resource management, mobility and user data transport between the core network and UEs. The Node-B is mainly responsible for the air interface layer 1 (physical layer) processing and some basic radio resource management issues such as modulation, the inner loop power control, etc. In 3GPP release 5, the Node-B functionality is enhanced to provide fast scheduling and fast physical layer hybrid automatic repeat request (H-ARQ), which improve the throughput and delay to support High Speed Downlink Packet Access (HSDPA) connections. As Figure 4.1 shows that data packets from upper layer are segmented or concatenated in the RLC layer and then mapped or multiplexed into transport blocks (TBs) at the MAC layer. Transport Blocks are encapsulated into the Frame Protocol (FP) frame. The FP frame is transported to the Node-B via the underneath transport network layer (TNL). The standard transport technologies such as ATM, IP and Ethernet etc, are deployed at the TNL level. The FP layer and the TNL provide the connection between the MAC layer and the physical layer (except for the HSDPA standard). For the HSDPA standard, the FP layer and the TNL provide the a connection between MAC-d and MAC-hs layer (as shown as Figure 3.1)

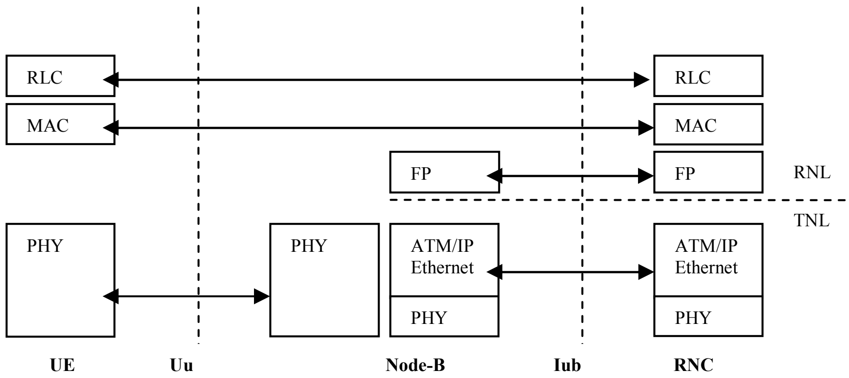


Figure 4.1 User plane protocol stack for the I_{ub} interface

The FP defines the structure of frames and basic in-band control procedures for every type of transport channel. The DCH FP carries transport block set (TBs) and outer loop power control information across the I_{ub} interface between the RNC and the Node-B. It also supports the transport channel and node synchronization mechanism [46]. There are two types of DCH FP frames: data frame and control frame. The payload of control frame contains commands, and measurement reports related to transport bearer and the radio interface physical channel but not directly related to specific radio interface user data. The payload of the data frame contains radio interface user data, quality information for transport blocks and for the radio interface physical channel during the transmission time interval (for uplink only), and an optional field. The user data frame is used to transparently transport the transport blocks between the RNC and the Node-B. Coordinated DCH channel with the same transmission time is allowed to multiplex into same transport bearer. The TBs of all the coordinated DCHs for one transmission time interval are included in one frame. The structure of downlink data frame is shown in Figure 4.2 [46].

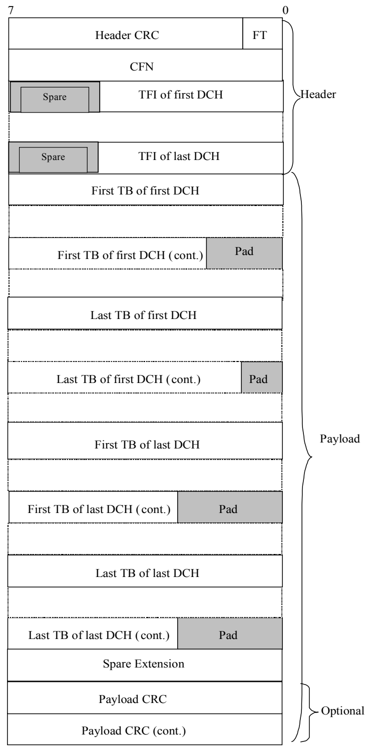


Figure 4.2 Downlink FP data frame structure

The number of Transport Format Indicator (TFI) fields is the same as the number multiplexed DCH channels in the same transport bearer. The correspondent TFI defines the size and the number of TBs for each DCH channel. Bit padding is used when the TB length does not consist of integer number of bytes. The header CRC calculation covers all bits in the header, starting from bit 0 in the first byte (FT field) up to the end of the header. The payload CRC calculation includes all bits in the data frame payload, starting from bit 7 in the first byte up to bit 0 in the byte located before the payload CRC field. The FT field indicate whether the frame is a control or a data frame. .

The HS-DSCH Data Frame carries HS-DSCH channel data over the I_{ub} link. The Figure 4.3 shows the position of the MAC-d PDUs which are mapped onto the HS-DSCH data frame. The DRT (Delay Reference Time) is used for dynamic delay measurements, which is an important parameter for the I_{ub} link traffic congestion control [35].

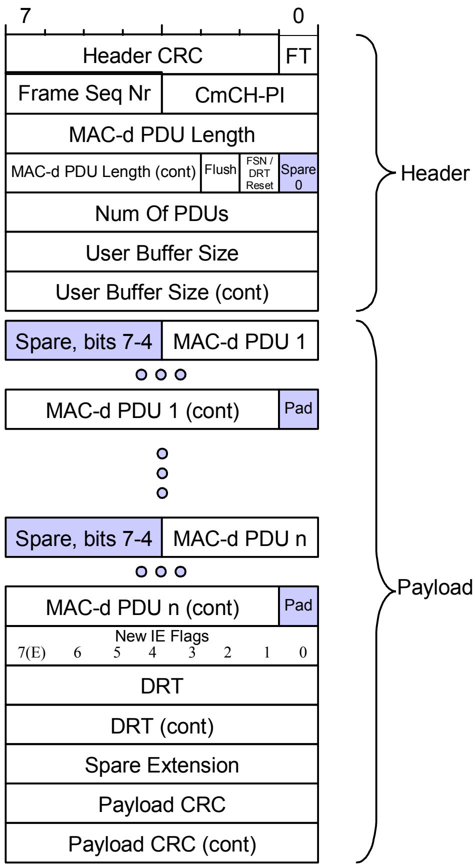


Figure 4.3 HS-DSCH FP data frame structure

A flow control algorithm is used to control traffic flow between MAC-hs and MAC-d layers. At the beginning of a call, the RNC send a request to the Node-B for the HS-DSCH capacity allocation. The Node-B allocate capacity to the RNC, and indicates the maximum MAC-d PDU length, allowed number of PDUs, HS-DSCH interval and repetition rate, etc, to control user data flow according to the capacity allocated by the Node-B. The HS-DSCH data is mapped onto a HS-DSCH Data Frame and transmitted immediately from the RNC to the Node-B. Currently the HS-DSCH channel is used to transport data traffic only. When the HS-DSCH is operating at the theoretical peak rate, the highest bursty traffic flowing is brought over the I_{ub} link.

4.3 Transport layer protocol

As description in previous sections, the transport network protocols used in the UTRAN have not been designed specifically for UMTS networks. Instead, the 3GPP have selected transport protocols from existing transport protocol suites. The selection of transport network protocols for UMTS purposes has focused on the ability to multiplex traffic from different end-users and to be more precise from different UMTS bearers taking into account their QoS characteristics [11]. The convergence of fix and mobile networks, and similar services expectation for both type of networks are the main factors for selecting the ATM, IP and Ethernet protocols for the UMTS network. We consider Plesiochronous digital Hierarchy (PDH)/ Synchronous Digital Hierarchy (SDH) transport technology as layer one transport protocol for TNL, due to their dominance in current public networks. ATM, IP and Ethernet transport frames can be transmitted over the well-established PDH/SDH infrastructure [47]. The followings describe these technologies to be used to transport FP frames over the I_{ub} interface.

4.3.1. AAL2/ATM Protocol

The AAL2 is one of the adaptation layers used in the ATM protocols structure. AAL layers are used to support various traffic services in an ATM network where the adaptation layer adapts high layer protocol parameters appropriately to match with lower layer parameters. There are four types of AALs used in the ATM standard which are known as AAL1, AAL2, AAL3/4 and AAL5. The AAL1 supports synchronous mode, connection-oriented connection and constant bit rate. The AAL2 offers synchronous mode, connection-oriented connections with variable data rate. The AAL3/4 offers connectionless and asynchronous connection also with a variable data rate. The AAL5 offers asynchronous mode, connection-oriented connection with a variable data rate. The AAL2 and AAL5 are the most appropriate alternatives in the UMTS transport network. The AAL2 layer is suitable for the CS user plane connections, and the AAL5 is suitable for control plane and the PS user plane data exchange. Within the UTRAN the AAL5 layer is deployed to carry all control data on I_u, I_{ur} and I_{ub} interfaces, as well as the PS domain user data on the I_u interface. In contrast, the AAL2 layer is used to carry user data on all interfaces, except the PS domain user data at the I_u interface. [11]

The AAL2 layer offers bandwidth-efficient transmission of low-rate, short, and variable length packets in delay sensitive applications. More than one AAL2 user information stream can be transported on a single ATM connection. To support specific AAL2 user services, or groups of services, different SSCS (Service Specific Convergence Sub-layer) protocols are defined by ITU [48]. The SSCS may also be null when it merely providing for the mapping of the equivalent AAL primitives to the AAL2 CPS primitives and vice versa. With Segmentation and Reassembly Service Specific Convergence sub-layer (SEG-SSCS) applied for a SSCS, the high layer data unit exceeding 45 bytes (or 64 bytes) is segmented into maximum length of 45 bytes (or 64 bytes) packet and 3 bytes header are added to form a CPS-Packet. The CPS-Packet header includes Channel Identifier (CID 8 bits), Length Indicator (LI 6 bits), User-to-User Indication (UUI 5 bits) and Header Error Control (HEC 5 bits). Reassembly service is provided on the reverse link. The AAL2 makes use of the service provided by the underlying ATM layer. Multiple AAL2 connections may be associated with a single ATM layer connection, allowing multiplexing at the AAL2 layer level. Multiplexing in the AAL2 layer occurs in the Common Part Sub-layer (CPS). Timer-CU is used for the AAL2 multiplexing to assure that packed packets don't wait longer than the value of Timer-CU before being scheduled for transmission. The CPS-Packets are mapped onto the CPS-PDU payload. The CPS-PDU consists of one octet start field (STF) and a 47-octet payload. The 48-octet CPS-PDU becomes the ATM-SDU [49]. The DCH or HS_DSCH FP data frames are segmented, and formed into AAL2 CPS-Packets. CPS-Packets are mapped into a CPS-PDU, and then encapsulated into ATM cells as shown in Figure 4.4.

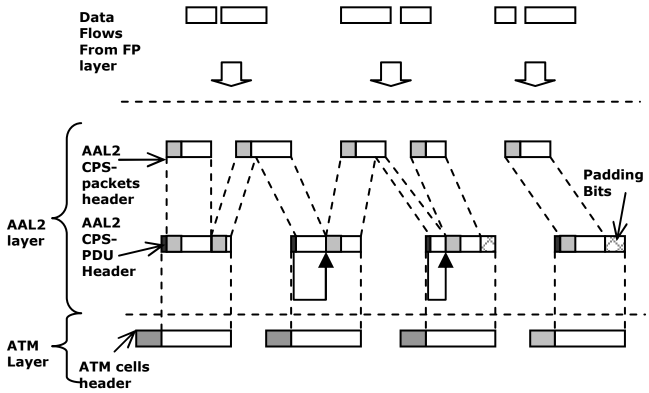


Figure 4.4 AAL2 multiplexing structure

4.3.2. IP Protocol

In order to create a basis for full-scale utilisation of IP internetworking technology as the common transport protocol for the future evolution of UMTS networks, it became necessary to study the IP option for both signalling and user data transport in all UTRAN interfaces. The dominance of the IP protocol in data networks and the efforts in incorporating the IP QoS provisioning mechanisms pave the way for the IP to be deployed in the UMTS as a transport solution. In the 3GPP release 5 three user plane solutions are proposed, those includes the CIP (Composite IP), LIPE (Lightweight IP Encapsulation) and PPP-MUX (Point-to-Point Protocol Multiplexing) solutions. A CIP based solution is described in this section. To support efficient usage of the I_{ub} link, the CIP (Composite IP) option can be used. The CIP allows multiplexing of variable CIP packets to be transmitted in one variable size CIP container. As Figure 4.5 showing, a large FP PDU is segmented in order to avoid IP fragmentation and to keep the transmission delay low. A CIP container can pack multiple CIP packets to improve the transmission efficiency. A CIP container is encapsulated into an UDP/IP packet. The UDP/IP header could be compressed to improve the bandwidth efficiency. A UDP/IP packet is combined with the HDLC/PPP header and tail bits and then transmitted over a PDH/SDH network [10].

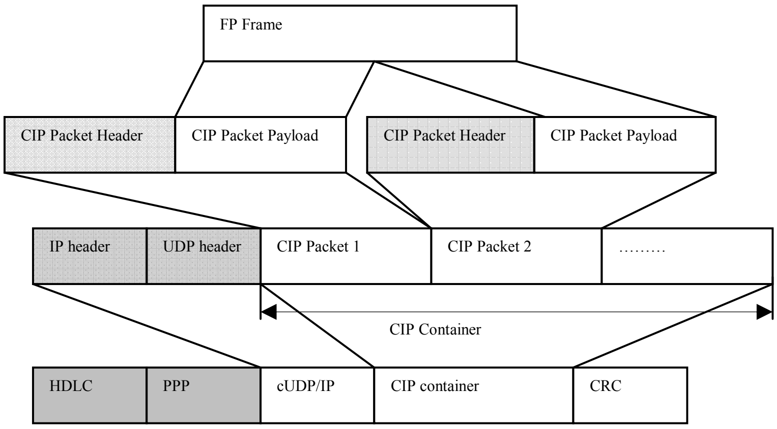


Figure 4.5 CIP/UDP/IP/PPP/HDLC multiplexing structure

4.3.3. IP/ Ethernet Protocols

Carrier Ethernet services could be delivered over a native Ethernet based metropolitan and access networks, and can also be supported by other transport technologies and protocols such as SDH, DWDM, MPLS, etc. The Carrier Ethernet possesses five attributes which are standardized service, scalability, service management, reliability and QoS. These attributes distinguish the Carrier Ethernet from the LAN based Ethernet because the LAN based Ethernet is a best effort networking technology. The Ethernet Forum (MEF) defines the User network Interface (UNI) Type 1 for the initial phase of the deployment of the Carrier Ethernet. The UNI Type 1 focuses on the Ethernet deployment using existing customer equipments which requires no changes and use the existing IEEE Ethernet interface [50]. With its simplicity and rigours specifications, and the dominance in data networks and efforts in incorporating OAM Scaling, QoS provisioning mechanisms have pave the way for Carrier Ethernet to be deployed in UMTS as an alternative transport solution. There are a number of frame encapsulation technologies available which includes the high-Level Data link Control (HDLC), Link Access Procedure for SDH (LAPS/x.86), and Generic Framing Procedure (GFP). The GFP technique provides a consistent and predictable payload throughput because the flag substitution is not performed in the data stream. We choose the GFP as the preferred encapsulation method. The preamble and SFD (start of frame delimiter) field of the Ethernet frame are not encapsulated into the GFP frame. The CIP (Composite IP) option is used to support efficient usage of the I_{ub} link. Figure 4.6 shows the CIP/IP/Ethernet/GFP frame multiplexing structure [51]. GFP PDUs can be carried over the PDH/SDH networks.

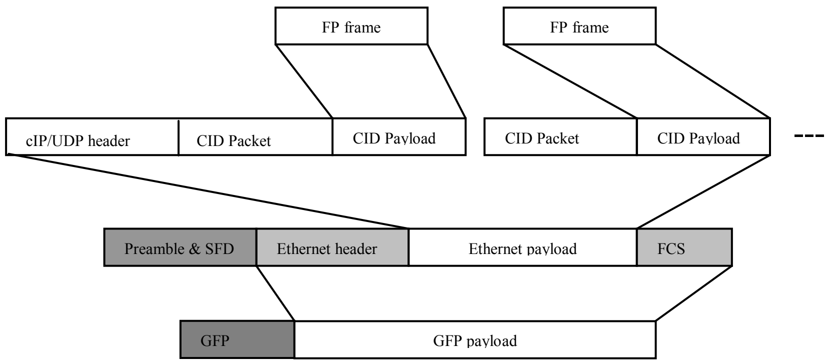


Figure 4.6. CIP/IP/Ethernet/GFP multiplexing structure

4.4 Traffic Model

This chapter analyse the performance of the UTRAN for multimedia traffic using voice, data and video traffic models. This section describes the traffic models used in this simulation work. The 3GPP standard defines voice and web data traffic models at the source level. Voice and web browsing data source models have the distinctive “ON/OFF” characteristic. In [52], the traffic model is structured in three levels: session level, burst level and packet level (shown as Figure 4.7). The session level traffic model generates user data activities. Session arrival pattern and session duration are modelled by statistical distribution functions using relevant parameters at this level. An active session generates two states (ON or OFF) in the burst level. Each state generates its own packet using statistical distributions.

After the traffic data is passed though several protocol layers, user traffic will be reshaped because of the frame handling procedures. In the different layers of the stack corresponding protocol headers are added before the transmission. These issues are incorporated in the simulation model. In the following sections we describe various traffic generators used in the investigation of transport layer performance.

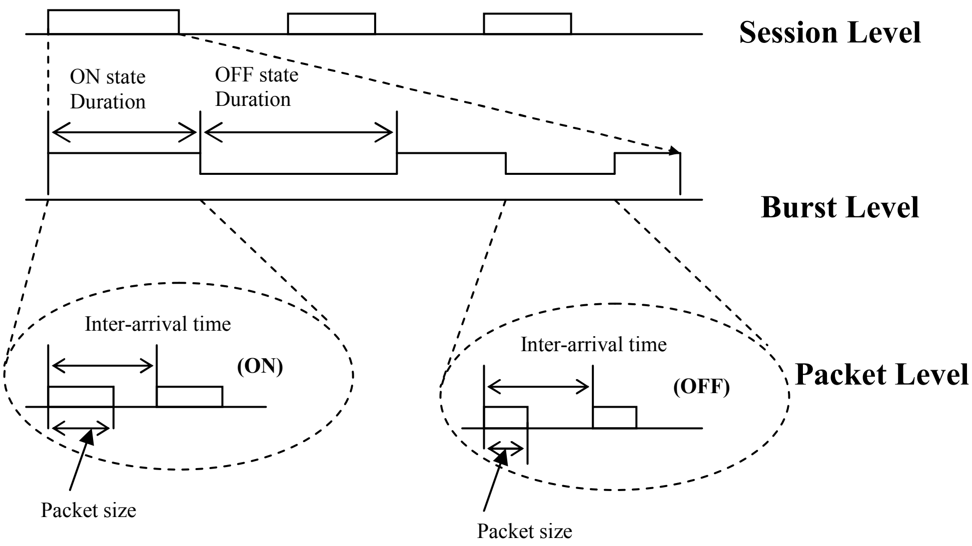


Figure 4.7 Traffic source model

4.4.1. Voice Traffic

The AMR (Adaptive Multi-rate) speech coder is employed in the UMTS network to offer voice services. The AMR is a multi-rate speech coder, which is consists of a source controlled rate scheme including a voice activity detector and a comfort noise generation system, and an error concealment mechanism to combat the effects of transmission errors and lost packets [53]. The multi-rate speech coder is a single integrated speech codec with eight source rates ranging from 4.75 kbps to 12.2 kbps, and a low rate background noise encoding mode. The speech coder is capable of varying its bit-rate every 20 ms speech frame. Depending on the air interface loading and the quality of the speech connection, the RAN could control the bit rate of an AMR speech connection.

The recorder bits are further divided into three indicative classes according to their subjective importance [54]. This class division is only informative and provides supporting information for mapping this generic format into specific formats. These three classes of speech bits can then be subject to different error protection in the network. Class A contains the bits most sensitive to errors and any error in these bits typically result in a corrupted speech frame which should not be decoded without applying appropriate error concealment technique. The CRC bits in the AMR Auxiliary Information protect this class. The Class B and Class C fields contain bits where increasing error rates gradually reduce the speech quality, but decoding of an erroneous speech frame is usually possible without annoying artefacts. Class B bits are more sensitive to errors than Class C bits. The importance ordering applies also within the three different classes and there are no significant step-wise changes in subjective importance between neighbouring bits at the class border.

The Table 4.1 shows the number of speech bits in Class A, Class B and C fields for each AMR codec mode.

Table 4.1 Number of bits in Classes A, B, and C for each AMR codec mode

Frame Type	AMR codec mode	Total number of bits	Class A (bits)	Class B (bits)	Class C (bits)
0	4.75	95	42	53	0
1	5.15	103	49	54	0
2	5.90	118	55	63	0
3	6.70	134	58	76	0
4	7.40	148	61	87	0
5	7.95	159	75	84	0
6	10.2	204	65	99	40
7	12.2	244	81	103	60

The comfort noise information is transmitted to the receiver side by sending the silence descriptor (SID) frame on a regular basis. The comfort noise bits are all mapped to the Class A field of the AMR Core Frame.

The AMR (Adaptive Multi-rate) codec model is used to simulate voice connections at the source level Speech frame or silence descriptor (SID) frame is generated after every 20 ms. Speech frames are created in ON periods, the SID frame are created in OFF periods. The distribution of ON and OFF periods are exponentially distributed. Mean value of both ON and OFF state duration is 3 second. The voice packet (speech frame, or SID frame) is split into several flows, and transmitted on the radio channel after every transmission time interval (TTI) of 20 ms. The RLC protocol is used to offer transparent services for voice traffic. Flows are packed into a data frame at the FP (Frame Protocol) layer. The FP Data Frame header, tail and flows header are added with the payload. Parameters for voice traffic (12.2 kbps AMR) entering TNL are shown as Table 4.2 [10]. Four flows (DCH0-3) are used to support Class A, Class B, Class C and signalling bits.

Table 4.2 Parameters for voice traffic for the 12.2 kbps AMR in ON state

Parameter	ON state	OFF state
Header CRC, CFN	2 bytes	2 bytes
TFI	4 X 1 byte	4 X 1 byte
Class A + Padding	11 bytes	5 bytes
Class B + Padding	13 bytes	0 bytes
Class C + Padding	8 bytes	0 bytes
Signalling (every 300 ms)	0 or 10 bytes	0 bytes
Payload CRC	2 bytes	2 bytes
Total size of FP frame	40 or 50 bytes	13 bytes
Mean value state duration (exponential)	3 second	3 second

4.4.2. Web Data Traffic

The UMTS traffic data model is used to develop the web browsing application [52]. The web browser generator generates a number of packets during the ON period which corresponds to the download of web pages. ON periods are geometrically distributed. The burst size distribution is controlled by the Pareto distribution. The reading time (OFF period) between two consecutive bursts is also geometrically distributed with a mean value of 12 second. The generated web page file size is Pareto distributed with a parameter α= 1.1, mean file size of 12,000 bytes, minimal file size of k=1858 bytes and the maximum file size m= 5000000 bytes. The probability density function of the Pareto distribution is shown in (4.1). The reading time (the OFF time between consecutive bursts) is exponentially distributed with a mean value of 12 seconds.

$$f(x) = \begin{cases} \dfrac{\alpha \cdot k^{\alpha}}{x^{\alpha+1}}, & k \le x \le m \\ \dfrac{k^{\alpha}}{m^{\alpha}}, & x > m \end{cases} \qquad (4.1)$$

This source model can't be directly used at the I_{ub} interface because traffic is shaped by radio protocols. The upper layer packets are split by the RLC layer into predefined Transport Block (TB) sizes. Two bytes header is added with a transport block. When payload size is smaller, padding bits are used to generate minimum size TBs. The MAC layer sent a number of TBs for a connection on the radio channel without exceeding the peak data rate in a TTI (40ms). All TBs in a TTI are put into one data frame at the FP layer. To simplify the simulation, we assume the web data source corresponds to the peak rate, and traffic packets are transmitted at the peak rate over the radio channel. The reference [10] defines the FP data frame size for different peak data rates. Each data frame includes a three byte header. The mean of ON period is calculated from using file sizes for various peak transmission data rate (file size/ peak data rate). Parameters for web data traffic entering the TNL are shown in Table 4.3.

Table 4.3 Parameters for web data traffic

Data rate	State	Packet size (constant)	Mean value of state duration
64 kbps	ON	339 bytes	1.5 second (Pareto)
	OFF	0 bytes	12 second (Exponential)
144 kbps	ON	759 bytes	0.667 second (Pareto)
	OFF	0 bytes	12 second (Exponential)
384 kbps	ON	2019 bytes	0.25 second (Pareto)
	OFF	0 bytes	12 seconds (Exponential)

4.4.3. Streaming Video Traffic

Streaming refers to the ability of an application to play synchronised media streams like audio and video streams in a continuous manner. These streams are being transmitted to the client over a data network. The 3GPP Packet Switched Streaming Session (PSS) provides a framework for Internet Protocol (IP) based streaming applications in 3G networks [55]. The PSS enable mobile streaming applications, which include streaming of news using still images and speech at very low bit rates, music listing at various bit rates and qualities, video clips and watching live sports events [56].

The 3GPP specified a PSS client supporting video application using the ITU-T H.263 decoder and the MPEG-4 Visual Simple Profile Level 3 decoder. The H.263 coder supports video compression for video-conferencing and video-

telephony applications. The H.263 coder was developed to stream video using a bandwidth ranging from 20Kbs to 24Kbps. Comparing to the H.261coder the H.263 coder uses half the bandwidth to achieve the same video quality. As a result, the H.261 coder has been largely replaced by the H.263 coder. The H.263 coder deploys the RTP protocol to transport video streams. [57]
Due to the use of a highly efficient codec algorithm the MPEG-4 and the H.263 encoded video are expected to account for large portions of transmitted traffic in future wireline and wireless networks [58]. Both encoding standards employ the discrete cosine transform (DCT) to reduce the spatial redundancy in the individual video frames, and a predictive encoding technique is used to reduce the temporal redundancy that is the temporal correlation between successive video frames. Following section briefly introduce the MPEG-4 and H.263 video compression technique. These technique are further elaborated in the reference [58].

a) MPEG-4 Video Compression

The MPEG-4 is a very efficient video coding technique, which provides different bit rates using variable bit rate output for a range of application environments. The MPEG-4 is an object based video coding. Each scene is composed of individually code video objects (VOs). Each VO may have several video objects layers (VOLs). Each VOL consists of an ordered sequence of video object planes (VOPs). The encoder processes the shape, motion, and texture characteristic for each VOP. Each video frame is divided into macro blocks (MBs). VOPs are arranged in a periodic pattern referred to as a group of pictures (GoP). The MPEG-4 coding technique provides a number of error resilience and error concealment features to combat the frequent transmission errors typical in the wireless communication.

b) H.263 Video Compression

For H.236 video coder the bit rate of the compressed video stream is controlled by adjusting several encoder parameters (e.g., quantizer scales and frame rate). The H.263 offers four advance coding options. Unrestricted motion vectors, advance predication and the PB frame are options that improve the inter-picture prediction. The fourth option uses more efficient arithmetic coding technique instead of using the variable length coding. A PB frame consists of a P frame which predicts from the preceding P frame, and a B frame which bi-directionally predicts from the preceding P frame and the P frame that is part of the PB entity. These four options improve the video quality at the expense of increased video codec complexity.

Unlike bit streams which give the actual bits carrying the video information, the video traces only give the number of bits used for the encoding of the individual video. A video trace gives defined quantities typically in an ASCII file with one line per frame [59].

Output rate control strategies for video coding are classified into two classes; constant bit rate (CBR) and variable bit rate (VBR). The main application of the CBR rate control is encoding for transmission bit stream over a constant rate link under strict end-to-end delay constraint. VBR video rate control strategies are used when the constant transmission rate constraint of an application or low-delay constraint is relaxed. A UMTS WCDMA bearer with strict QoS guarantee requires rate-smoothing techniques which allow streaming of variable rate encoded content at a constant transmission rate. When transmitting at a variable rate an encoded video content over the UMTS network, the transmission rate of a video stream is adjusted according to the available bandwidth of the UMTS streaming bearer, which corresponds to the negotiated guaranteed bit rate.

The MPEG-4 video traces with a constant bit rate were used to investigate the performance of the transport network layer. Video traffic generation process is shown in Figure 4.8. Several MPEG-4 single layer video trace files with rate control are used as video traffic model input. Via parsing trace file, traffic source model generates video frames at the rate of 30 frames per second. Similar to the web data packets, the RLC splits the video frame into several TB size segments (excluding header bytes). The RLC layer operates in the transparent mode (TM). The MAC layer sent TBs on the radio channel at the peak data rate in every TTI (40ms). Data Frames are created after every TTI (40ms) to form video streaming traffic on the FP layer, and then enters the AAL2 layer for generating ATM cells or relevant IP, Ethernet layer.

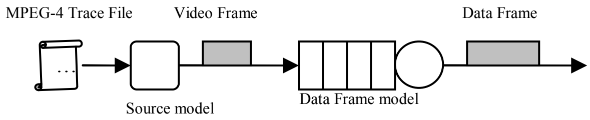


Figure 4.8 Video FP data frame generation

4.5 Simulator Architecture

Each user session (voice or data or video) is assumed to make use of one DCH channel at the FP layer. The resulting DCH FP data frames form traffic streams, which are carried across the I_{ub} link. The voice traffic is model as a symmetric data stream for each user session. The web data and streaming video traffic are asymmetric in nature. The downstream traffic (RNC to Node-B) volume is heavier than the upstream traffic volume. The downstream traffic is simulated for an asymmetric traffic flow.

On the transmitter side (at the RNC side) (illustrated in Figure 4.9), the traffic source generates the DCH FP traffic (voice, data or video). FP frames are segmented and encapsulated into AAL2 PDUs or into CIP PDUs. The AAL2 multiplexer aggregates all same type traffic streams into one ATM VCC. The CIP multiplexer aggregates traffic streams into UDP/IP PDUs, and then these PDUs are mapped onto HDLC or Ethernet packets. In the simulation E1 physical transmission links (2048bit/s) are used [60]. The physical line server's service rate is set to 1920 bit/sec.

On the receiver side (Node-B) (shown as Figure 4.10), the ATM, IP or Ethernet de-multiplexer branch the ATM cells, IP or Ethernet packets to different traffic groups. CPS-PDUs or CIP-PDUs are demultiplexed by the AAL2 or CIP de-multiplexer into individual DCH FP data frames. The extra delay will be introduced by the reassembly process. The transmission periodicity and phase of voice, data and video streams could be different. The phase of traffic sources are assumed to be uniformly distributed. See Figure 4.11.

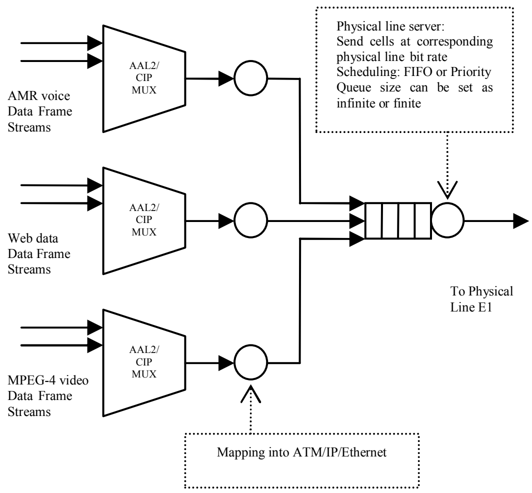


Figure 4.9 Transmitter side of Simulator (RNC side)

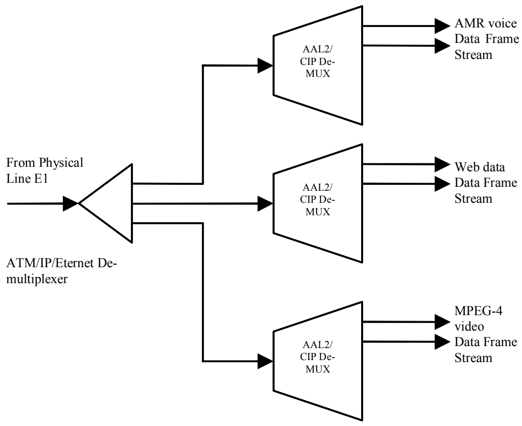


Figure 4.10 Receiver side of Simulator (Node-B side)

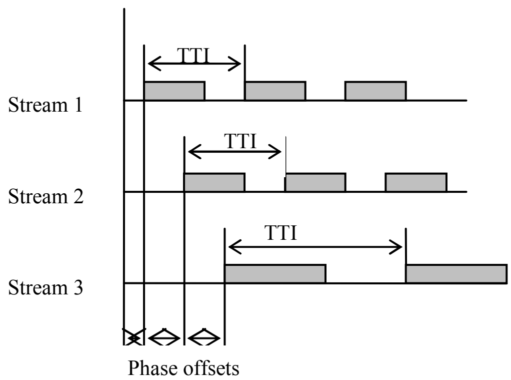


Figure 4.11 an example of packet over DCH

The parameters for the simulator are shown as Table 4.4:

Table 4.4 the simulator parameters

Traffic Type	AMR 12.2 kbps voice; 64kbps web data; 64kbps and 128 kbps video stream
Transport Protocols	AAL2/ATM, IP and Ethernet
Physical Link Bandwidth	E1 (2.048 Mbps)
Scheduling Algorithm	FIFO, Priority
User Numbers	Variable
Queue Size	Unlimited or Limited

4.6 Performance Analysis and Comparison

Transport network capacity should guarantee the end users' QoS to meet the service level agreement once connections are admitted and established. In this section, we used the OPNET based simulator and traffic models described in previous sections to evaluate the performance of the AAL2/ATM layers when used in the I_{ub} link. The following sections investigate the performance of ATM, IP and Ethernet based I_{ub} links to support DCH and HSDPA traffic. The results could be used for the transport network capacity planning and optimisation.

4.6.1. AAL2/ATM performance analysis

In this section, the study focuses on the AAL2/ATM protocols to support multimedia traffic over the I_{ub} link. The impact of the Timer-CU on QoS and bandwidth utilization are also investigated. The relationship between ATM cells and AAL2 packets loss ratio for different traffic load is also examined. Later an algorithm to improve web data service quality is also studied in this section.

a) Performance Impact of Timer-CU

The Timer-CU (Timer combined use) is used for the AAL2 multiplexing to assure that packed packets don't wait longer than the value of Timer-CU before being scheduled for transmission. When the timer expires, the padding octets are mapped onto unused payload of the CPS-PDU. The filling ratio of the CPS-PDU is defined as equation (4.2). Higher filling ratio means higher bandwidth utilization since less padding bits are added.

$$filling\ ratio\ =\ \frac{payload - padding}{payload} \qquad (4.2)$$

Firstly, we investigate the performance of a packet lossless system by setting the physical link server queue size to the infinite value. The Timer-Cu value is set to 0.5ms, 1ms, 2ms and 3ms separately. For each Timer-CU value, the number of user data streams is increased from 10 to 130. Figure 4.12 shows that the Timer-CU significantly affects the packet filling ratio, especially in a low traffic load. Lower Timer-CU led to lower filling ratio due to short waiting time. Figure 4.13 and Figure 4.14 shows that at low traffic load the significant proportion of the delay is caused by the AAL2 multiplexing operation. A larger Timer-CU value leads to a long multiplexing and de-multiplexing delay, and consequently to the connection QoS could degrade. A lower Timer-CU value leads to use of more padding bits during the multiplexing operation resulting poor bandwidth utilization. With increasing traffic load, the AAL2 multiplexing delay could decrease. The transmission delay could decrease for large Timer-CU (2, 4 ms) values until a high traffic load lead to a high server queue delay becomes the main component of the total delay. The Timer-CU value is not an important parameter to impact the performance of the I_{ub} link for a high traffic load. But for a very low Timer-CU (0.5 ms) value leads to a padding operation during the multiplexing in a high traffic load. This will directly lead to a peak cell rate (PCR) increase, which causes high sever queue delay.

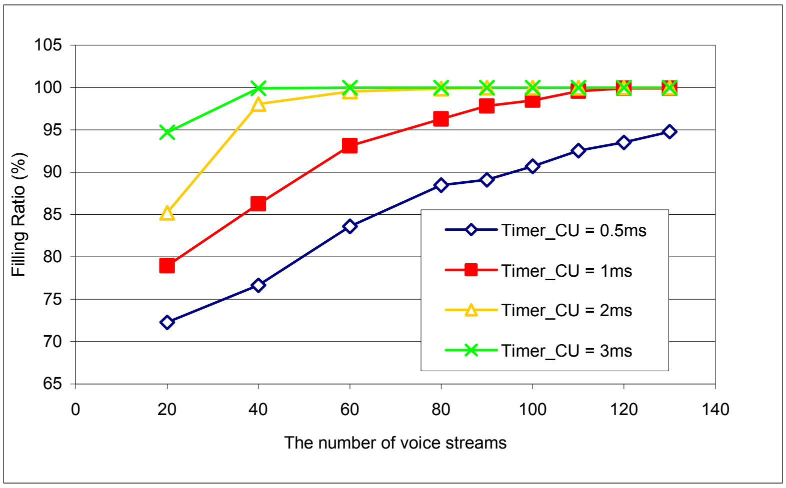


Figure 4.12 Filling ratio of CPS-PDU vs. no of voice stream

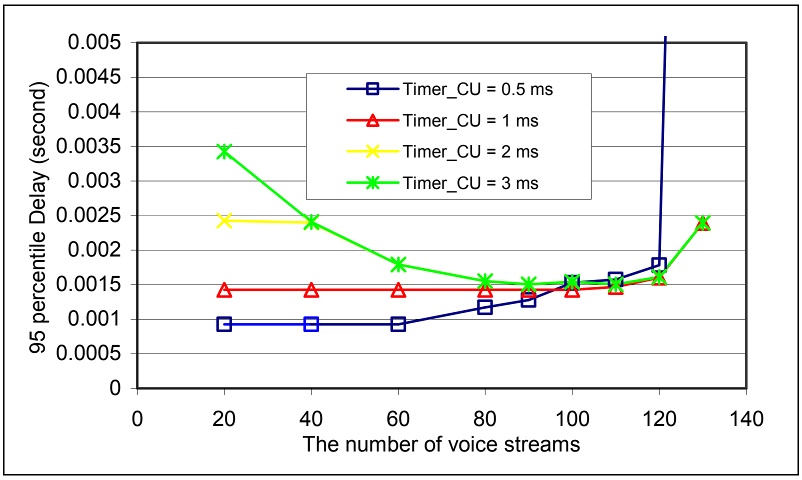


Figure 4.13 95-percentile delay of voice Data Frame

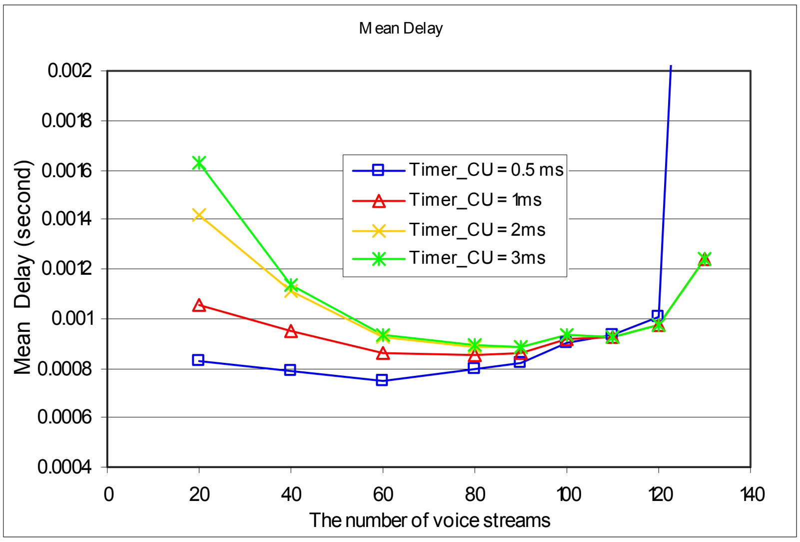


Figure 4.14 Mean voice Data Frame transmission delay

b) Loss Ratio of ATM Cell and AAL2 Packet

In a loss scenario, the physical link server queue size of each link is allocated with a 10 cell buffer size and with a Timer-CU value of 2 ms. When the queue is full any new incoming ATM cell would be discarded, which could lead to the loss of data frame encapsulated in an ATM cell. Packet loss results are shown in Figure 4.15. For voice traffic, data frame loss ratio is equal to AAL2 Packets loss ratio. Since one ATM cell may contain more than one AAL2 packet, one ATM cell loss could lead to more than one AAL2 packet loss. Due to this reason the AAL2 packets loss ratio increases faster than ATM cell losses with the increasing traffic intensity. AAL2 connections QoS may seriously degrade if the ATM cell loss rate is very high. Also, one should keep in mind that any packet loss on the I_{ub} link could increase the total end-to-end delay thus reducing the voice connection quality.

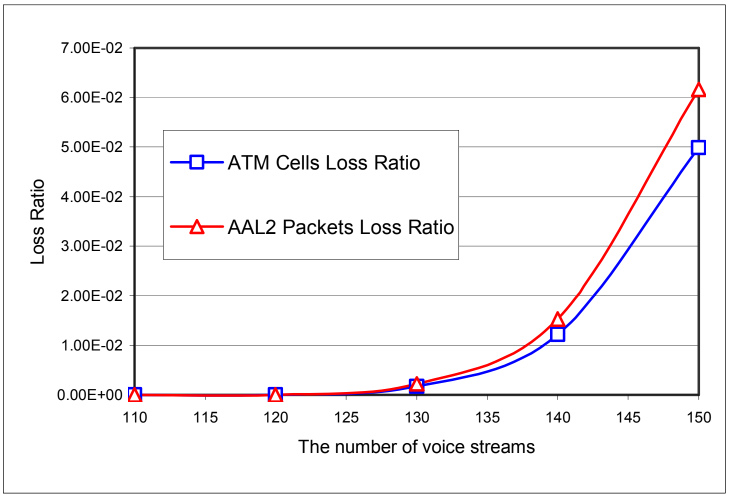


Figure 4.15 AAL2 packet and ATM loss ratio

c) A priority mechanism to the retransmitted traffic

As described in Chapter 2, the RLC Acknowledge Mode (AM) is the normal mode for packet-services such as internet browsing and email transfer. Error correction and in-sequence delivery of higher layer PDUs function is used in the AM mode to keep integrity and the error free condition. An Automatic Retransmission Request (ARQ) mechanism is used as one of the error correction measures. The acknowledged mode data (AMD) PDUs are placed in the retransmission buffer. AMD PDUs buffered are either deleted or retransmitted based on the status report from the receiver. The transmitter deletes all positively acknowledged AMD PDUs and retransmits the negatively acknowledged AMD PDUs only. Decreasing the retransmission delay on the I_{ub} link will improve the QoS of data connection and reduce the buffer size of the RNC and the UE.

We assume that packet received in error due to radio channel conditions are independent identically distributed (i.i.d). The retransmitted web data PDUs also follow the same pattern. We use the simulation model with voice, data and video traffic sources implementing both priority and non-priority techniques. For different traffic categories, different VCCs are used to transmit traffic. Hence, each class of VCC can be treated separately to address the QoS needs of different classes of traffic. Table 4.5 shows the traffic load distribution and priorities attached to different classes. As shown in the table that a value of 3 indicates the highest priority and value of 1 indicates the lowest priority. The retransmitted data traffic is configured to two different multiplexing options. Using the first multiplexing option the retransmitted data is packed with newly generated data into the same FP frame, and transmitted using the same VCC. In the second option, retransmitted data is packed in a separate FP frame, and transmitted using a different VCC and the priority order. Using the FIFO multiplexing scheme, the first option is deployed. In the priority multiplexing scheme, two options are used.

Table 4.5 Combined traffic parameters

Traffic Type	AMR 12.2 kbps Voice	64 kbps Data	64 kbps Video	128 kbps Video
No of sources	60	25	3	2
Priority	3	1	2	2

Figure 4.16 shows the I_{ub} link delay for different retransmission rates. Retransmission rate indicates the proportion of packets that are retransmitted on the air interface. Retransmission on the air interface is mainly controlled by the radio channel conditions. Higher retransmission rate means higher traffic load at the RNC because the RLC layer is operating between the RNC and the UE encompassing U_u and I_{ub} interfaces. Figure 4.16 shows that by offering different

service priorities as indicated in Table 4.5, the overall delay of voice and video traffic could be reduced at the expense of higher data packet delays. Figure 4.17 shows the mean delay for all services including retransmitted data traffic. The graph shows that for retransmitted PDU if additional priority is offered then retransmitted PDU delay decreases significantly (92% reduction at 10% retransmission ratio), which can significantly improve the end to end QoS of a data connection. In this simulation, only the data traffic is needed to be retransmitted because of RLC working in the AM mode for data traffic. The graph also shows the additional priority for retransmitted traffic has less impact on the delay of voice (0.6% increase) and video traffic (3%-10% increase). The additional priority offered to retransmitted traffic reduces the buffer requirements of a RNC and UEs because of the short waiting time.

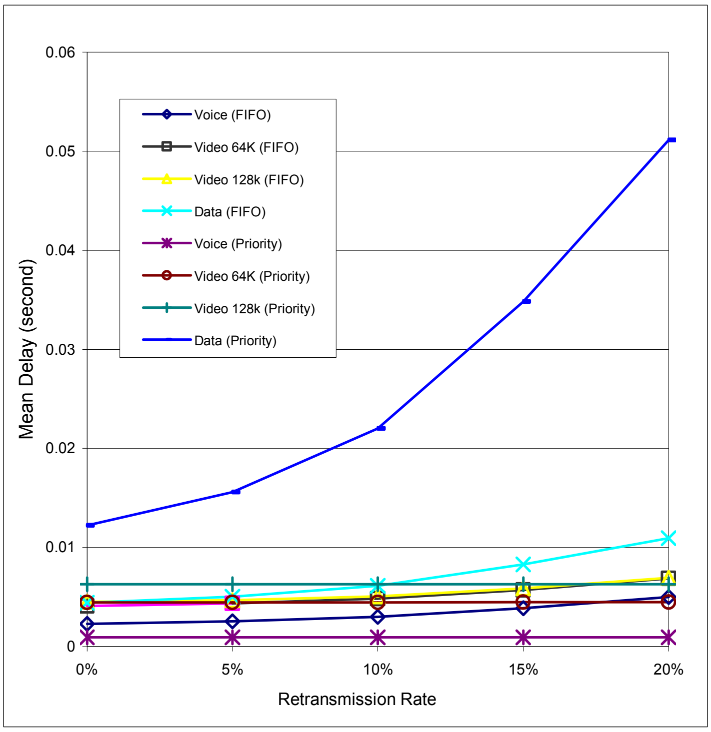
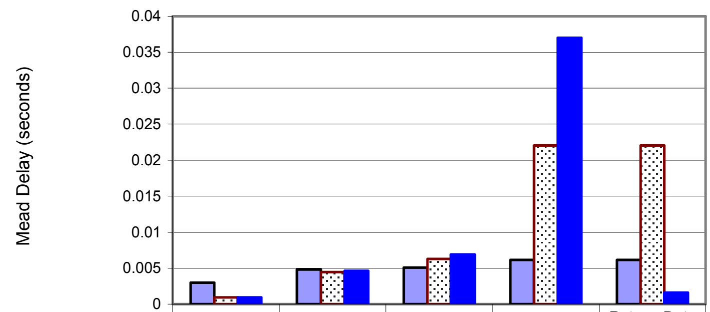


Figure 4.16 Performance comparison of FIFO vs. Priority transmission techniques

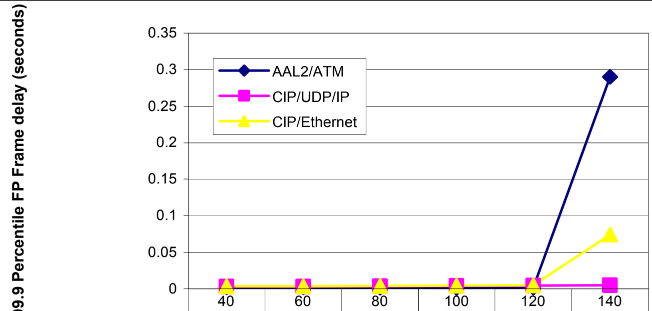


	Voice	Video 64K	Video 128K	Data 64K	Retran. Data PDU
FIFO	0.002992021	0.004836098	0.005073879	0.006142947	0.006142947
Priority	0.000926509	0.00446902	0.006294887	0.022023182	0.022023182
Retran. Data PDU Priority	0.000932368	0.004646218	0.006925602	0.03699836	0.001608467

Figure 4.17 Mean delay for different priorities at 10% re-transmission ratio

4.6.2. Performance Comparison

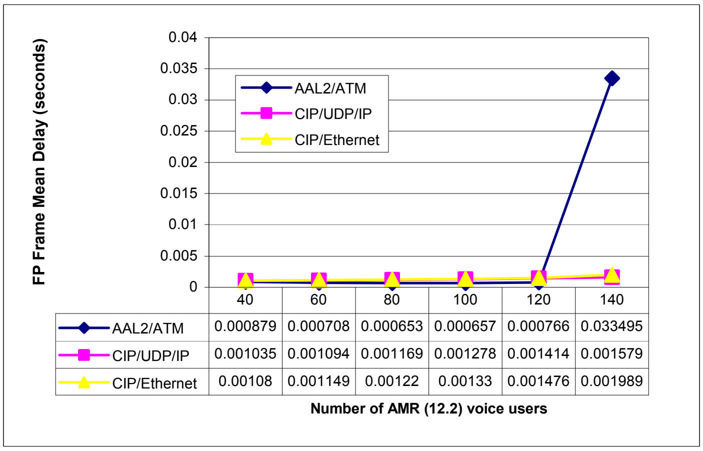
We find that the performance of an AAL2/ATM based link degrades faster than IP and Ethernet links with the increasing number of users (as shown in Figure 4.18 to Figure 4.23. Performance of an Ethernet based I_{ub} link is similar to the IP based link. For DCH voice and data traffic, due to small ATM payload size, the ratio of the payload to total transmitted bits change slightly with the increasing number of users. On the contrary, efficiencies of IP and Ethernet based links increases significantly due to the use of large payloads. Also, on the IP and Ethernet based links relatively fewer header bits are transmitted compared to an ATM based links. Results also show that IP and Ethernet based links appear to perform better for heavy traffic loading.



	40	60	80	100	120	140
AAL2/ATM	0.0022	0.0022	0.00183	0.00141	0.00177	0.29
CIP/UDP/IP	0.0032	0.0032	0.00344	0.00375	0.00424	0.00483
CIP/Ethernet	0.003542	0.003595	0.00384	0.004133	0.004745	0.07433

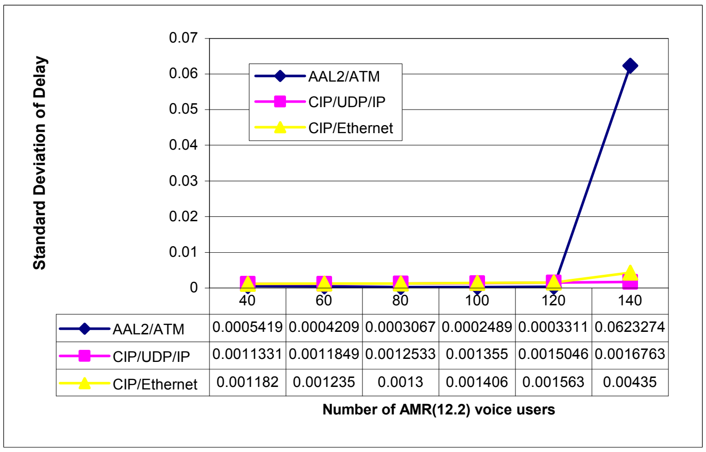
Number of AMR (12.2) voice users

Figure 4.18 99.9 percentile of voice FP frame delay



	40	60	80	100	120	140
AAL2/ATM	0.000879	0.000708	0.000653	0.000657	0.000766	0.033495
CIP/UDP/IP	0.001035	0.001094	0.001169	0.001278	0.001414	0.001579
CIP/Ethernet	0.00108	0.001149	0.00122	0.00133	0.001476	0.001989

Figure 4.19 Mean delay of voice FP frame



	40	60	80	100	120	140
AAL2/ATM	0.0005419	0.0004209	0.0003067	0.0002489	0.0003311	0.0623274
CIP/UDP/IP	0.0011331	0.0011849	0.0012533	0.001355	0.0015046	0.0016763
CIP/Ethernet	0.001182	0.001235	0.0013	0.001406	0.001563	0.00435

Figure 4.20 Standard deviation of voice FP frame

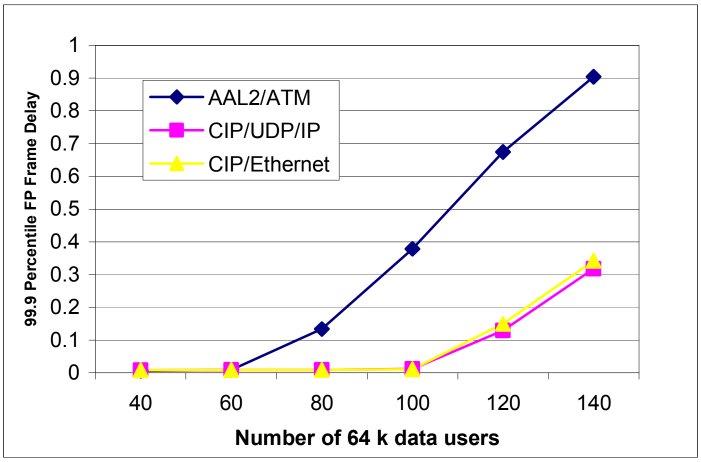


Figure 4.21 99.9 percentile of 64 kbps data FP frame delay

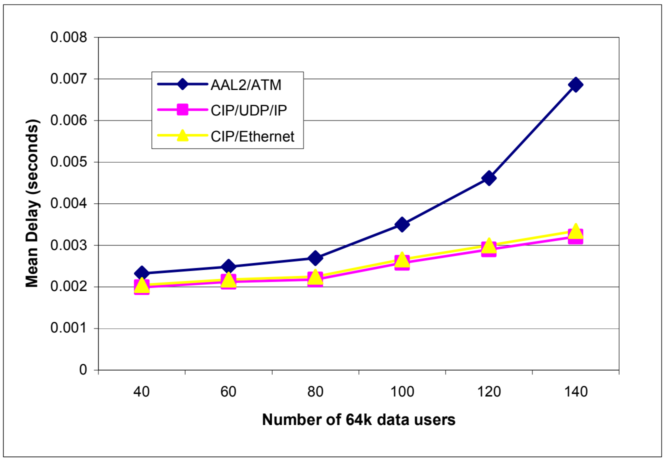


Figure 4.22 Mean delay of 64kbps data FP frame

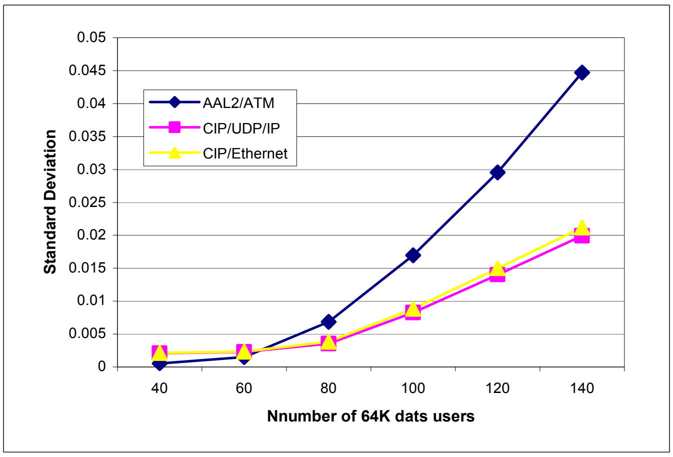


Figure 4.23 Standard deviation of 64kbps data FP frame

Next we examine the performance of HSDPA traffic on the I_{ub} link. In this simulation we used HSDPA UE category 12 to analyze the efficiency of various I_{ub} links. When the HS-DSCH is operating at the theoretical peak rate, the highest bursty traffic flowing is brought over the I_{ub} link. When Category 12 UEs are used, in each TTI (2ms), up to 10 RLC/MAC-d blocks are sent, if there is an active user. Table 4.6 shows the size of a FP frame for category 12 UE operating at the peak rate.

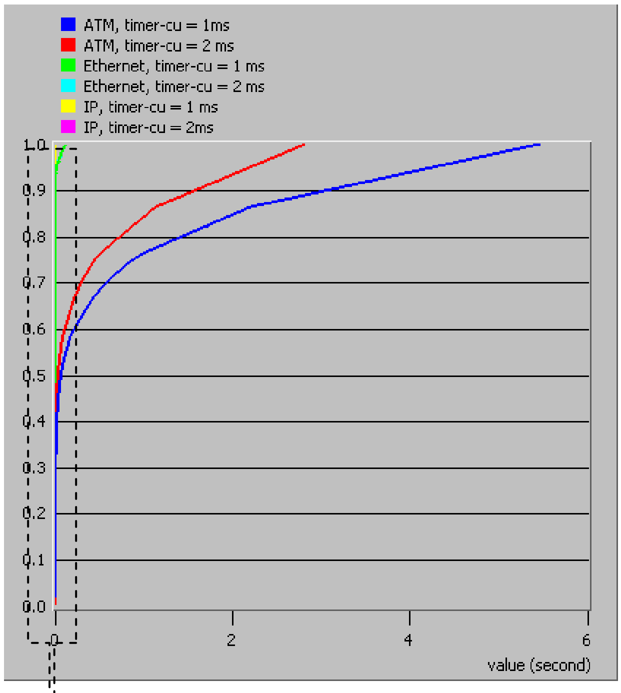
A RLC block size of 320 bits is used in the simulation with a 16 bit header as defined by the UE 12 category specification. Ten RLC PDUs are sent per 2ms TTI when a HS-DSCH link is operating at the theoretical peak rate. The total FP frame size of 443 bytes is used. To transmit the FP frame on an ATM based link the FP frame is segmented in to number of blocks and encapsulated into cells using the frame packing algorithm described in Figure 4.4. Similarly an Ethernet and IP packet is generated using the packing algorithm described in Figure 4.5 and Figure 4.6 respectively. Efficiency of an I_{ub} link transport mechanism will depend on the ratio of the payload to total transmitted bits which includes header bits. For an ATM based link the number of header bits transmitted in each FP frame will be higher compared to an IP or an Ethernet based link. To transmit a 443 byte FP frame an ATM based link needs to transmit 11 ATM cells giving overall link efficiency of 76%. On the other hand if the same FP frame is transmitted using a CIP/UDP/IP frame then only one IP packet is transmitted with an overall link efficiency of 97%. Similarly the CIP/Ethernet/GFP solution offers overall link efficiency of 92%.

Table 4.6 FP Frame Fields (byte)

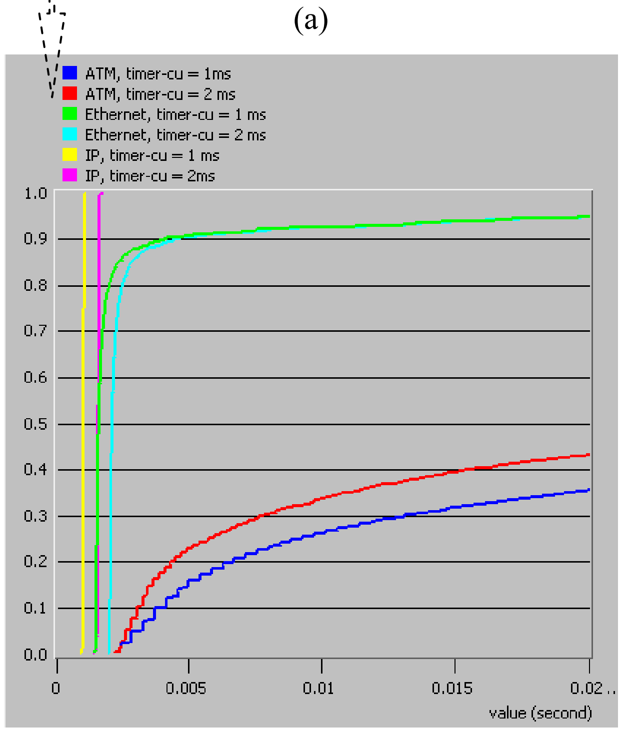
FP header	7
MAC-d	10 × 42
Spares bit and Padding	10 × 1
new IE flag, DRT, and CRC etc	6
Total	443

Figure 4.24 shows the CDFs (cumulative Distribution Function) of the FP frame delays for ATM, IP and Ethernet based links respectively for the HSDPA traffic. An E1 link is used between the RNC and node-B. Figure shows 95% delay is less than 3, 0.02 and 0.0025 seconds for ATM, Ethernet and IP based I_{ub} links respectively when Timer-CU is set to 2 ms. Average delay of the Ethernet based link is much shorter than the ATM based link. The reduction of delay is mainly due to lower segmentation and queuing delays. Average delay of IP based link is shorter than Ethernet based link because header size of an IP packet is short than Ethernet frame when an IP header compression technique is used. Figure also shows that a shorter Timer-CU delay increases the ATM link delay, on the other hand IP and Ethernet link delays are reduced. The main reason for increased ATM link delay is more extra padding bits are used for transmission of cells due to a shorter waiting time. Padding bits increase traffic load. For the IP and Ethernet link multiplexing delay is the major delay component which reduces

due to the use of a shorter Timer-CU value. The consequence of the higher delay on ATM link means the node-B can't sustain the peak rate on the air interface.



(a)



(b) Enlarging the dashed part of (a)

Figure 4.24 . CDF of delay on ATM, IP and Ethernet based I_{ub} link for HSDPA

4.7 Summary

The chapter first discussed the UTRAN and the I_{ub} link protocol structure used to transmit multimedia traffic in UMTS networks. The FP layer and the transport network layer provide the connection between the MAC layer and the air interface physical layer. The user data frame is used to transparently transport the transport blocks (TBs) between the RNC and the Node-B. Coordinated DCH channels with the same transmission time are allowed to multiplex into one transport bearer.

When AAL2/ATM transport technology is used to transmit UMTS voice and data traffic via the I_{ub} link, the Timer-CU value has important impact on the performance of the link under low traffic load conditions, since multiplexing and de-multiplexing delay dominate the total delay. Low Timer CU values could decrease the packets delay to. With the increasing of traffic load, the timer-CU value becomes less important. But for very low Timer-Cu values the packet delay rapidly increase due to the use of padding bits to multiplex different streams even in a high traffic load situation. This directly increases the peak cell rate (PCR), which causes a higher delay. The AAL2 packets loss ratio increases faster than ATM cells with the increasing traffic intensity. If the retransmission traffic is offered appropriate priority then the end-to-end delay can be significantly reduced thus improving the QoS of data traffic. For future mobile networks it is necessary to accurately assess the needs of a RAN. Significant bottleneck could be generated by RAN and core networks. Retransmission priority could improve the overall delay of data traffic.

For DCH voice and data traffic, due to small ATM payload size, the ratio of the payload to total transmitted bits change slightly with the increasing number of users. On the contrary, efficiencies of IP and Ethernet based links increases significantly due to the use of large payloads. Also, on IP and Ethernet based links relatively fewer header bits are transmitted compared to an ATM based links. Results also show that IP and Ethernet based links appear to perform better under heavy traffic load situation. For the HSDPA traffic, average delay of the Ethernet based link is much shorter than an ATM based link, average delay of an IP based link is shorter than Ethernet based link because of the header size of an

IP packet is shorter than the Ethernet packet header when an IP header compression technique is used. The Ethernet based link is an alternative transport technique that can be used for the UTRAN I_{ub} link due to its flexibility, economy and bandwidth efficiency.

Chapter 5

Performance Analysis of the HSDPA Radio Access Network

5.1 Introduction

The HSDPA (High Speed Downlink Packet Access) is one of the initial evolutionary steps towards a high capacity data centric network. HSDPA connections have been designed to offer high data rate transmission on the downlink. Currently the HSDPA networks can offer peak transmission rate of 28.8 Mbps and in very near future the peak data rate will reach 40 Mbps. A peak data rate connection requires extremely favourable radio transmission conditions. It is unlikely that all users in a cell would be able to get such a high data rate, even if only one user was using the cell at a time. In a HSDPA network user data rate will vary with transmission conditions, and also due to time and code sharing among users. Since the average traffic volume on the I_{ub} link is usually less than the peak rate over the air interface, it is not economical or efficient to reserve the I_{ub} link bandwidth at the potential peak rate for the duration of a call. The overall throughput and quality of service (QoS) of a HSDPA connection will depend on the selection air interface condition and the transport link conditions.

The HSDPA architecture incorporates a downlink buffer and scheduler in the Node-B to control the throughput of downlinks. A flow controller is generally employed on the I_{ub} link between a radio network controller (RNC) and a Node-B to avoid buffer overflow or buffer starvation. If an I_{ub} link cannot provide high enough instantaneous bandwidth then sufficient number of packets may not be supplied to the Node-B from the RNC to match the service rate of the downlink buffer at a particular instant. In such an event it is possible that the buffer might run out of queued packets and the peak transmission rate over the air interface cannot be sustained [6]. A peak data rate transmission is only possible when an HSDPA receiver reports an appropriate CQI (Channel Quality Indicator) based on the SNR value of a link. A HSDPA connection can be seen as a two hops link consist of a RNC to Node-B link and a Node-B to UE link. Hence, the overall throughput of a HSDPA connection will depend on the capacity of both links.

The instantaneous bandwidth of the I_{ub} link is affected by the choice of transport protocol architecture and the physical transmission rate. The traffic flow between a RNC and a Node-B is controlled by the frame transmission protocol on the I_{ub} link. As described in Chapter 4, currently several transport protocols exist which can be deployed on the I_{ub} link. The 3GPP has chosen AAL2/ATM and IP based transport network to transmit multimedia traffic on the I_{ub} interface to support packet based services. Recently the Carrier Ethernet protocol has been suggested for the deployment in wide area networks due to its relatively lower cost, simplicity and flexibility. The introduction of the Carrier Ethernet protocol in the RAN will introduce some flexibility in the UMTS network. This discussion show that it is important to optimise the parameters of the HSDPA RAN and the air interface by studying the interdependency of these parameters and designing these links based on the study. This chapter investigates the interdependency of HSDPA the air interface and RAN parameters using analytical and simulation studies.

Reviewing the literature in this area, several works have been found, which have concentrated on the design of the I_{ub} link flow control and analysed their efficiencies for the HSDPA traffic. Legg presented a strategy for the I_{ub} link flow control in which the Node-B scheduler could serve the maximum number of bits for every run of the scheduler. At the same time the flow control algorithm would have enough time to transfer another block of bits to the Node-B in time for the next transmission opportunity, thus preventing any stalling of transmission, whilst maintaining the minimum number of bits in the Node-B buffer to prevent any further stalling. Legg's study showed that the scheduler in the Node-B has a major influence on the amount of memory used at the Node-B [37]. Necker et. al. investigated the impact of the I_{ub} link flow control on the HSDPA system performance, and correlated larger protocol delay and resource grant update period between the Node-B and the RNC. The study showed that the deterioration in IP packet delay characteristics since the spatial separation of a RNC and a Node-B imposes significant signalling constraints. At the same time the control dead time limits the performance of the flow control mechanism [61]. Toskala et. al. studied the effect of the I_{ub} link efficiency on the HSDPA air interface [62]. The study shows a significant improvement in the I_{ub} link efficiency when the HSDPA air interface is used, compared to dedicated channels when delivering similar services. Above studies tried to optimise the performance of the I_{ub} link based on the RAN requirements. In this work we optimise the performance of the I_{ub} link based on air interface and RAN requirements. This study changes the RAN design paradigm and introduces the joint optimisation of UTRAN parameters.

This chapter is organized as follows. In section 5.2, an I_{ub} link flow control algorithm and its corresponding Markov chain model is presented. The interactions between HSDPA air interface and RAN parameters are analysed in the section. A simulation model is introduced in section 5.3 which is used to further analyse the parametric relationship. Using both analytical and simulation models the HSDPA RAN performance is analysed in section 5.4. Section 5.5 present conclusions.

5.2 HSDPA Air Interface and RAN Parametric Interactions

As described in previous chapter, to prevent any Node-B buffer overflow or buffer starvation a flow controller is employed between a RNC and a Node-B to control traffic on the link. The RNC sends capacity requests to the Node-B to request the HS-DSCH capacity. The Node-B allocates capacity to the RNC indicating the maximum MAC-d PDU length, numbers, HS-DSCH interval and repetition etc. to control the user data flow. The 3GPP technical report regarding to Evolved UTRAN (E-UTRAN) has specified the transmission delay between a UE and the RNC should be lower than 10 ms [63] [64]. The delay limit on the downlink is expected to be even shorter than the uplink. When capacity allocated by the Node-B is higher than the I_{ub} link capacity, the risk of congestion on the I_{ub} link could increase, which leads to packet delays increasing and higher probability of packet loss, which in turn could cause RLC layer retransmission, and degradation of HSDPA performance. To ensure a lower transmission delay between the RNC and a UE on the downlink it is necessary to employ a flow control algorithm between these two nodes, which allocates transmission resources on the I_{ub} link to pro-actively avoided congestion on the transport network. .

5.2.1. I_{ub} link Flow Control Algorithm

In the introduction section of this chapter the importance of the I_{ub} flow control algorithm is introduced which need to be used to maintain a HSDPA connection throughput and appropriate QoS. Figure 5.1 shows the basic architecture of the proposed joint flow control algorithm. As shown in the figure, the RNC maintains a large buffer of information packets received from the core network. The Node-B maintains a minimum buffer size which enables the use of peak transmission rates for HSDPA users on the air interface. In the proposed algorithm for each HSDPA connection is allocated with a fixed buffer space in the main Node-B buffer. A connection buffer is allocated based on the connection's peak data rate requirement. The flow control algorithm continuously updates the Node-B buffer based on the HSDPA connection throughput and the connection demand. We assume that *n* represents a connection buffer size in the Node-B buffer, where *n* corresponds to the total number of MAC-d PDUs that could be sent over the air interface at the peak rate in every TTI (2ms) under the best transmission channel condition. The I_{ub} link bandwidth is assumed as *m*, where *m* is the number of MAC-d PDUs sent in each TTI over the link. After each air-interface transmission event on the air interface, the Node-B immediately requests the RNC to send more MAC-d PDUs to replenish its buffer. The request size R_s in number of PDUs is equal to the free buffer size. If the requested size is larger than the I_{ub} link bandwidth *m*, i.e. $R_s>m$, then only *m* MAC PDUs are delivered to the Node-B buffer since attempting to deliver a larger value would lead to the I_{ub} link congestion and even greater delay in refilling the buffer. The value of R_s could vary between 0 to *n*. To avoid low I_{ub} link utilization we maintain the following relationship $m<n$. Using the above flow control algorithm the Node-B buffer is used to adjust the HSDPA user throughput on the air interface based on the radio link condition. A packet scheduler at the Node-B will be used to transmit data on HSDPA connections which will determine the throughput of a connection. It is possible that a peak air interface transmission rate may not be sustainable due to the I_{ub} link and the air interface transmission rate mismatch. It is worthwhile to mention here that the air interface throughput is radio channel condition dependent and varies over time.

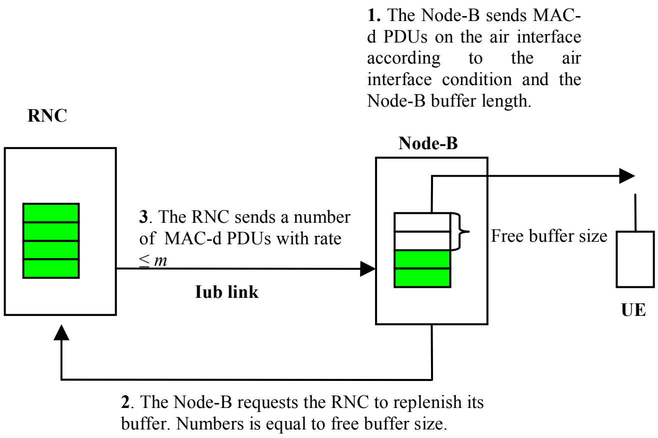


Figure 5.1 flow control between Node-B and RNC

5.2.2. HSDPA Air Interface Transmission Efficiency

In this section we introduce a new performance measure which shows the dependency of the air interface and the I_{ub} link parameters. The new performance measure is the HSDPA transmission channel efficiency η which is the ratio of the actual number of MAC PDUs available for transmission in the Node-B buffer to the potential maximum number of MAC PDUs that can be carried by the air interface condition per TTI, as shown in equation (5.1). The HSDPA transmission efficiency takes into account the air interface condition, buffer length, and the I_{ub} link effective bandwidth. For example, on one occasion, if 8 PDUs of a HSDPA connection is available in the Node-B buffer but the radio link condition permits transmission of 10 PDUs then we obtain a η value of 0.8 (8/10). In that case when a transmission opportunity arises the Node-B cannot transmit the maximum number of PDUs. This situation can be avoided by increasing the fixed allocation of bandwidth on the I_{ub} link but that will reduce the I_{ub} link utilization.

$$\eta = \frac{Actual\_MAC\_PDU\_Tx}{Max\_MAC\_PDU\_air} \qquad (5.1)$$

5.2.3. Analytical Model

In this section a Markov chain based analytical model is set up to examine the interdependency of air interface and the I_{ub} link parameters. Due to the time-variant nature of radio channels, the transmission data rate of the air interface is highly variable. We assume that the transmission rate (expressed in number of MAC-d PDUs per TTI) of a HSDPA connection is distributed between 0 and a maximum number n. The value of n depends on the UE category. Let r represent the air interface rate, which the radio channel condition can support in terms of number of PDU. Let R_i represents the probability of the air interface transmission rate, where i represent the number of PDUs transmitted at that rate. After the Node-B transmits PDUs on the air interface, it is assumed that the Node-B buffer immediately requests the RNC to send packets to replenish its buffer before the next transmission opportunity because the flow control algorithm introduced in previous section restrict the number of MAC-d PDUs in per TTI not more than the I_{ub} bandwidth. It guarantees low delay of the I_{ub} link. Let l_n be the number of MAC-d PDUs in the Node-B buffer measured after the buffer is replenished and the next transmission opportunity over the air interface in the n-th TTI is pending. Since the I_{ub} link bandwidth is m, at least m PDUs are available to be sent from the Node-B buffer after the replenishment of the buffer. The buffer length could vary between m and n, i.e. $m \leq l_n \leq n$. We can form a Markov chain using the values of l_n. Consider a stationary state where all $\{l_n\}$ have the same probability distribution. Let l denote the random variable in MAC-d PDU numbers in the Node-B buffer. The transition matrix is shown in equation (5.2).

Event $l_n = m$ happen

When $l_{n-1} = m$, the air interface rate $n \geq r_{n-1} \geq m$;

Or, when $l_{n-1} = m+1$, the air interface rate $n \geq r_{n-1} \geq m+1$;

$\vdots$

Or, when $l_{n-1} = n$, air interface rate $r_{n-1} = n$;

Event $l_n = m+1$ happen

When $l_{n-1} = m$, the air interface rate $r_{n-1} = m-1$;

Or, when $l_{n-1} = m+1$, the air interface rate $r_{n-1} = m$;

$\vdots$

Or, when $l_{n-1} = n$, air interface rate $r_{n-1} = n-1$;

$\vdots$

Event $l_n = n$ happen

When $l_{n-1} = m$, the air interface rate $0 \leq r_{n-1} \leq 2m-n$;

Or, when $l_{n-1} = m+1$, the air interface rate $0 \leq r_{n-1} \leq 2m-n+1$;

$\vdots$

Or, when $l_{n-1} = n$, air interface rate $r_{n-1} \leq m$;

Here we assume $m \leq n \leq 2m$,

(Note: if $n > 2m$, I_{ub} link bandwidth m is set as less than half air interface peak rate n, the transition from Node-B buffer occupancy status $l = m$ to $l = n$ doesn't happen)

Then the transition matrix P is shown below[65], [66].

$$\begin{pmatrix} \sum_{i=m}^{n} R_i & R_{m-1} \ \dots \ R_{2m-n+1} & \sum_{i=0}^{2m-n} R_i \\ \sum_{i=m+1}^{n} R_i & \ddots & \sum_{i=0}^{2m-n+1} R_i \\ \vdots & & \vdots \\ R_n & R_{n-1} \cdots R_{n-m-1} & \sum_{i=0}^{m} R_i \end{pmatrix} \quad (m \le n \le 2m) \tag{5.2}$$

Let, $\pi_i = P\{l = m+i\}, \qquad i = 0,....,n-m$ (5.3)

In the stationary state, we can write $\pi = \pi P$, so

$$\pi_0 = \pi_0 \sum_{i=m}^{n} R_i + \pi_1 \sum_{i=m+1}^{n} R_i + ... + \pi_{n-m} R_n$$
$$\pi_1 = \pi_0 R_{m-1} + \pi_1 R_m + ... + \pi_{n-m} R_{n-1}$$
$$\vdots$$
$$\pi_{n-m-1} = \pi_0 R_{2m-n+1} + \pi_1 R_{2m-n} + ... + \pi_{n-m} R_{n-m-1}$$
$$\pi_{n-m} = \pi_0 \sum_{i=0}^{2m-n} R_i + \pi_1 \sum_{i=0}^{2m-n+1} R_i + ... + \pi_{n-m} \sum_{i=0}^{m} R_i$$

$$\pi_0 + \pi_1 + \cdots + \pi_{n-m} = 1 \tag{5.4}$$

When $l = m$,

if $r \le m$, $\eta_{l=m} = 100\%$;

if $r = m+1$, $\eta_{l=m} = \dfrac{m}{m+1}$;

$\vdots$

if $r = n$, $\eta_{l=m} = \dfrac{m}{n}$;

$\vdots$

When $l = n\text{-}1$,

if $r \le n\text{-}1$, $\eta_{l=n-1} = 100\%$;

if $r = n$, $\eta_{l=n-1} = \dfrac{n-1}{n}$;

When $l = n$,

$\eta_{l=n} = 100\%$;

So, the average HSDPA transmission efficiency is given by Equation (5.5):

$$\eta = \pi_0(\sum_{i=0}^{m} R_i \times 1 + R_{m+1} \times \frac{m}{m+1} + ... + R_n \times \frac{m}{n}) +$$
$$\pi_1(\sum_{i=0}^{m+1} R_i \times 1 + R_{m+2} \times \frac{m+1}{m+2} + ... + R_n \times \frac{m+1}{n}) \qquad (5.5)$$
$$+ .... + \pi_{n-m} \times 1$$

5.2.4. Air Interface Model

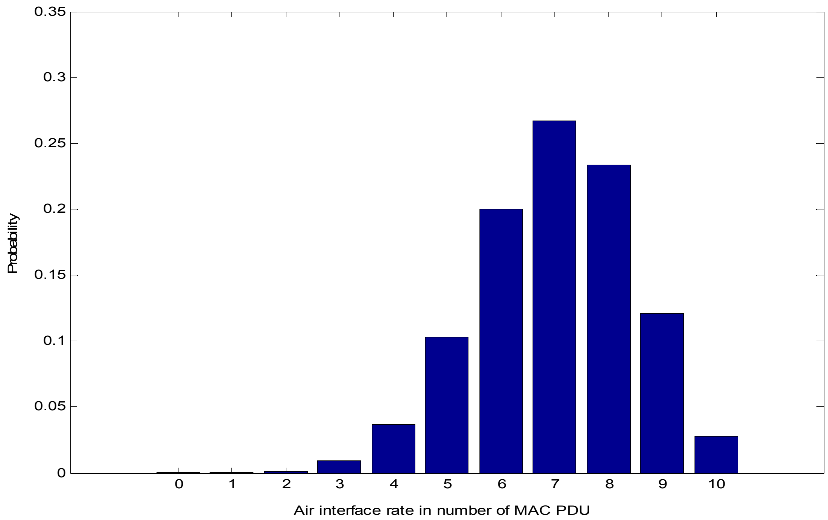
From the statistical point of view, the air interface capacity of a cell will be approximately normally distributed according to the Central Limit Theorem. Since the distribution of air interface rate is discrete, to simplify our analysis, we approximate the HSDPA air interface data rate using the following binomial distribution (5.6) [67], [68].

$$R_i = \binom{n}{i} p^i (1-p)^{n-i} \quad 0 < i \le n; \quad 0 < p \le 1 \qquad (5.6)$$

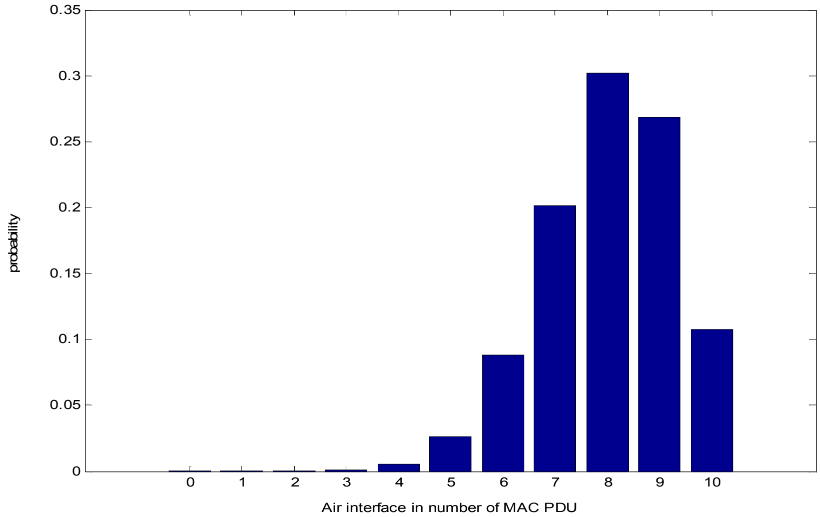
$$mean = n \cdot p \qquad (5.7)$$

$$HSDPA\ average\ throughput\ at\ RLC\ layer \qquad (5.8)$$
$$= n \times p \times RLC\_Packet\_size / TTI$$

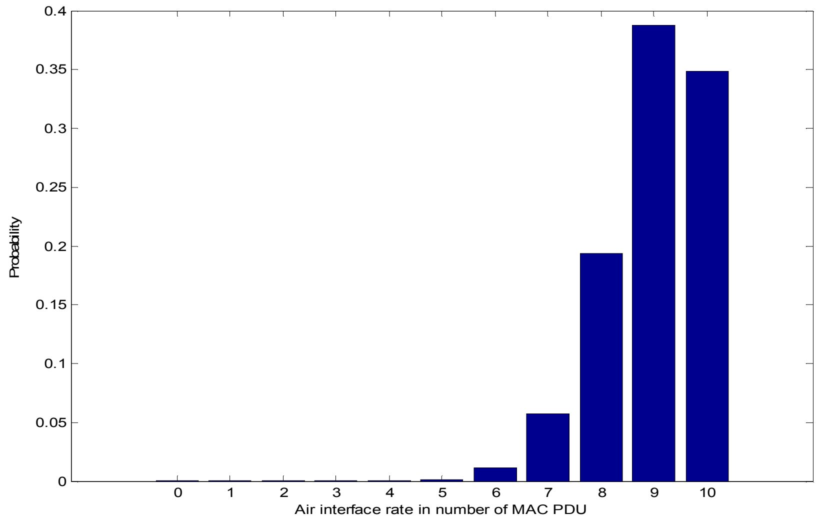
Where R_i represents the probability of the successful PDU delivery rate over air interface, i represents the number of PDUs transmitted at that rate, and n is equal to the maximum number of MAC-d PDUs that can be sent in every TTI. The value of p is determined by physical air interface conditions. In this analysis RLC packet size is fixed at 320 bits. The retransmission isn't considered to simplify the analysis. We assume that a very strong and controlled coverage in a small cell, and low user mobility could lead to higher value of p, which represents higher air interface transmission rate for users. The best radio channel condition is represented by the value of $p = 1$. Figure 5.2 shows the probability distribution of air interface transmission rate for $p = 0.7$, 0.8 and 0.9, and $n = 10$ for a category 12 UE. These plots show that a HSDPA connection may transmit a number of PDUs per TTI at different times. These plots show that for $p = 0.8$ there is a high probability that a UE will transmit 7 to 9 PDUs per TTI more often than other values. The peak of the plot will shift towards left for lower values of *p*. Similar results were reported by an Ericsson research group that measured HSDPA network performance in different transmission conditions [69]. Figure 5.3 shows the result of stationary testing with good signal strength close to the transmission station. Figure 5.4 shows stationary test result with poor signal strength at the edge of a cell. The test area for measurements consisted of 40 HSDPA sites, which generally separated by a site to site distance of 500 m. The test environment was typical urban with a mix of offices, shopping and restaurant areas, living areas, and an open area. Category 12 terminals are deployed for the measurement. All measurements were made on layer-2 (MAC-hs) which, given a 10% retransmission rate. The maximum bit rate is 1.5 Mbps. The median bit rate 0.9 Mbps was achieved in the stationary testing with poor signal strength at the edge of the WCDMA coverage area.



(a) $p = 0.7$



(b) $p = 0.8$



(c) $p = 0.9$

Figure 5.2 Probability distribution of the air interface transmission rate represented in number of MAC-d PDUs for UE Category 12

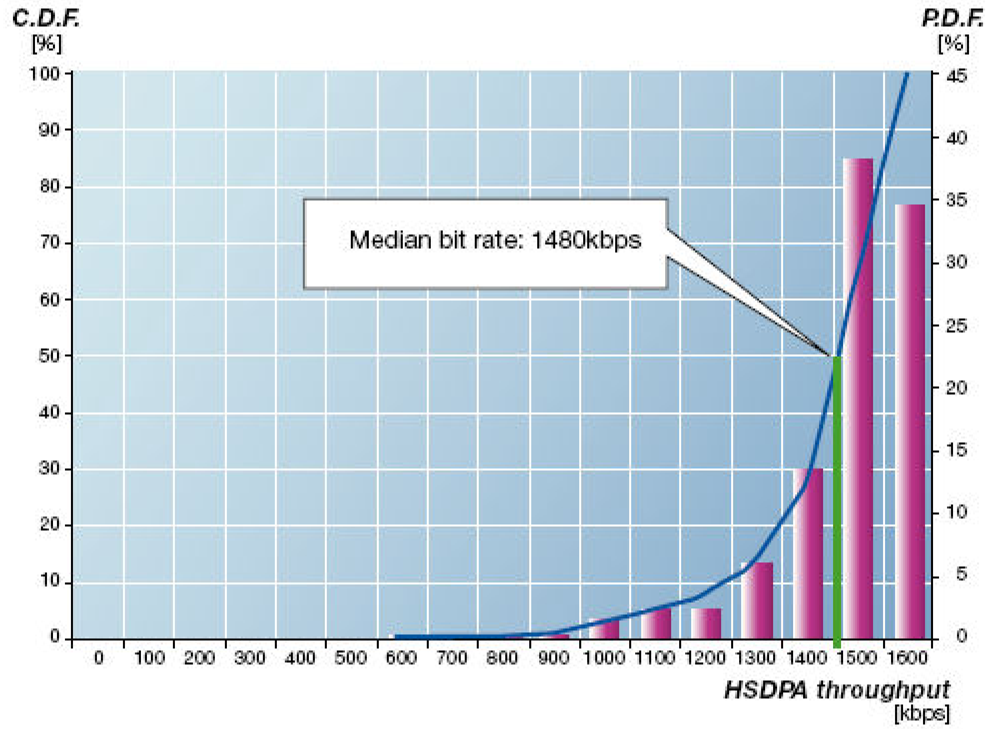


Figure 5.3 HSDPA bit-rate: stationary during good radio conditions [69]

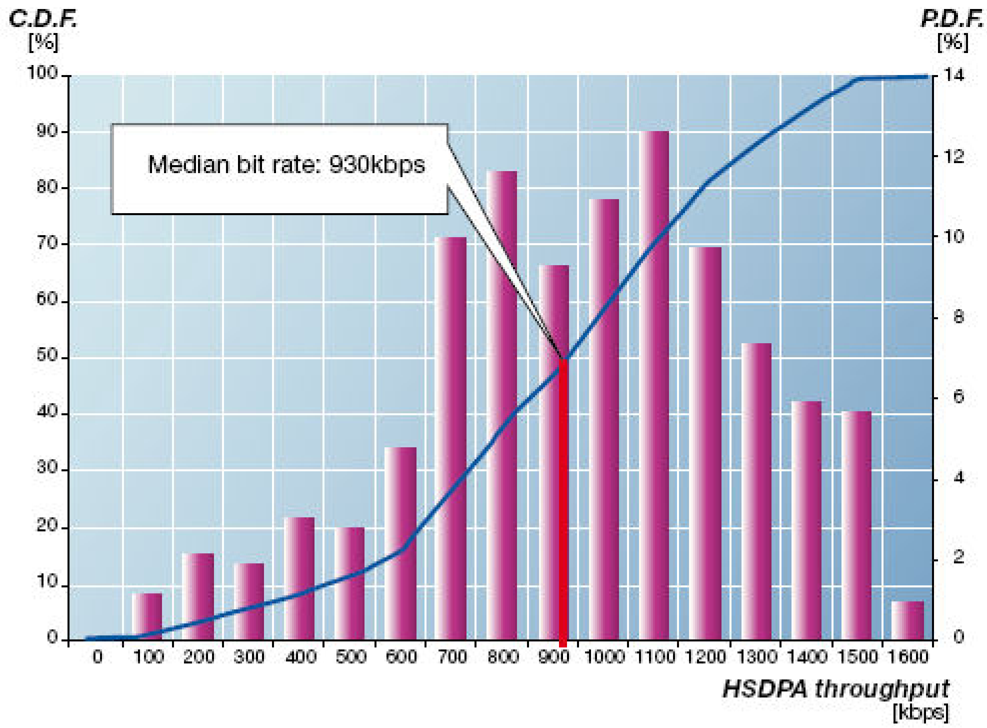


Figure 5.4 HSDPA bit-rate: stationary during poor radio conditions [69]

5.2.5. Example Calculation

In this section, we use a single HSDPA cell, which supports Category 12 and 5/6 UEs separately, as an example to analyze the relationship between the HSDPA transmission efficiency and the I_{ub} link bandwidth. A fixed RLC/MAC PDU size of 320 bits is used. We assume that an AAL2/ATM based I_{ub} link connects the RNC and the Node-B.

d) Example 1:

First we use we use a scenario where only Category 12 UEs are supported in a single HSDPA cell. For the category 12 UE, the maximum number of RLC/MAC PDU for the peak rate transmission is set to 10. The physical link data rate of an AAL2/ATM based I_{ub} link of 2.048 Mbps is used. Considering RLC/MAC, AAL2/ATM and FP frame headers, the value of m is set to 8 MAC-d PDUs per TTI.

Using the analytical technique in the above section, the HSDPA transmission efficiency η for different cell parameters could be calculated. For a category 12 UE, the air interface transmission rate is distributed between 0 and 10 in number of MAC-d PDUs. These PDUs are transmitted from the Node-B according to the current air interface data rate. After every successful transmission the Node-B buffer is replenished by the RNC. Since the I_{ub} link effective bandwidth m is 8 MAC-d PDU per TTI, at least 8 PDUs are waiting in the Node-B buffer for transmission in the next turn. The size of the Node-B buffer l could vary from 8 to 10. A transition matrix can be constructed using the above values. Figure 5.5 illustrates the state transition diagram.

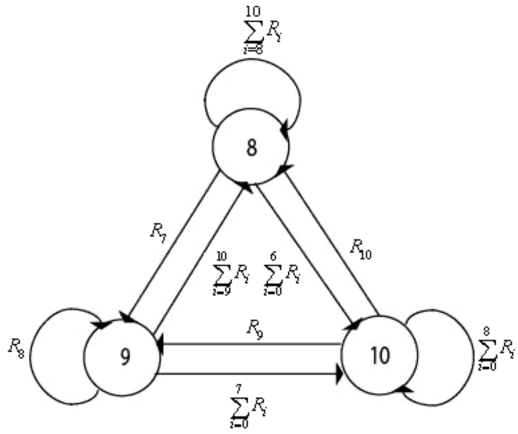


Figure 5.5 state transition diagram

The transition matrix P is shown by the equation (5.9).

$$P = \begin{pmatrix} \sum_{i=8}^{10} R_i & R_7 & \sum_{i=0}^{6} R_i \\ \sum_{i=9}^{10} R_i & R_8 & \sum_{i=0}^{7} R_i \\ R_{10} & R_9 & \sum_{i=0}^{8} R_i \end{pmatrix} \qquad (5.9)$$

Using equation (5.6), the value of R_i for different air interface rate r is calculated. When $p = 0.8$, and $n = 10$, the transition matrix P shown below.

$$P = \begin{pmatrix} 0.6778 & 0.2013 & 0.1209 \\ 0.3758 & 0.3020 & 0.3222 \\ 0.1074 & 0.2684 & 0.6242 \end{pmatrix}$$

$$\pi_0 = P\{l = 8\}$$
$$= 0.6778\pi_0 + 0.3758\pi_1 + 0.1074\pi_2$$
$$\pi_1 = P\{l = 9\}$$
$$= 0.2013\pi_0 + 0.3020\pi_1 + 0.2684\pi_2$$
$$\pi_2 = P\{l = 10\}$$
$$= 0.1209\pi_0 + 0.3222\pi_1 + 0.6242\pi_2$$

$$\pi_0 + \pi_1 + \cdots + \pi_{n-m} = 1$$

Hence,

$$\pi_0 = 0.406,\ \pi_1 = 0.249,\ \pi_2 = 0.345$$

Using the equation (5.5), the average HSDPA transmission efficiency η is calculated as shown below.

$$\eta = \pi_0(\sum_{i=0}^{8} R_i \times 1 + R_9 \times \frac{8}{9} + R_{10} \times \frac{8}{10}) + \pi_1(\sum_{i=0}^{9} R_i \times 1 + R_{10} \times \frac{9}{10}) + \pi_2 \times 1$$
$$= 97.65\ \%$$

Similarly the value of η can be calculated for different values of p which represents different transmission conditions. Table 5.1 presents the value of η for different Node-B buffer sizes and transmission channel conditions.

Table 5.1 Calculation Results (UE Cat.12)

p		0.5	0.6	0.7	0.8	0.9	1.0
π	l=8	0.001	0.009	0.072	0.406	0.902	1
	l=9	0.010	0.046	0.148	0.249	0.075	0
	l=10	0.989	0.945	0.780	0.345	0.023	0
η (%)		100%	99.99	99.82	97.65	89.57	80

e) Example 2:

In this example the same calculation is repeated for the Category 5/6 UEs. For category 5/6 UE, the maximum number of RLC/MAC PDU transmitted at a peak rate is set to 21. The physical link data rate of an AAL2/ATM based I_{ub} link is 4.096 Mbps (2 X E1 links). The I_{ub} link bandwidth m is set to 18 MAC-d PDUs per TTI. The size of Node-B buffer l is in state l = 18, 19, 20 and 21.

The transition matrix P is shown as the equation (5.10).

$$P = \begin{pmatrix} \sum_{i=18}^{21} R_i & R_{17} & R_{16} & \sum_{i=15}^{0} R_i \\ \sum_{i=19}^{21} R_i & R_{18} & R_{17} & \sum_{i=16}^{0} R_i \\ \sum_{i=20}^{21} R_i & R_{19} & R_{18} & \sum_{i=17}^{0} R_i \\ R_{21} & R_{20} & R_{19} & \sum_{i=18}^{0} R_i \end{pmatrix} \qquad (5.10)$$

The following Table 5.2 shows results.

Table 5.2 Calculation Results (UE Cat.5/6)

p		0.5	0.6	0.7	0.8	0.9	1.0
π	l=18	0.0000	0.0000	0.0009	0.0457	0.7570	1
	l=19	0.0000	0.0003	0.0058	0.0774	0.1321	0
	l=20	0.0001	0.0021	0.0231	0.1412	0.0675	0
	l=21	0.9999	0.9976	0.9702	0.7357	0.0433	0
η (%)		100	100	100	99.91	95.41	85.71

Above results show that for an ATM based I_{ub} link the higher HSDPA efficiency η can be achieved for lower values of p. The value of η decreases with the increasing value of p because for a better channel condition the average value of R_s (the request size) becomes larger than m, hence, the I_{ub} link is unable to supply sufficient number of MAC-PDUs within the specified duration. The effect of the I_{ub} link characteristics is also influenced by the scheduling mechanism at Node-B. For example, a round robin (RR) scheduler will impose less restriction on the I_{ub} link than a proportional fair (PF) scheduling scheme. When using a PF scheduler the same UE can receive multiple transmission opportunities in successive TTIs. These issues are further studied in the next chapter.

5.3 Simulation Model

To verify analytical model an OPNET based simulation model has been developed to analyze the effect of the I_{ub} link characteristics on HSDPA transmission efficiency. In the following sections the simulator structure and three I_{ub} link transport protocols are briefly described.

5.3.1. Simulator Architecture

The simulator consists of an I_{ub} interface and the I_{ub} physical link, a traffic generator, an air interface link and the proposed flow controller as shown in the Figure 5.6. It is assumed that the Core Network (CN) can deliver enough packets to the RNC and there are always enough packets waiting in the RNC buffer during the active period of a data burst. The flow controller is implemented as description of section 5.2.1. The Node-B allocates each HSDPA connection a transmission bandwidth according to the radio channel condition which will reported by an UE using the CQI data. For example, some users may experience good transmission conditions and hence, able to receive data at its peak rate; whereas other UE may receive data at an average transmission rate because of poor channel condition. The simulator model the air interface using the binomial distribution function as described earlier in the section 5.2.4. In the simulator each HSDPA connection data is drained by their respective traffic sink shown on the right of the Figure 5.6. These traffic sinks simulate variable radio channel conditions for different HSDPA connections. After every TTI the Node-B sends a request packet to the RNC for more user data packets to replenish its empty or partially empty buffer. The size of the request will depend on the air interface throughput. On the RNC side the RLC/MAC layer header are added to form MAC PDUs, and then packed into a FP frame. A FP frame is either segmented into a number of AAL2/ATM cells or assembled into a large IP or Ethernet frame. At the Node-B RLC/MAC PDUs are extracted and stored in the Node-B buffer [35] [34]. Key parameters for simulator are shown in Table 5.3.

Table 5.3 Key Simulation Parameters

Simulation Parameter	**Value/features**
UE category	5/6 and 12
No. of UE's	15
I_{ub} Link Data Rate	2.048 Mbps(Cat 12), 4.096Mbps (Cat 5/6)
I_{ub} Link Transport protocol	AAL2/ATM,UDP/IP and Ethernet
Maximum Buffer allocation for each connection	21 PDU/connection for Cat 5/6 UE, 10 PDU/connection for Cat 12 UE
Iub link capacity	18 PDU for Cat 5/6 UE, 8 PDU for Cat 12 UE
Scheduling Algorithm	RR
Air Interface model	Binomial distributed
Traffic model	Web Browsing, pareto distributed based on the UMTS traffic model
PDU size	320 bits

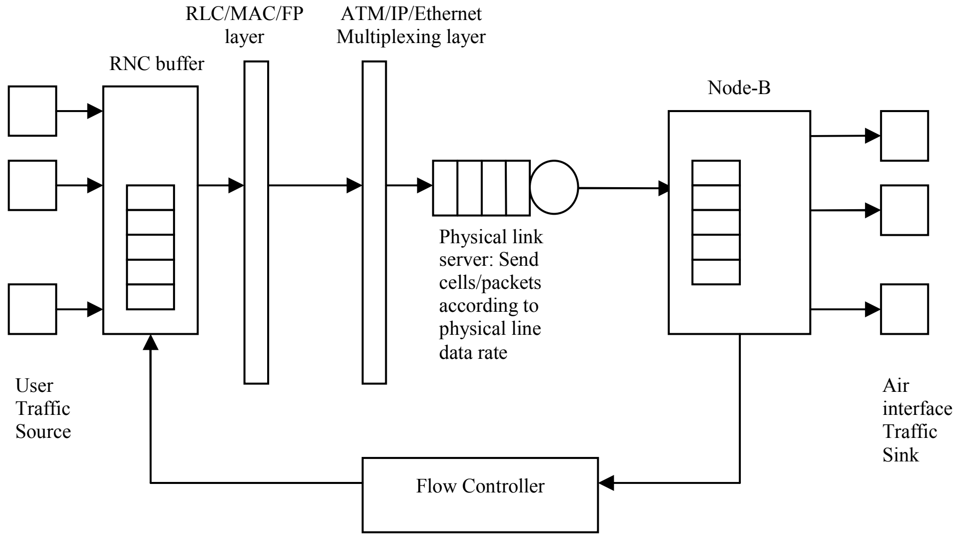


Figure 5.6 The simulator structure

5.3.2. I_{ub} Link Simulation

In this simulation three different I_{ub} transport protocols were simulated to transmit HS-DSCH FP (High Speed Dedicated Shared Channel Frame Protocol) frames between the RNC and the Node-B to carry user information and control signals. Various frame mapping and multiplexing structures for the I_{ub} link are described in section 4.3 which are used in the simulation model. An AAL2 based ATM connection is simulated to transfer HS-DSCH FP frames by mapping these frames on ATM cells using the AAL2 CPS (Common Part Sub-layer) [49]. For the IP based I_{ub} link simulation the CIP (Composite IP) option was used. The CIP allows multiplex variable size CIP packets to be transmitted in one variable size CIP container [10]. To improve the transmission efficiency we also implemented the UDP/IP header compression technique. The third I_{ub} transport protocol used to simulate the I_{ub} link is the Carrier Ethernet protocol. In this case the Generic Framing Procedure is used to map the HSDPA FP frame on an Ethernet frame for transmission from the RNC to Node-B [50].

5.3.3. Traffic Model

In this work the simulation model initially concentrated on a bursty web traffic transmission to study the effect of the I_{ub} links on the HSDPA transmission efficiency. The UMTS traffic model is used to develop the web browsing application [10]. As described in the section 4.4.2, the simulated web page file size is Pareto distributed with a parameter $\alpha = 1.1$. The traffic generator uses a mean file size of 12,000 bytes with a minimal file size of k=1858 bytes and the maximum file size m = 5000000 bytes. The reading time (the OFF time between consecutive burst) is exponentially distributed with a mean value of 12 seconds.

5.4 Interdependency between RAN and Air Interface

The simulation model was used to examine the interdependency of RAN and HSDPA connection parameters. Earlier discussions showed that the RNC to UE being a two hop links has a number of parameters which could influence the performance of the link. Using the simulation first the effect of the I_{ub} link capacity on the HSDPA efficiency factor is examined for different transmission parameters, and then validates analytical model results. The simulation model was then used to study the impact of RAN based transport protocols on HSDPA performance.

5.4.1. Impact of RAN Capacity on the HSDPA Connection QoS

Using the analytical technique in section 5.2, we can calculate the HSDPA transmission efficiency η for a category 12 UE with different air interface conditions and effective I_{ub} link capacity. Analysis results are shown as Figure 5.7. The figure shows that the value of η increases with the I_{ub} link bandwidth. The figure also shows that the degree of improvement depends on the air interface condition. In a realistic HSDPA network the air interface throughput will be determined by the radio link condition and available I_{ub} bandwidth. Hence, a fixed I_{ub} bandwidth allocation will either reduce the HSDPA air interface efficiency or it will reduce the I_{ub} link utilization.

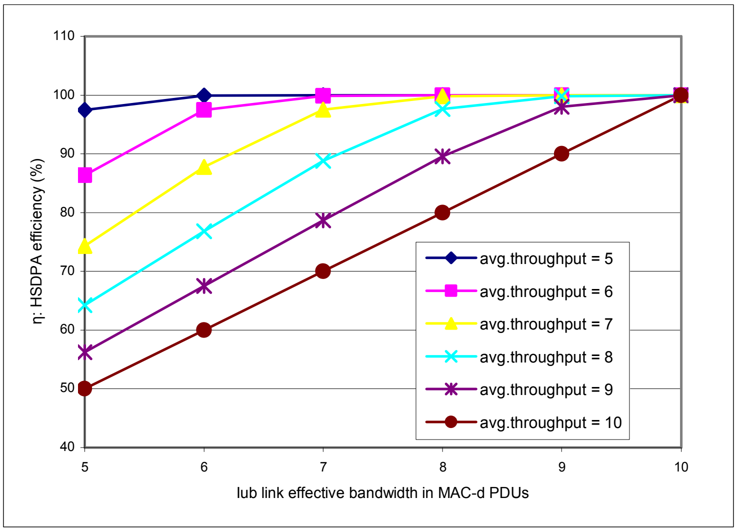


Figure 5.7 the interdependency between the I_{ub} effective link bandwidth and the HSDPA air interface efficiency. HSDPA average throughput is measured in number of MAC-d PDUs per TTI.

In this section, simulation results are compared with the analytical results. Figure 5.8 shows the relationship between the value of p (represents the radio transmission condition) and the HSDPA air interface efficiency η for Category 5/6 and Category 12 UEs. The figure shows that both analytical and simulation results are in close agreement. The figure shows that the value of η decreases as the transmission channel quality improves because the I_{ub} link creates the bottleneck that couldn't deliver enough MAC-d PDUs from RNC to Node-B to sustain high rate air interface packet transmission. In this simulation for Category 12 UE we used a 2.048 Mbps AAL2/ATM based I_{ub} which supports 8 MAC PDUs per TTI transmission between the RNC and the Node-B. For Category 5/6 UE the I_{ub} transmission link rate of 4.096 Mbps were used. The effective bandwidth of I_{ub} link is 18 MAC PDUs. Figure 5.8 shows that both category connections behaves similarly however, the Category 5/6 connections offer slightly better performance rate since the ratio of the Node-B buffer replenishment rate to drain rate is higher for Category 5/6 UE. For example, for the p value of 0.9, the ratio is 18/(21*0.9) = 0.952 for Category 5/6 UE whereas 8/(10*0.9) = 0.89 for Category 12 UE. Next we observe the relationship between the transmission channel condition and the combined air interface and the I_{ub} link throughputs. Figure 5.9 shows that the HSDPA air interface throughput can linearly increase with the increasing value of *p*. However, due to the I_{ub} link constraints, the air interface throughput remains limited at 1.3 Mbps for a 2.048 Mbps I_{ub} link based on the ATM transport protocol. On the other hand the graph also shows that the I_{ub} link utilization which is defined as the ratio of actual delivered MAC PDUs to the I_{ub} link effective bandwidth (in MAC PDUs) remains low for lower values of *p* because of lack of demand by the air interface. In this case if a high bandwidth the I_{ub} link is used then for an average transmission channel conditions the I_{ub} link utilization will drop. These results show that simply increasing the I_{ub} link capacity is not sufficient to increase the overall HSDPA air interface efficiency. Figure 5.10 shows both analytical and simulation results of the Node-B buffer occupancy. The graph shows that for the *p* value of 0.7 the buffer occupancy level is 10 MAC PDU for about 80% of the time whereas for p=0.9 for about 90% time the buffer length is 8 the minimum buffer size. The average HSDPA throughput (UE Category 12) is 7, 8 and 9 PDUs per TTI for p = 0.7, 0.8 and 0.9 separately. The effective bandwidth of AAL2/ATM based I_{ub} link is 8 PDUs per TTI when the E1 physical link is deployed. The reason for shorter buffer length for p= 0.9 is that the Node-B drain rate by the air interface is higher than the buffer replenishment rate by the I_{ub} link. The buffer length is measured after every transmission on the air interface and its subsequent buffer replenishment by the RNC. The graph shows that p= 0.8 is a balanced region where drain rate is matched by the replenishment rate. Figure 5.11 shows the distribution of PDU numbers in the Node-B buffer for Category 5/6 UEs when a 4.096 Mbps AAL2/ATM based I_{ub} link is used. The effective bandwidth of I_{ub} link is 18 PDUs. The result indicates that the buffer occupancy level is 21 MAC PDU for about 97% and 73% of the time for the *p* value of 0.7 and 0.8 respectively. The reason for high buffer occupancy for p=0.7 and 0.8 is that the Node-B drain rate (average drain rate is 14.7 and 16.8 PDUs per TTI for Category 5/6 separately) by the air interface is lower than the buffer replenishment rate (18 PDUs per TTI) by the I_{ub} link. The graph shows that the buffer occupancy level is 21 MAC PDUs for less than 7% of time when p = 0.9, since the replenishment capability is lower than the drain rate.

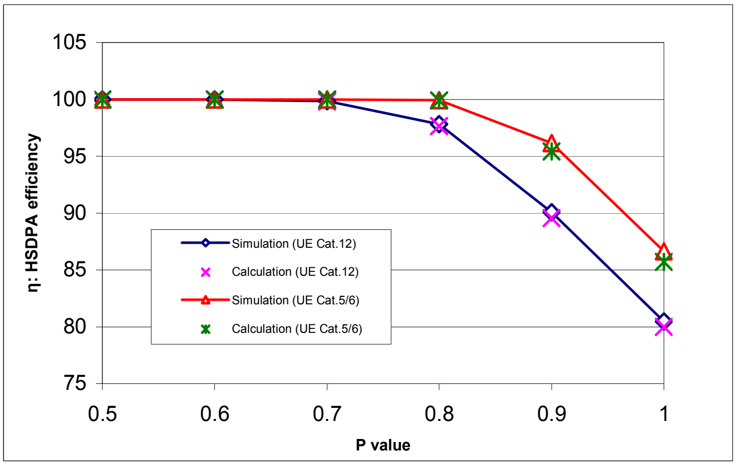


Figure 5.8 HSDPA efficiency for different values of *p*.

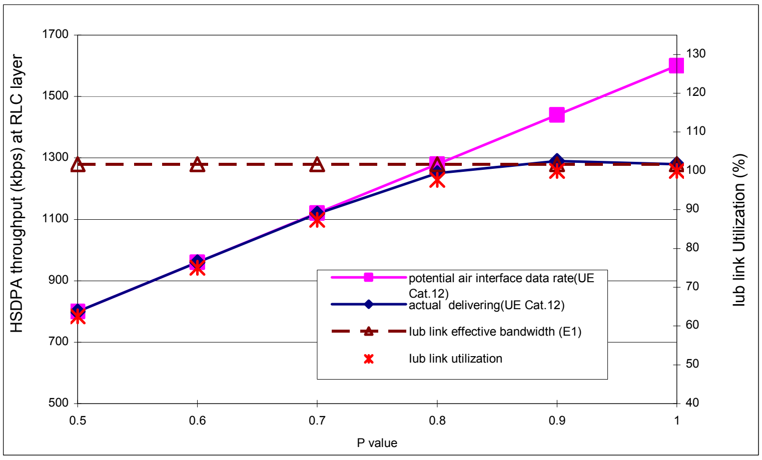


Figure 5.9 Relationship between the HSDPA average throughput (RLC layer) and I_{ub} link utilization

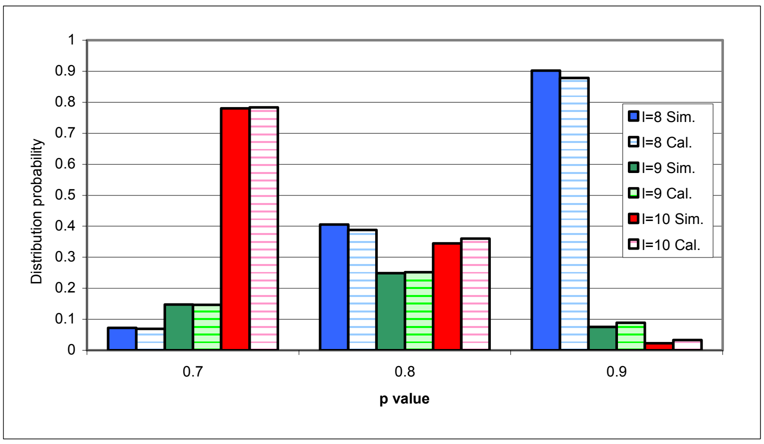


Figure 5.10 Distribution number of PDUs in the Node-B buffer for UE Cat.12

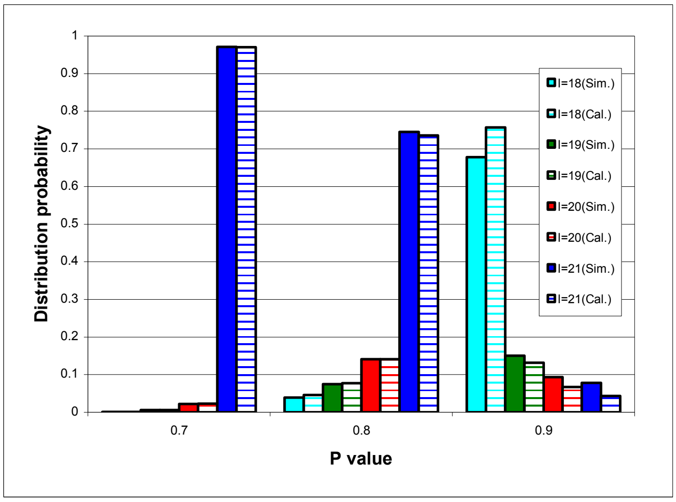


Figure 5.11 Distribution of PDU number in node-B buffer for UE Cat.5/6 (calculation vs. simulation)

5.4.2. Effect of Transport Protocols on the HSDPA Connections

Three transport protocols based I_{ub} links are used in the simulation model to transmit HS-DSCH FP (High Speed Dedicated Shared Channel Frame Protocol) frames between the RNC and the Node-B. The FP frame to transport frame mapping procedures have been described in the section 4.3.

The HSDPA transmission efficiency of AAL2/ATM, IP and Carrier Ethernet based I_{ub} link are compared. AAL2/ATM, IP and Carrier Ethernet protocols are separately used as the transport layer protocol in various simulation runs. Only Category 12 UEs are used in these simulations. We used the IPv4 protocol for IP based transport layer. To improve the transmission efficiency we also implemented UDP/IP header compression technique. Figure 5.12 shows the proportion of header bits used by the transport protocol to deliver the MAC-d PDUs in FP frame. These header bits are necessary control bits but consume a significant proportion of the I_{ub} link bandwidth. As the figure shows that for high bandwidth connections ATM based I_{ub} links consumes about 25% of bandwidth to transmit header bits thus reducing the effective I_{ub} link bandwidth to about 1.5 Mbps from 2.048 Mbps. On the other hand IP and Ethernet based I_{ub} links require relatively fewer header bits thus offering high effective link bandwidth. Figure 5.13 shows the HSDPA efficiency figure for different transport protocols. Result shows that the CIP solution offers the best HSDPA efficiency figures for both average and good channel conditions. Results also show that the increased I_{ub} link data rate improves the value of η for different transmission channel conditions. Use of IP protocol on the I_{ub} link will reduce the latency of a RAN and will improve overall QoS of a HSDPA network.

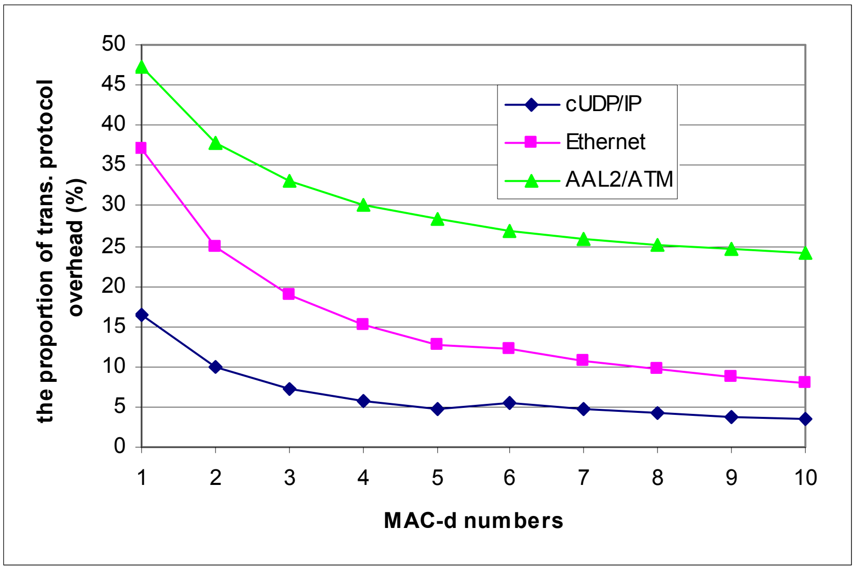


Figure 5.12 Efficiency comparison of transport layer protocols

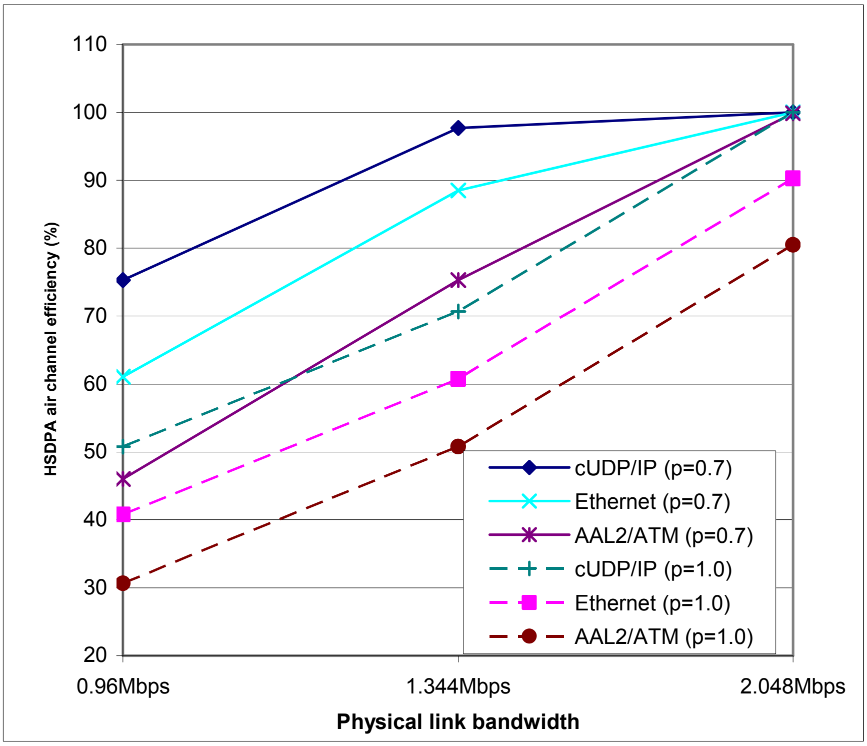


Figure 5.13 HSDPA air interface efficiency for different I_{ub} physical link data rate

5.5 Summary

In this chapter, novel analytical and simulation techniques have been developed to study the relationships between the HSDPA air interface and the I_{ub} link characteristics. For next generation HSDPA network services it is necessary to properly dimension the radio access network to support high data rate multimedia services over the air interface. However, the I_{ub} link congestion can arise if the I_{ub} link is not properly designed. This is a real issue that is gaining prominence with ever increasing air interface data rates. A HSDPA air interface throughput can increase with the improvement of air interface transmission conditions. However, due to the I_{ub} link constraints the air interface throughput could remain limited unless the I_{ub} link is properly designed. The HSDPA transmission efficiency increases with the increasing I_{ub} link effective bandwidth. On the other hand, if a high bandwidth I_{ub} link is used then for average transmission channel conditions the I_{ub} link utilization could drop and therefore reducing the network resource utilisation. So, simply increasing the I_{ub} link capacity is not sufficient to increase the overall HSDPA link efficiency. Transport layer protocols also play an important role on HSDPA performance. When ATM and IP transport protocols are compared it can be seen that the IP based transport protocol provide significant improvement of the air interface efficiency for a given I_{ub} link bandwidth.

Chapter 6

Effect of Scheduling and Adaptive Joint Resource Allocation Technique for the HSDPA RAN

6.1 Introduction

Due to the time-varying nature of the radio channel, it is difficult to specify a capacity of a shared channel that could depend on which users are being served in a particular time [70]. The packet scheduling algorithm at the Node-B and the I_{ub} link characteristics could affect the HSDPA air interface performance. Hence, it is one of the major objectives of this chapter to study the impact of I_{ub} link capacity on the HSDPA air interface for different type of Node-B schedulers. This chapter also introduces a joint resource allocation technique for the UTRAN to support the HSDPA traffic. Chapter 5 analysed the performance of the I_{ub} link and investigated the dependency of the I_{ub} link performance on the air interface characteristics. Results showed that the radio channel characteristics and transport layer protocol can significantly affect the performance of the I_{ub} link and hence, the performance of the UTRAN. In this chapter we have developed a UTRAN design paradigm where a joint resource allocation technique is used to allocate the transmission resources of the I_{ub} link to maximise the I_{ub} link utilization as well as minimising the transmission resource requirements of the I_{ub} interface. The joint resource allocation algorithm also encompasses the packet scheduling algorithms to optimise the performance of the UTRAN.

High HSDPA traffic intensity could increase the risk of congestions in a RAN leading to increased packet delays and higher probability of packet loss, which in turn could trigger a large number of RLC (Radio Link Control) layer retransmissions consequently degrading the HSDPA network performance [71]. Bajzik et. al. [71] introduced a cross-layer backpressure algorithm to prevent congestions and to reduce the RLC layer retransmissions. Using the algorithm the RNC and the Node-B are connected via a Constant Bit Rate (CBR) AAL2/ATM virtual circuit connection (VCC). All connections are multiplexed onto the VCC by using a strict priority AAL2 multiplexer. The HSDPA MAC-d flows are mapped on to a low priority connection and the remaining real time or non-real time connections are directed to a high priority buffer. The MAC-d layer schedules the PDUs according to the Node-B allocated capacity. The algorithm controls the load of the low priority AAL2 buffer to prevent the delay of the AAL2 layer not to exceed the maximum allowed delay. The queue length in the low priority AAL2 buffer is measured periodically. The backpressure algorithm sends a control command to the MAC-d layer to decrease the sending rate of the MAC-d flows when the buffer length threshold crossing is detected at the low priority buffer. This algorithm avoids the traffic congestion on the I_{ub} link by controlling the traffic flow. However, using the above algorithm the Node-B probably may not get enough packets to sustain the transmission rate on the HSDPA air interface and unable to fully exploit the air interface resources. The 3GPP document contained in the reference [38] presented a HSDPA congestion control technique to improve the traffic flow on the I_{ub} link. As described in section 3.6, Chapter 3, the suggested congestion control technique can't act as a flow control to pro-actively prevent an imminent congestion rather used as an "emergency break" in order to keep the system at a stable state when any congestion is detected. Legg [37] presented an I_{ub} link flow control strategy. The Node-B allocates credits to the RNC equal to the memory allocation at the Node-B. The RNC can then send PDUs to the Node-B. The number of PDUs should not exceed the allocated credit limit. When the Node-B schedules *n* PDUs, it sends an HSDSCH CAPACITY ALLOCATION command to the RNC for *n* credit to top up the number of PDUs at the Node-B up to the memory allocation level. The I_{ub} link capacity constraint isn't considered in Legg's study. In this chapter, an adaptive RAN resource management algorithm is deployed to improve the RAN resource utilization and HSDPA air interface performance.

This chapter first introduces different key scheduling techniques used in a HSDPA network. Section 6.2 discusses various packet scheduling techniques and their impact on the I_{ub} link. Section 6.3 introduces the developed joint adaptive resource allocation technique for the I_{ub} interface. Both sections present simulation and analytical results to analyse various UTRAN related performance figures. Results presented in this chapter can be used as guidelines for an HSDPA network design. Summary of findings are presented in the section 6.4.

6.2 Packet Scheduling Technique

A packet scheduler is a key component of the HSDPA network. Different types of packet scheduling algorithms are currently being considered for mobile communication networks which allocate transmission resources based on transmission conditions, traffic classes and priorities, and connection requirements [33]. Air interface characteristics are influenced by several time and location dependent parameters such as signal attenuation, fading, interference, and noise that results in variable transmission channel capacities [72]. There are a number of packet schedulers which try to exploit the variation of instantaneous channel conditions to maximize the network throughput while satisfying users' QoS requirements. This particular approach may lead to some trade-off problems between radio resource allocation efficiency and levels of satisfaction among channel users [73]. There are other classes of schedulers which may not always tries to exploit the transmission channel characteristics but allocate resources based on traffic QoS requirements and fairness of resource allocations among users [33]. Selection of a packet scheduler could depend on a number of factors including HSDPA connection QoS requirements, traffic type and the RAN design. By employing different radio resource management (RRM) algorithms, a packet scheduler can be configured to provide users with different peak to average ratios of data rates over the air interface. An HSDPA network potential achievable cell throughput depends on the scheduler deployed in the Node-B. A cell throughput variation would lead to different demand for the I_{ub} link capacity to support different types of traffic. In the following the basic characteristics of a number of key packets scheduler and their influence on the I_{ub} link design are presented.

6.2.1. HSDPA Packet Scheduling Algorithms

In this section we examine the effect of three packet scheduler algorithms (RR, Max-C/I and PF) on the performance of the I_{ub} interface. The Round Robin (RR) is simple, channel-independent packet scheduling algorithm, which serves users in a cyclic order, ignores the channel quality conditions and ensures a fair resource distribution among all users. The RR algorithm also ignores the QoS requirements of traffic sources and hence, allocates transmission resource in a simple round robin fashion. Although the RR scheduling algorithm apparently offer the fairness by equally allocating the air interface transmission time but the actual throughput of different users will be determined by the radio transmission channel condition. In a good channel condition a RR scheduler could offer fairness to all users.

The maximum carrier-to-interference ratio (Max-C/I) serves the user with largest instantaneous supportable data rate to maximize a cell throughput. This algorithm is well suited for high data rate terminals with a good transmission channel. Using the Max-C/I scheduler a UE can transmit a peak rate. However the main disadvantage of this algorithm is the lack of fairness, the users with poor channel quality conditions will get less or no transmission opportunities. Similarly the users at cell edges may not receive any transmission opportunities or throughput.

The proportional fair (PF) provides a trade-off between fairness and achievable cell throughput [33], [73]. The PF scheduler selects the user with the largest relative channel quality:

$$P_i(n) = \frac{R_i(n)}{T_i(n)} \qquad (6.1)$$

Where $P_i(n)$ denotes the user i priority in the *n-th* TTI, $R_i(n)$ is the instantaneous data rate experienced by user i if it is served, and $T_i(n)$ is the user i average throughput in past TTIs. The classic method to average the user throughput is shown as equation (6.2).

$$T_i(n) = \frac{A_i(n)}{n} \qquad (6.2)$$

Where $A_i(n)$ represents the amount of successfully transmitted data by user i in the past. In [74] and [75], a exponential decay filter was used for the averaging process. The average throughput of the user i is calculated recursively using the following equation.

$$T_i(n) = (1 - \frac{1}{N_i})T_i(n-1) + \frac{1}{N_i}R_i(n) \qquad (6.3)$$

Where N_i is the time constant in TTI numbers, it represents the memory of the exponential decay filter. The average throughput is updated every TTI. If a user is not to be served then the $R_i(n)$ is set equal to zero, otherwise the $R_i(n)$ is equal to actual throughput in the *n-th* TTI.

If the user average throughput is calculated by equation (6.2) then the average throughput decrease when there are no data to transmit because of empty Node-B buffer. This reduction of the user throughput during the traffic inactivity periods artificially increases the priority of the user using the PF algorithm. It avoids the influence of burstiness of data sources for the priority computation to calculate the user average throughput using the equation (6.3). The average throughput only is updated when the user have data queued for transmission in the Node-B buffer [33].

In a HSDPA network each user may experience different channel conditions due to their locations in a cell. Each users channel quality can be estimated by measuring the QoS of a connection as well as by reading the CQI (Channel Quality Indicator) of a connection [32]. A scheduler can serve the user with favourable channel condition and avoid serving users that experience deep fades, due to the time shared nature of the HS-DSCH. A degree of selection (multi-user) diversity gain can be introduced, which subsequently benefit the system throughput [33]. The multiuser diversity gain increases with number of users.

The packet schedulers discussed in this section not only affect the throughput of the HSDPA air interface but also affect resource allocation of the UTRAN. As already discussed, a HSDPA connection utilises the Node-B buffer to support packet transmission on the downlink which is replenished by the RNC via the I_{ub} link. In order to maintain high throughput on the downlink the Node-B buffer should be replenished with a minimum delay using the flow control algorithm. In the previous chapter the effect of variable channel throughput on the HSDPA efficiency factor η have been discussed. The following section examines the effect of various packet schedulers on the I_{ub} link resource requirements and its transmission efficiency.

6.2.2. Impact of Node-B scheduling algorithms on the I_{ub} Link Performance

In this section, the impact of Node-B scheduling algorithms on the I_{ub} link is analysed using the simulation model described in the chapter 5. In the simulation three scheduling algorithms RR, Max-CQI and PF are used separately at the Node-B to transmit packets. The radio channel condition represented by p of each UE is controlled by a uniform distribution where values could vary between 0.5 and 1.0. We assume low users mobility in order to allow the scheduler to be able to track the channel fading variations. The simulation model uses Category 12 users and a 2.048 Mbps I_{ub} link to connect the Node-B and the RNC.

Figure 6.1 shows the cumulative distribution function (CDF) of potential HSDPA air interface throughput and the actual delivered throughput using RR, Max-CQI and PF scheduling algorithms. The figure shows CDF of each HSDPA connection throughput in terms of number of MAC PDU's per TTI. The result was obtained by simulating 10 HSDPA connections. The figure shows that the slope of the Max-CQI curve is very steep where more than 90% of the connections transmit at the peak rate using the scheduling algorithm, since it always serves the user with largest instantaneous supportable data rate**.** The Max-CQI algorithm fully exploits the available capacity thus offering the highest potential HSDPA throughput for the connections with high CQI values. On the contrary, the slope of RR is less steep. The RR offers a low potential average throughput, because it ignores the radio channel quality and allocates transmission resources in a cyclic manner. In this case some of the connections which experience lower SNR values can transmit packets at a much lower data rate. The PF provides a trade-off between fairness and achievable cell throughput. In this case modest throughput gain is achieved compared to the RR. Due to the I_{ub} link capacity constraint the link couldn't deliver enough MAC-d PDUs from RNC to Node-B to sustain high rate air interface packet transmission, the actual delivered HSDPA throughputs are lower than potential ones for a good radio channel condition. The figure shows the actual throughput is lower than I_{ub} link allocated capacity (8 PDUs/TTI) for a saturated connection. Figure 6.2 shows the average cell throughput for 2, 6 and 10 HSDPA connections using RR, Max-CQI and PF scheduling algorithms. The figure shows that the potential cell throughput increases with increasing number of connections for Max-CQI and PF algorithms. For Max-CQI scheduler, the potential throughput increases from 8.8 to 9.94 PDUs/TTI. This increase is achieved due to the multi-user diversity gain. The multi-user diversity gain for the Max-CQI scheduler is highest because the user with the largest instantaneous supportable data rate is getting more opportunities to send its data. For the PF scheduler, the potential throughput increases from 8.4 to 9.3 PDUs/TTI. The throughput for the RR scheduler is constant for a varying number of users. There is no multi-user diversity gain for the RR algorithm since it serves users in a cyclic order and ignores the channel quality. The I_{ub} link constrains the actual average cell throughput limiting the capacity to 8 PDUs/TTI. Figure 6.3 shows the HSDPA efficiency η for different scheduling algorithms and user numbers. Results indicate the impact of the I_{ub} link capacity constraint is quite significant when the user number increases to 10 and the Max-CQI scheduler is used. A Max-CQI scheduler fully exploits the available air interface capacity and high multi-user diversity gain. The HSDPA air interface efficiency η drops from 94.38% to 80.56% with the increasing user numbers. In this case the HSDPA air interface efficiency value reaches 80% which means 20% of the air interface capacity can't be utilised due to the I_{ub} link bottleneck. Under this condition the I_{ub} link will not be able to update the Node-B in one TTI time which resulting fewer numbers of transmitted PDUs on the radio link. The PF algorithm may offer subsequent transmission opportunities at a much lower frequency whereas the RR scheduler doesn't offer any subsequent transmission opportunities. The above results illustrate that the I_{ub} link capacity constraint offsets the cell throughput gain by deploying an effective scheduling algorithm and the multi-user diversity. These results were obtained for a fixed 2 Mbps ATM/AAL based I_{ub} link. The value of η can be improved for the Max-C/I scheduler either by over dimensioning I_{ub} link capacity or by matching the air interface capacity with the I_{ub} link capacity.

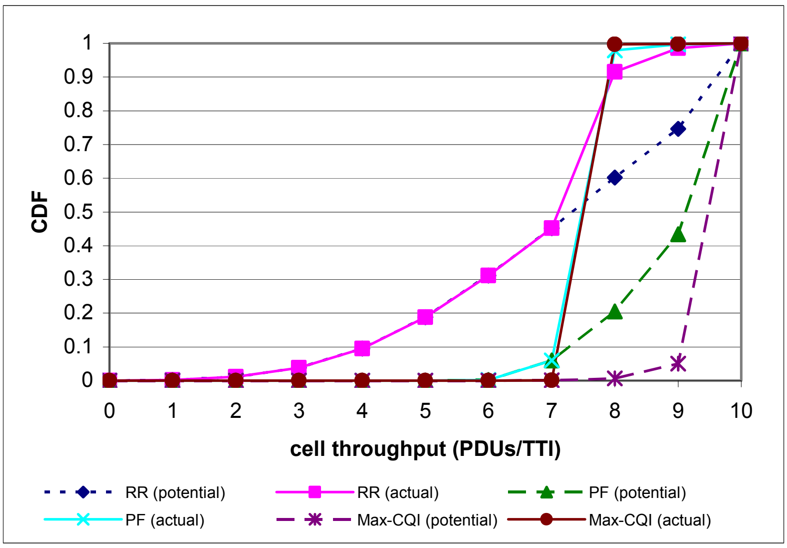


Figure 6.1 Potential and actual HSDPA cell throughput

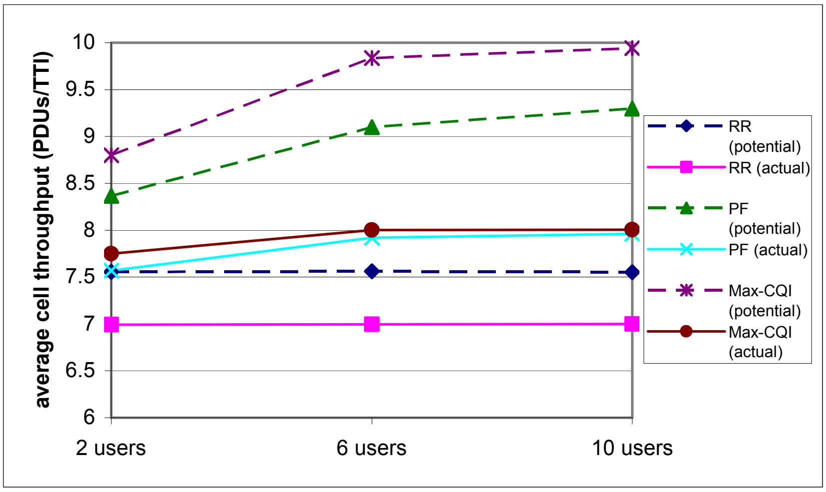


Figure 6.2 Average cell throughput; comparing potential and actual throughput

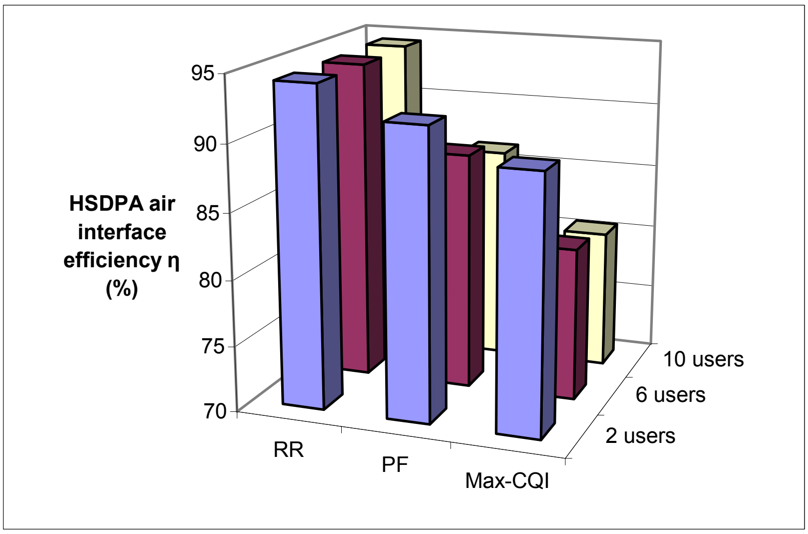


Figure 6.3 HSDPA efficiency distribution for different scheduling algorithms

In the next simulation high rate UEs are deployed. We use Category 10 UEs to replace Category 12 UEs in the simulation. The Category 10 UE can support up to 13.3 Mbps peak data rate at the RLC layer. Ten UEs were used in this simulation. The I_{ub} link data rate is increased from six E1 (12.288 Mbps) to eleven E1 (22.528 Mbps) in the step of one E1 (2.048 Mbps). All E1 links are grouped using the Inverse Multiplexing technique on the ATM (IMA) link [76]. The IMA protocol uses multiple E1 links as a group to provide an aggregate higher bandwidth logical link whose data rate is approximately the sum of individual link rates.

Figure 6.4 indicates actual average cell throughput is close to the potential throughput with the increasing the I_{ub} link capacity. In this case more packets can be delivered from the RNC to Node-B to meet the need of HSDPA air interface. The actual cell throughput increases with the I_{ub} link bandwidth constraint easing. When the I_{ub} link capacity is increased to the peak rate of UE, the efficiency of HSDPA air interface is up to 100 percent as shown in the Figure 6.5. The value of η increases with the increasing of I_{ub} link data rate. The rate of rise depends on the packet scheduling algorithm. As shown in the Figure 6.5 that the rate of rise of RR and PF schedulers are quite moderate compared to the Max-CQI scheduler. The Max-CQI scheduler offers much lower efficiency at a lower data rate but the efficiency increases at a faster rate than other schedulers. The reason for the lower efficiency at the lower I_{ub} link data rate is that in case of a Max-CQI scheduler the user with the largest instantaneous supportable air interface data rate get transmission opportunities in subsequent TTI's thus introducing high bandwidth requirement for the I_{ub} link which the link can't cope at a lower data rate. The RR scheduler serves users in a cyclic order, ignores the channel quality conditions, which introduces a moderate bandwidth requirement for the I_{ub} link.

The above simulation results show that the selection of a scheduler algorithm and the I_{ub} link capacity are very important factors which could significantly influence the RAN performance. These two factors have to be comprehensively considered when designing a RAN.

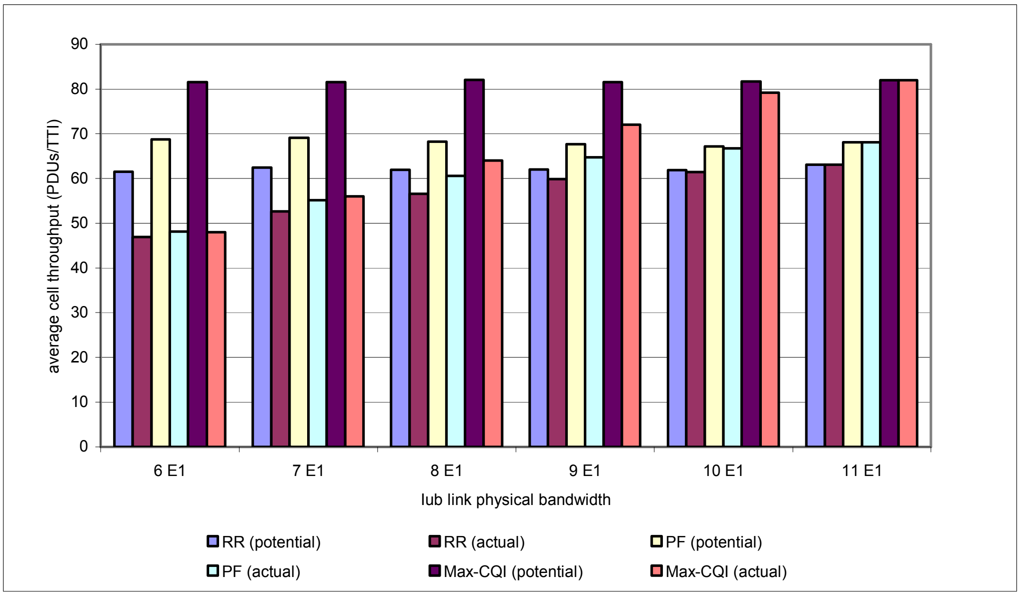


Figure 6.4 Average cell throughput (UE Category 10)

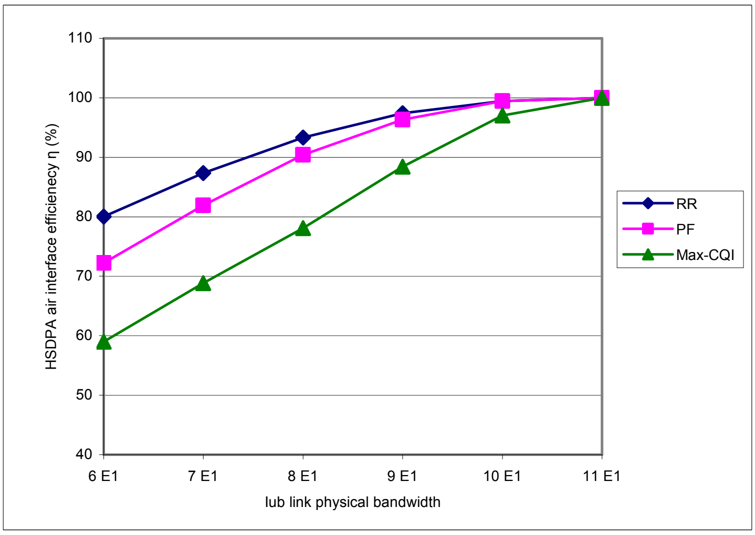


Figure 6.5 HSDPA air interface efficiency (UE category 10)

6.3 Joint Resource Management Technique

This section presents a dynamic bandwidth management technique to improve the transmission efficiency of a HSDPA network by adaptively allocating radio access network resources. In Chapter 5, Figure 5.7 shows that the HSDPA transmission efficiency for the category 12 UE for different air interface conditions and I_{ub} link capacities. The figure shows that the value of η increases with the I_{ub} link data rate. The figure also shows that the degree of improvement depends on the air interface condition. In a realistic HSDPA network the air interface throughput will vary depending on the transmission channel condition. Hence, a fixed I_{ub} bandwidth allocation will either reduce the air interface efficiency or it will reduce the I_{ub} link utilization. To avoid above problems we propose a simple but effective RAN resource allocation technique. The figure shows that for an HSDPA connection the HSDPA air interface efficiency could reach 97% when the I_{ub} link effective bandwidth is increased to its air interface average throughput. In the other word, if a HSDPA connection's air interface average throughput is known, and the I_{ub} link bandwidth can be allocated accordingly to the air interface condition to achieve 97% air interface efficiency. The resource allocation algorithms are compared from the I_{ub} link efficiency point of view in this section.

6.3.1. Proposed Joint Resource Management Technique

Based on above observations a simple adaptive joint resource allocation technique is proposed in this section. The proposed algorithm allocates the I_{ub} link capacity based on the air interface average throughput. The proposed resource management algorithm will be located at a Node-B which will measure every HSDPA connection throughput. Based on the air interface throughput the resource allocator at Node-B will allocate the corresponding capacity on the I_{ub} link in each TTI.

a) Estimating HSDPA air interface average throughput

The resource allocator estimates each HSDPA connection average throughput using a simple averaging equation. Equation (6.4) shows the averaging process. The connection throughput is measured for every data bursts which contain a number of packets. The running average of the air interface connection throughput is used to allocate capacity of that particular connection. The average throughput value is calculated for every data bursts separately i.e. when a connection goes to the idle mode then its average throughput value of the link is reset to zero.

$$\overline{throughput} = \frac{1}{N}\sum_{i=1}^{N} X_i \qquad (6.4)$$

Where X_i represents the throughput in number of MAC PDUs for the *i-th* transport opportunity of a connection and N represents the transport opportunity numbers.

When $X_1, X_2, ... X_i \sim binomial\ (n, p)$, Round Robin scheduling technique is used [78],

$$\begin{aligned}
&\mathrm{var}(\overline{throughput}) \\
&= \mathrm{var}(\frac{1}{N}\sum_{i=1}^{N} X_i) \\
&= (\frac{1}{N})^2 \sum_{i=1}^{N} \mathrm{var}(X_i) \\
&= (\frac{1}{N})^2 Nnp(1-p) \\
&= \frac{np(1-p)}{N} \\
&= \frac{np(1-p)}{N} \qquad (6.5)
\end{aligned}$$

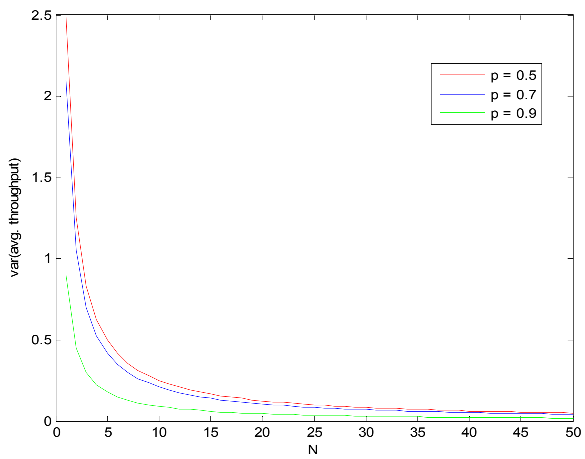


Figure 6.6 the variance of average throughput

Figure 6.6 shows the plot of the equation 6.5. Figure shows that with the increasing value of N the calculated value of a connection throughput converges.

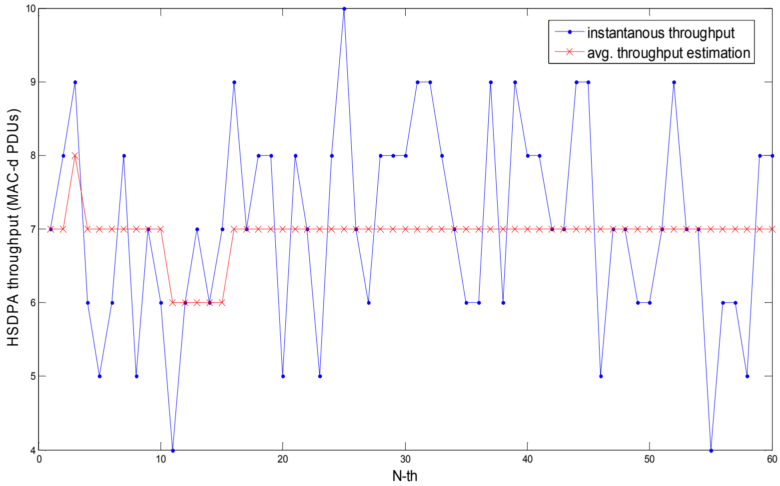


Figure 6.7 Instantaneous air interface throughput and the running average throughput value

Figure 6.7 shows the estimation process that the HSDPA connection throughput varies quite considerably over a 120 ms period (60 TTI). If the RNC to Node-B connection bandwidth is allocated based on the instantaneous throughput of the air interface then the air interface utilization as well as the I_{ub} link utilization could keep significantly high. It will be impossible to allocate I_{ub} resources based on the instantaneous throughput of the air interface because there is always a time lag in allocating the I_{ub} link resources based on the air interface condition. In previous chapter it was shown that if the I_{ub} link capacity is allocated based on the peak rate of the air interface connection then we may get high HSDPA efficiency but the I_{ub} link utilization drops significantly. Figure 6.7 shows that the link throughput is much lower than the peak capacity for the considerable amount of time. Based on this observation it will be prudent to allocate the I_{ub} link capacity based on the average HSDPA link throughput. The running average value of the HSDPA throughput then the both links throughput can be matched thus increasing the overall utilization.

b) Adaptive I_{ub} Link bandwidth allocation

To implement the adaptive bandwidth allocation technique we implement an I_{ub} link resource allocator as shown in Figure 6.8. The resource allocator measures each downlink connection buffers to generate the running average throughput and then using the value to allocate link capacity. The adaptive resource scheduler calculates each capacity $\overline{m}$ after every air interface transmission of the corresponding connection using the equation (6.4). The Node-B calculates the value of R_s by measuring the buffer length. The Node-B then sends a control message by incorporating the value of the I_{ub} link capacity and the request size R_s to the RNC. The I_{ub} link bandwidth allocated by the RNC is adjusted accord to the value of R_s. The required aggregate I_{ub} link bandwidth converges with the connection time. Using the above I_{ub} link bandwidth allocation technique it is possible to optimally match the air interface throughput rate and the I_{ub} link throughput. The required aggregate I_{ub} link bandwidth is calculated by equation (6.6).

$$Bandwidth = \sum_{i=1}^{N_{total}} s_i \times \overline{m_i} \qquad (6.6)$$

s_i represents the HSDPA connection i repetition rate, which depend on the scheduling technique used by the Node-B. N_{total} represents the total number of HSDPA connections mapped on the I_{ub} link.

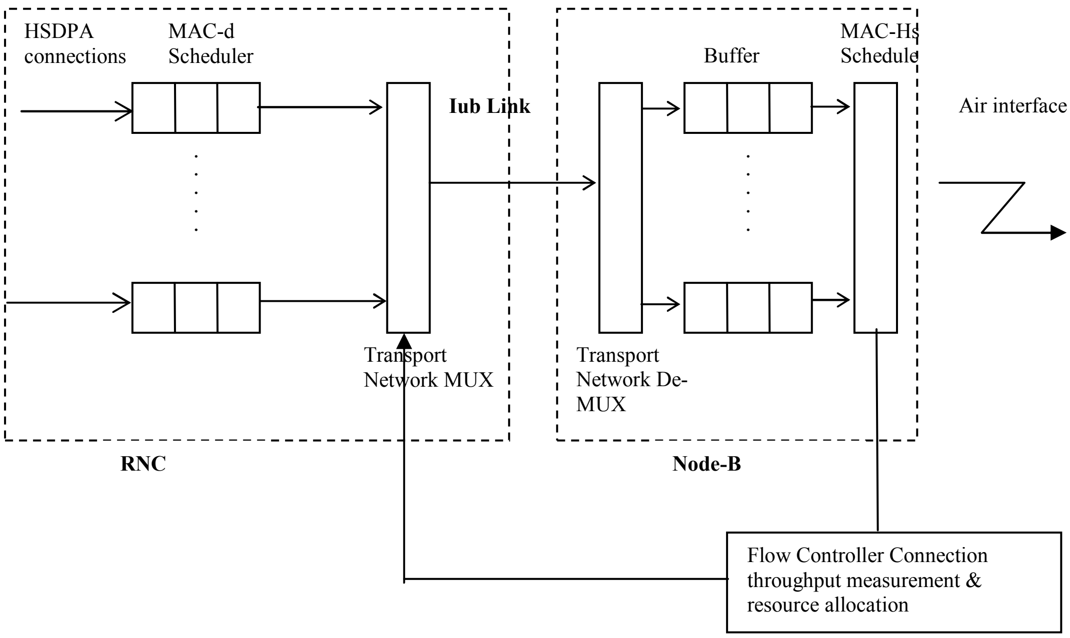


Figure 6.8 Adaptive bandwidth allocation

6.3.2. Performance Evaluation of the Proposed Algorithm

The simulation model described in Chapter 5 is modified according to the Figure 6.8 to include the adaptive bandwidth allocation algorithm. We obtain several simulation results using the modified model to analyze the performance of the adaptive resource management algorithm. Figure 6.9 compares the relationship between the transmission channel condition and the HSDPA air interface efficiency η for fixed and adaptive I_{ub} link bandwidth allocation techniques. The horizontal axis shows the average air interface throughput in terms of number of MAC PDUs/TTI. The figure shows that when the I_{ub} link capacity is allocated using the peak rate value of the air interface the value of η remains steady near 100% whereas the adaptive bandwidth allocation technique reduces efficiency values ranging from 96% to 100% depending on the channel condition. Figure 6.10 shows the I_{ub} link utilization for different air interface channel conditions. The graph shows that the adaptive bandwidth allocation technique significantly improves the I_{ub} link utilization (up to 50%) compared to the peak rate capacity allocation technique. It can be seen clearly from the plot the for an average channel condition the adaptive resource scheduler offer significant I_{ub} link capacity savings. It is possible to achieve high I_{ub} link utilisation when the adaptive resource allocation algorithm is used.

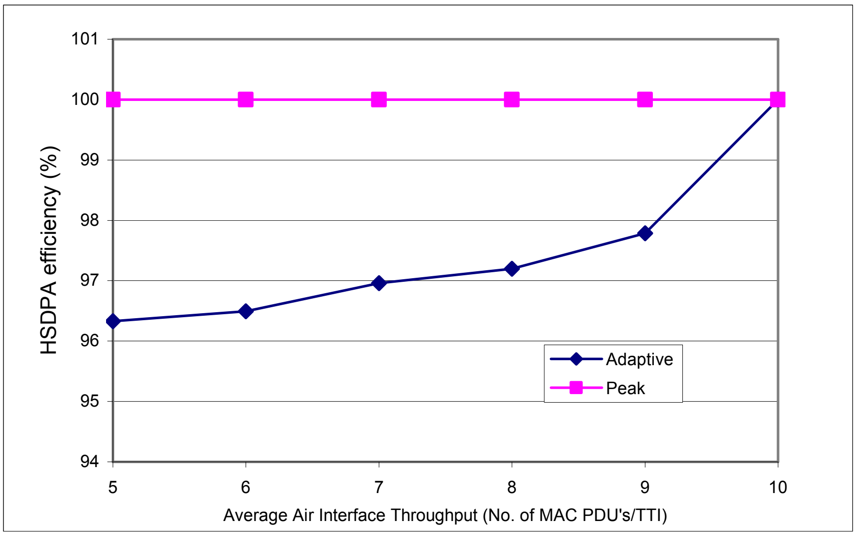


Figure 6.9 HSDPA air interface efficiency for different channel conditions and I_{ub} link allocation techniques

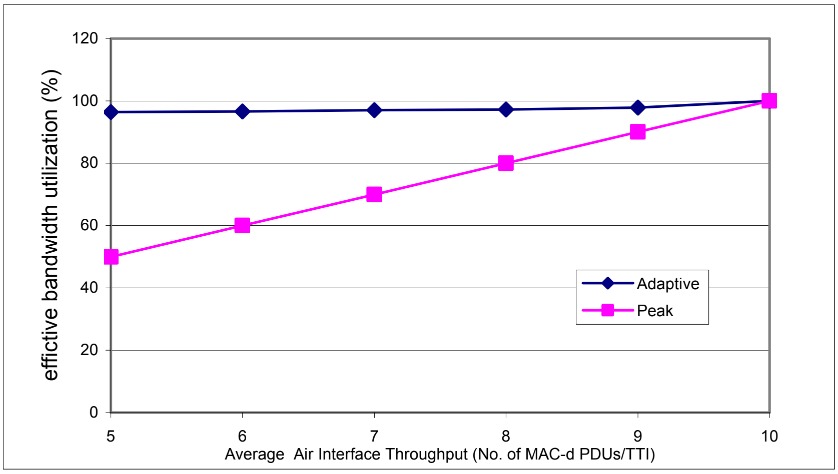


Figure 6.10 I_{ub} link effective bandwidth utilization for different channel conditions and link allocation techniques

Next, we investigate the aggregate I_{ub} link bandwidth requirements for different transport protocols. The performance of the adaptive resource allocation technique is analysed using AAL2/ATM, IP/Ethernet and UDP/IP transmission protocols. To obtain these simulation results we used total of ten UEs of Category 12. In the simulation the value of p (represents air interface transmission condition) of each UE is controlled by a uniform distribution where the p value could vary between 0.5 and 1.0. In the simulation the Round Robin (RR) scheduling technique is used at the Node-B. For this simulation first we allocate each I_{ub} link capacity using the peak rate value of its connection on the air interface. For the Category 12 UE, its peak transmission rate is 10 MAC-d PDU per TTI. In the next simulation run we use the proposed adaptive resource allocation technique. Figure 6.11 shows the HSDPA air interface efficiency and the I_{ub} link bandwidth utilization figures. Comparing these results we can see that the value of η drops by 2% when the adaptive resource allocation technique. Whereas when the I_{ub} link utilisation figure is compared the result shows that utilisation figure goes up by 25% and reaches nearly 100%. These results show the clear advantage of the feedback based joint resource allocation technique. Although there is a small price to pay which is the signalling required to obtain the estimate of the air interface throughput and to allocate the I_{ub} link capacity.

However, the signalling for the capacity allocation is combined with the data transmission process hence, the net signalling load is very low.

We further investigate the effect of the adaptive resource allocation technique on the Carrier Ethernet (IP/Ethernet) and UDP/IP transport protocol based I_{ub} links. Figure 6.12 presents the result for three transport protocols. Result shows that to support same amount of air interface traffic the I_{ub} link capacity requirement for the adaptive resource allocation decrease by 0.53 Mbps (from 2.33 Mbps to 1.8 Mbps) comparing to peak rate allocation when the AAL2/ATM transport protocol is used. The I_{ub} link data rate drop by 0.43 Mbps for the IP/Ethernet and UDP/IP protocols. Since the IP/Ethernet and UDP/Ethernet based links require relatively fewer header bits compared to an AAL2/ATM link hence, these links can offer further reduction of the I_{ub} link bandwidth. The graph shows that the IP/Ethernet and UDP/IP based I_{ub} link requires 0.3 and 0.4 Mbps lower transmission bandwidth respectively than an AAL2/ATM based. This is a significant savings of the I_{ub} link data rate and the savings becomes more prominent when the air interface is increased. Figure 6.13 shows the CDF of actual and potential cell throughput when AAL2/ATM transport protocol is deployed. Two curves are much closed that show the simple adaptive algorithm can be used to match the air interface throughput with the I_{ub} link data rate.

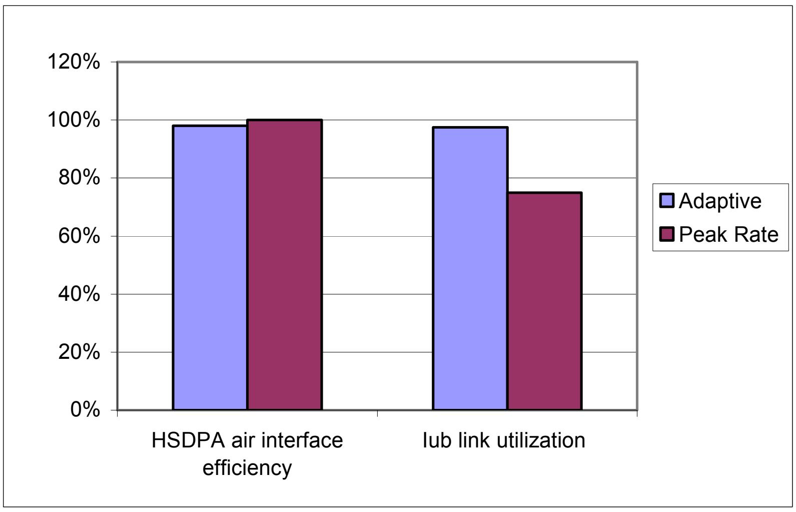


Figure 6.11 HSDPA air interface efficiency and I_{ub} link bandwidth utilization for the adaptive resource allocation vs. peak rate allocation techniques

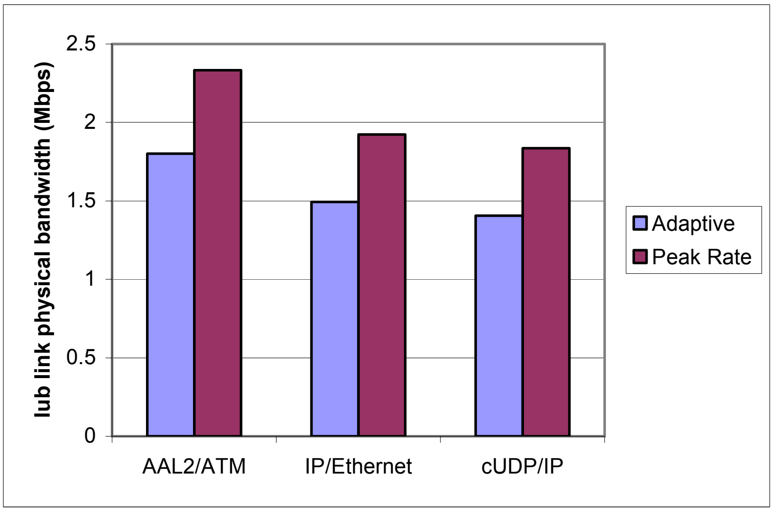


Figure 6.12 I_{ub} link bandwidth requirements for peak rate and adaptive allocation techniques using three transport protocols

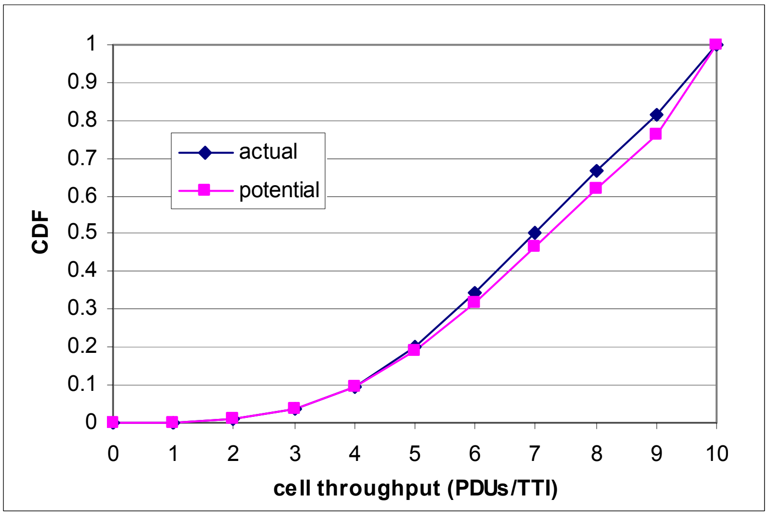


Figure 6.13 CDF of cell throughput (potential vs. actual)

Finally, we investigate the performance improvement for high data rate user equipment, UE Category 10, using the adaptive bandwidth allocation technique. In this simulation the performance of the adaptive resource allocation algorithm is evaluated for different channel conditions. To simulate different air interface channel conditions, the p value is varied for each group. The p value from 0.5 to 0.6 represents a poor air channel condition, in this channel condition very few opportunities will be available for HSDPA connections to transmit at a peak rate. The moderate channel condition is represented by the p values ranging from 0.5 to 0.8 and the good channel condition is represented by p values ranging from 0.5

to 1. Performance of the I_{ub} link is studied for different channel conditions for three transport protocols AAL2/ATM, IP/Ethernet and UDP/IP. Figure 6.14 shows the utilization of I_{ub} link bandwidth for different transmission conditions. With the adaptive algorithm, the I_{ub} link utilization reaches to 100 percent for all channel conditions. The utilization value for the adaptive resource allocation case is increased from 24% (in a good channel condition) to 44% (in a poor channel condition) comparing to the peak rate bandwidth allocation. This result indicates the I_{ub} link utilization improvement is significant especially for poor air channel conditions. Figure 6.15 illustrates HSDPA air interface efficiency which shows that the value of η is only 2% lower than the peak rate allocation scenario. Results of Figure 6.14 and Figure 6.15 prove the effectiveness of the proposed feedback based adaptive resource allocation technique. The I_{ub} link bandwidth constraint on air interface resource efficiency can be neglected. Figure 6.16 shows the I_{ub} link data rate for varying transmission channel conditions for different transport protocols. Compared to the peak rate bandwidth allocation the adaptive resource allocation technique reduces the transmission rate requirements of the I_{ub} link. The I_{ub} link physical transmission rate requirement decreases from 17.38 to 12.93Mbps (good air channel conditions), and to 9.75 Mbps (poor air channel conditions) when the AAL2/ATM transport protocol is used. For IP/Ethernet and UDP/IP transport protocols, the physical bandwidth requirement for peak rate allocation is 14.6 Mbps and 14.5Mbps respectively. For the adaptive allocation, the bandwidth requirements decrease 3.6 Mbps (good air channel conditions) and 6.4 Mbps (poor air channel conditions) for both of protocols. The above simulation results indicate the adaptive bandwidth management algorithm is very effective to improve the RAN resource utilization and reduce the operational cost of the UTRAN.

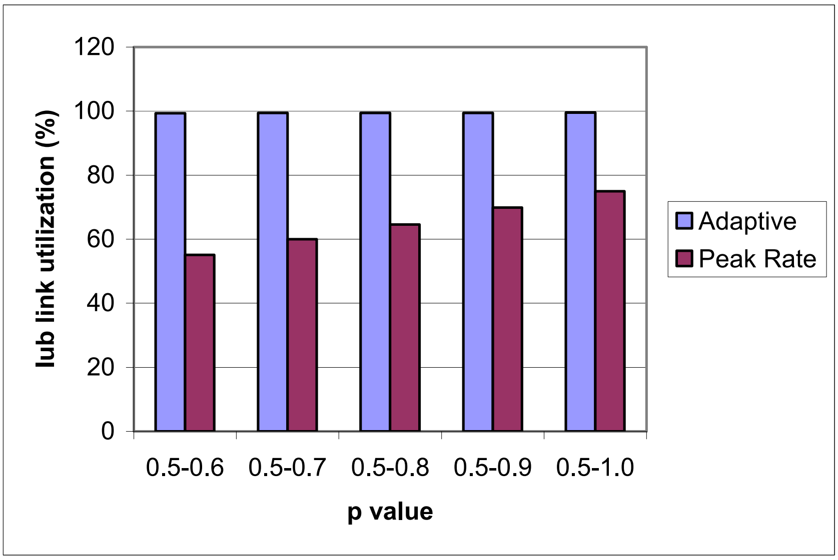


Figure 6.14 Iub link bandwidth utilization (UE category 10)

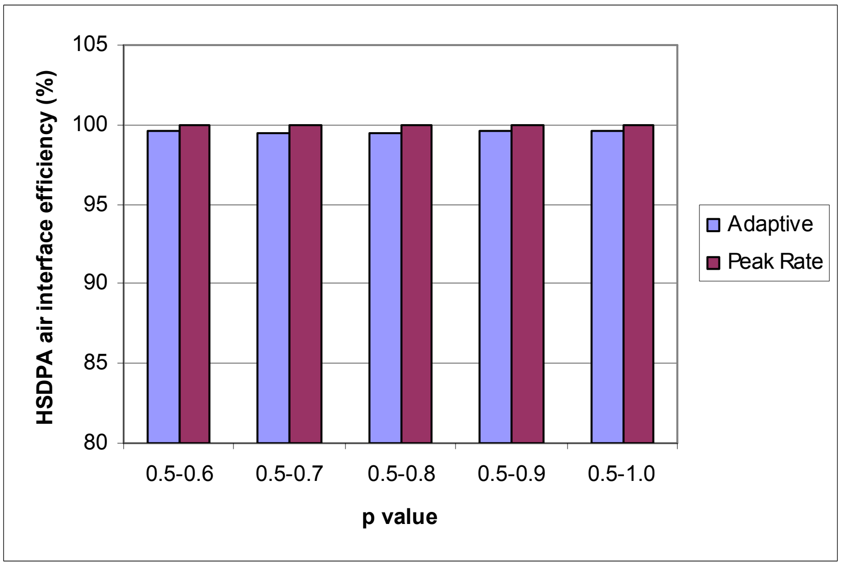


Figure 6.15 HSDPA air interface efficiency (UE category 10)

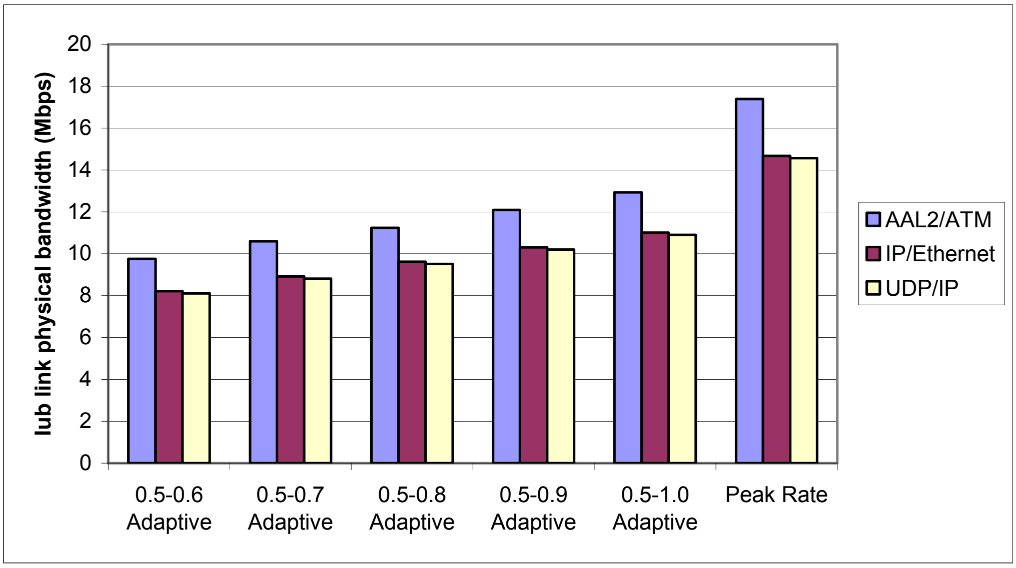


Figure 6.16 The required I_{ub} link physical bandwidth for three transport techniques (UE category 10)

6.4 Summary

In this chapter we studied the relationships between the HSDPA air interface and the I_{ub} link bandwidth allocation for different Node-B schedulers. The scheduler in the Node-B exploits the variation of instantaneous channel conditions to maximize the cell throughput while meeting users' QoS requirements. Multi-user diversity gain can be introduced by allocating channels to the users who are experiencing good transmission conditions. The Max-CQI fully exploits the available capacity thus offering the highest HSDPA air interface throughput and multi-user diversity gain. On the contrary, the RR offers a low throughput, because it serves user in a cyclic order and ignores the air interface channel quality. The PF provides a trade-off between fairness and achievable cell throughput. Modest throughput gain is achieved comparing to the RR. The scheduling algorithm used at a Node-B is also an important factor when allocating radio access network capacity. The impact of the I_{ub} link capacity constraint is quite significant when the Max-CQI scheduler is used because the Max-CQI offers the highest potential HSDPA air interface throughput, which needs higher I_{ub} link capacity to sustain the air interface rate. Selection of scheduling algorithms and the I_{ub} link capacity significantly influences the RAN performance.

By studying the interdependency between the downlink and the I_{ub} link characteristics we developed a simple adaptive I_{ub} link bandwidth allocation technique. We presented both analytical and simulation models to study the performance of both interfaces. Results show the adaptive algorithm clearly reduces the radio access network resource requirements while keeping the air interface efficiency and the I_{ub} link utilization high. From physical bandwidth requirement point of view, UDP/IP and Carrier Ethernet based I_{ub} links are more efficient than an AAL2/ATM based link. IP and Ethernet based I_{ub} links require relatively fewer header bits and can support variable packet sizes thus offers a high effective link bandwidth which improves the resource requirements of a RAN. Simulation and analytical results clearly shows that for high data rate HSDPA connections it is necessary to jointly optimise the air interface and radio access network radio resources.

For next generation HSDPA networks it is necessary to properly dimension radio access networks to support high data rate multimedia services over the air interface. Backhaul connections are evolving from E1/T1 circuit based transmission technology to an IP/Ethernet based packet technology. Dynamic resource allocation techniques could significantly improve a network utilization and performance because the algorithm tracks the rapid change in traffic conditions and allocate resources to maximize the network resource utilisation. The results can be used as a reference by the network designers.

Chapter 7

Conclusion and Further Works

7.1 Contributions and Conclusions: A review

The goal of the UMTS system is to provide users flexible delivery of multimedia service and other new services as the results of the evolution of telecommunications networks. An extensive variety of service with ever increasing demands for higher data rate and a wide diversity of QoS requirements bring the challenge to a UMTS system designer. The RAN contributes a major part cost for the UMTS system operation. To properly dimension a network capacity and to maximize the system resource utilization it is vital for mobile operators to sustain their business growth in a very competitive environment. It is very important to jointly optimize radio and network resource to maintain the end-to-end QoS for a range of traffic services. Investigating the performance of interdependency between air interface and the RAN, developing adaptive resource allocation algorithm to optimise the RAN performance for multimedia traffic are the main focus of this research work.

The ATM technique was firstly introduced in the 3GPP release 99 standard, for the UTRAN to deliver circuit switched and packet switched services. An IP based UTARN was introduced in the release 5 of the 3GPP standard as the first step towards an all IP service network. An Ethernet based RAN could be feasible due to Carrier Ethernet deployment for other fixed network applications. In this work comprehensive simulation models have been developed to investigate the performance of ATM, IP and Ethernet based UTRAN to deliver voice, data and video traffic in Chapter 4. Results show that IP and Ethernet based links perform better under heavy traffic load conditions. For the HSDPA traffic, average delay of an Ethernet based link is lower than the ATM based link. The Ethernet based link is an alternative transport technique for the UTRAN I_{ub} link due to its flexibility, economy and bandwidth efficiency. The effect of priority transmission in a RAN is also studied in Chapter 4. When the retransmission data traffic is offered appropriate priority then the end-to-end delay can be significantly reduced thus improving the QoS of data traffic. The Timer-CU is used to assure that packets in the AAL2 multiplexing queue don't wait longer before being scheduled for transmission in the AAL2 multiplexer. The Timer-CU value has an important impact on the performance of I_{ub} interface for a low traffic load condition. A low Timer-CU could lead to decreasing delay caused by packets waiting in AAL2 multiplexing queue. With the increasing traffic load condition the importance of the timer-CU value becomes less important. However, very low value Timer-Cu values could lead to rapid increase of total end-to-end delay since it still leads to more padding in multiplexing stage in a high traffic load situation. This phenomenon directly increases the peak cell rate (PCR), which introduces a higher delay.

Chapter 5 presents novel analytical and simulation techniques to study the relationships between the HSDPA air interface and the I_{ub} link characteristics. For next generation HSDPA services it is necessary to properly dimension radio access networks to support high data rate multimedia services over the air interface. However, the I_{ub} link congestion is an important issue that is gaining prominence with ever increasing air interface data rates. The HSDPA air interface throughput can linearly increase with the improvement of air interface radio transmission conditions. However, due to the I_{ub} link constraints, the air interface throughput remains limited. The HSDPA transmission efficiency increases with the increasing I_{ub} effective transmission rate. But on the other hand, if a high data rate I_{ub} link is used for an average transmission channel condition then the I_{ub} link utilization could drop significantly. So, simply increasing the I_{ub} link capacity is not sufficient to increase the overall HSDPA link efficiency. A flow controller is employed between the RNC and the Node-B to control traffic on the link to prevent any Node-B buffer overflow or any buffer starvation. The RNC sends a capacity request to the Node-B to request the HS-DSCH capacity. The Node-B allocates capacity to the RNC, indicating the maximum MAC-d PDU length, numbers, HS-DSCH interval and repetition etc. to control the user data flow. The risk of congestions on I_{ub} link will increase when capacity allocated by the Node-B is higher than I_{ub} link capacity. The congestion leads to packets delays increasing and higher probability of packet loss, which in turn could cause the RLC layer retransmission, and then degrade the HSDPA performance. A flow control algorithm which allocates transmission resources on the I_{ub} link to pro-actively avoid congestions on the transport network is deployed between these two nodes to ensure a lower transmission delay between the RNC and UEs on the downlink. A Markov model is developed to quantify the interdependence of HSDPA air interface and RAN I_{ub} link capacity. The RAN performance is examined in terms of air interface transmission conditions and the transport network capacity. A simulation model is created to investigate the interdependence and to validate the analytical model. These two models can be used to plan the RAN capacity of a HSDPA network. This particular study of interdependency between a RAN and the HSDPA air interface is a unique and a novel approach which was not considered before.

In Chapter 6, the relationships between the HSDPA air interface and the I_{ub} link characteristic is further studied for different Node-B packet schedulers. A packet scheduling algorithm used at a Node-B is also an important factor when allocating radio access network capacity, which directly impacts the network resource utilization and the QoS. There is no multi-user diversity gain when the RR algorithm is used since it serves users in a cyclic order and ignores the channel quality. The PF provides a trade-off between fairness and achievable cell throughput. The impact of the I_{ub} link capacity constraint is quite significant when the Max-CQI scheduler is used because the Max-CQI fully exploits the available capacity thus offering the highest potential HSDPA throughput, which led to increasing I_{ub} link capacity requirement. By studying the interdependency between the downlink and the I_{ub} link characteristics this work has developed an adaptive I_{ub} link bandwidth allocation algorithm to optimise the RAN performance. The resource allocator of the flow controller measures the air interface capacity that could be supported by each downlink connection and generates the running average of the air interface throughput to allocate the I_{ub} link capacity. We presented both analytical and simulation models to study the performance of the HSDPA air interface and the I_{ub} interface. Results show that the adaptive algorithm clearly reduces the radio access network resource requirements while keeping the air interface throughput high. Comparing the adaptive bandwidth allocation algorithm with the peak rate allocation of the I_{ub} link capacity it can be found that the adaptive allocation technique significantly improves the I_{ub} link utilization and reduces the RAN cost due to lower bandwidth requirements. Simulation and analytical results also clearly shows that for high data rate HSDPA connections it is necessary to jointly optimise the air interface and radio access network radio resources.

The research results can be a useful reference for network designers and planers.

7.2 Further Works

In the 3GPP release 6, High Speed Uplink Access (HSUPA) is standardised as the WCDMA uplink evolution technology. HSDPA uses an uplink Enhanced Dedicated Channel (E-DCH), which has some of features similar to the HSDPA, such as shorter 2-ms transmission time interval (TTI), Hybrid ARQ with incremental redundancy, etc. Similarly to the HSDPA, a Node-B based scheduler operates on a request-grant principle where the User Equipment (UE) requests the permission to sent packets and the scheduler determines transmit time and rate. Unlike the HSDPA, the HSUPA support the soft handover. The HSUPA standard supports peak data rate up to 5.76 Mbps. HSUPA and HSDPA standards are complimentary to one another. In this work, we only investigated the performance interdependency between the HSDPA air interface and the RAN without considering impact of the HSUPA standard. The HSUPA standard will affect the RAN in a different way. The HSUPA will operate at a lower data rate than the HSDPA also at the same the volatility of the radio transmission condition can be absorbed by the Node-B buffer. In the further work, the performance impact of HSDPA and HSUPA interfaces on the RAN could be jointly investigated. RAN resource could be again jointly optimised to improve both air interfaces performance by using predictive models based on the realtime measurements from the network.

Another relevant topic that requires further investigation is the RAN resource management for the E-UTRAN. By providing a combination of very high downlink/uplink speeds and much more efficient use of spectrum, the LTE standard could revolutionise the way in which existing services are supported, as well as enabling the delivery of a range of high-value services and applications such as high quality video streaming, super-fast browsing, and uploading content to social networking site, high quality video meetings, etc. Looking forward, the range of services that could be supported by the LTE downlink/uplink data rates the mobile user QoS experience could reach to same level as fixed broadband networks [79]. In a recently published 3GPP technical report the E-UTRAN architecture move from the current structure to a flat RAN architecture [63], [80]. The LTE architecture is deviating away from the conventional hierarchical radio access and core network architecture. Proposed 3G LTE architecture is consists of an access core gateway (ACGW) and a Node-B. The proposed ACGW merges with the 3GPP release 6 nodes such as gateway GPRS support node (GGSN), serving GPRS support node (SGSN) and RNC into a single entity. Most of the MAC layer functionalities will be relocated in the Node-B. The packet-centric link layer will be introduced in this structure for easy transmission of packets [81]. According to the packet size of each application, the RLC block size can be flexibly chosen to avoid unnecessary padding or segmentation [82]. IP packets will be mapped one-to-one on RLC PDUs. The architecture will support variable PDU sizes. Segmentation and concatenation will be unnecessary at the RLC layer, thereby avoiding padding operations. Each RLC PDU will correspond to exactly one IP packet. The scheduler in MAC layer can be expected to allow for more efficient scheduling decision because complete IP packets will be seen by the MAC layer. Hence, the FP frame layer is not required anymore. To meet LTE performance requirements, the RAN resource management such as scheduling, flow control and transport capability allocation between ACGW and Node-B etc., is taking vital roles for further research. Considering these new changes in the network architecture it will be useful to develop jointly optimised resource allocation algorithm for the packet based LTE access network similar the approach presented in this work.

References

[1] UMTS Forum http://www.umts-forum.org/.

[2] "3G/UMTS Towards mobile broadband and personal Internet," UMTS Forum White Paper, 2005.

[3] H. Holma and A. Toskala, *WCDMA for UMTS (Third Edition)*: John Wiley & Sons, 2004.

[4] "3G/UMTS Evolution: towards a new generation of broadband mobile services," UMTS Forum White Paper, 2006.

[5] 3GPP, http://www.3gpp.org/, 2007.

[6] H. Holma and A. Toskala, "HSDPA/HSUPA for UMTS," John Wiley and Sons, 2006.

[7] "Cost-Optimized Transport Evolution - making the right choice," Nokia Corporation White Paper, 2006.

[8] P. O. Andersson, H. Asp, A. Bolle, H. Leino, P. Seybolt, and R. Swardh, "GSM transport evolution," *Ericsson Review*, vol. No.1, pp. 26-31, 2007.

[9] R. Makke, S. Tohme, J.-Y. Cochennec, and S. Pautonnier, "Performance of the AAL2 protocol within the UTRAN," presented at Universal Multiservice Networks, 2002. ECUMN 2002. 2nd European Conference on, 2002.

[10] 3GPP, "IP transport in UTRAN (Release 5)," TR 25.933 V5.4.0, 2003.

[11] H. Kaaranen, A. Ahtiainen, L. Laitinen, S. Naghian, and V. Niemi, *UMTS Networks: Architecture, Mobility and Service (Second Edition)*: John Wiley & Sons, 2005.

[12] 3GPP, "Network architecture (Release 7)," TS 23.002 V7.1.0, 2006.

[13] J. De Vriendt, P. Laine, C. Lerouge, and X. X. A.-X. Xu, "Mobile network evolution: a revolution on the move," *Communications Magazine, IEEE*, vol. 40, pp. 104-111, 2002.

[14] A. Korhonen, Introduction to 3G Mobile Communications: Artech House, 2001.

[15] C. Smith and D. Collins, *3G Wireless Network*: McGraw-Hill, 2002.

[16] V. K. Garg, *Wireless Network Evolution 2G to 3G*: Prentice Hall PTR, 2002.

[17] 3GPP, "UTRAN overall description (Release 7)," TS 25.401 V7.3.0, 2007.

[18] A. García, M. Alvarez-Campana, E. Vázquez, and J. Berrocal, "Quality of Service support in the UMTS Terrestrial Radio Access Network," presented at Proceedings of the 9th HP Openview University Association Workshop (HPOVUA 2002), 2002.

[19] D. Wisely, P. Eardley, and L. Burness, *IP for 3G*: John Wiley & Sons, 2002.

[20] K. Venken, I. G. Vinagre, and J. De Vriendt, "Analysis of the evolution to an IP-based UMTS terrestrial radio access network," *Wireless Communications, IEEE [see also IEEE Personal Communications]*, vol. 10, pp. 46-53, 2003.

[21] 3GPP, "Radio Link Control (RLC) protocol specification (Release 6)," TS 25.322 V6.5.0, 2005.

[22] 3GPP, "Radio Interface Protocol Architecture (Release 7)," TS 25.301 V7.1.0, 2007.

[23] L. Zhang, F. Li, and J. Zhu, "Performance analysis of multiple reject ARQ theme at RLC layer in 3G," presented at Vehicular Technology Conference, 2002. Proceedings. VTC 2002-Fall. 2002 IEEE 56th, 2002.

[24] 3GPP, "Medium Access Control (MAC) protocol specification (Release 6)," TS 25.321 V6.7.0, 2005.

[25] 3GPP, "Physical channels and mapping of transport channels onto physical channels (FDD) (Release 6)," TS 25.211 V6.3.0, 2004.

[26] 3GPP, "Services and Service Capabilities (Release 8)," TS 22.105, V.8.4.0, 2007.

[27] R. Koodli and M. Puuskari, "Supporting packet-data QoS in next generation cellular networks," *Communications Magazine, IEEE*, vol. 39, pp. 180-188, 2001.

[28] M. Zivkovic and Y. Wang, "QoS attributes for packet switched services in 3rd generation mobile systems (UMTS)," presented at Telecommunications Quality of Services: The Business of Success, 2004. QoS 2004. IEE, 2004.

[29] 3GPP, "End-to-end Quality of Service (QoS) concept and Architecture (Release 6)," TS 23.207 V6.6.0, 2005.

[30] J. Laiho, A. Wacker, and Novosad, *Radio Netork Planning and Optimization for UMTS*: John Wiley & Sons (UK), 2002.

[31] 3GPP, "Quality of Service (QoS) concept and architecture (Release 6)," TS 23.107 V6.4.0, 2006.

[32] S. Parkvall, E. Englund, M. Lundevall, and J. Torsner, "Evolving 3G mobile systems: broadband and broadcast services in WCDMA," *Communications Magazine, IEEE*, vol. 44, pp. 30-36, 2006.

[33] P. Jose, "Packet Scheduling And Quality of Service in HSDPA," in *Department of Communication Technology*: Aalborg University, 2003.

[34] 3GPP, "High Speed Downlink Packet Access (HSDPA), Overall Description, Stage2," TS 25.308 V6.3.0, 2004.

[35] 3GPP, "UTRAN Iub Interface User Plane Protocols for Common Transport Channel Data Streams," TS 25.435 V6.3.0, 2005.

[36] S. Y. Yerima and K. Al-Begain, "An Enhanced Buffer Management Scheme for Multimedia Traffic in HSDPA," presented at Next Generation Mobile Applications, Services and Technologies, 2007. NGMAST '07., 2007.

[37] P. J. Legg, "Optimised Iub flow control for UMTS HSDPA," presented at Vehicular Technology Conference, 2005. VTC 2005-Spring. 2005 IEEE 61st, 2005.

[38] 3GPP, "Iub/Iur congestion control (Release 7)," TR 25.902 V7.0.0, 2006.

[39] B. Subbiah, "Transport architecture evolution in UMTS/IMT-2000 cellular networks," presented at Global Telecommunications Conference, 2000. GLOBECOM '00. IEEE, 2000.

[40] D. Allan, N. Bragg, A. McGuire, and A. Reid, "Ethernet as carrier transport infrastructure," *Communications Magazine, IEEE*, vol. 44, pp. 95-101, 2006.

[41] "WiMax - copper in the air," Ericsson AB 2006.

[42] O. Isnard, J.-M. Calmel, A.-L. Beylot, and G. Pujolle, "Handling traffic classes at AAL2/ATM layer over the logical interfaces of the UMTS terrestrial radio access network," presented at Personal, Indoor and Mobile Radio Communications, 2000. PIMRC 2000. The 11th IEEE International Symposium on, 2000.

[43] J. W. Chong, J. H. Chung, C. Y. Jung, H. Y. Hwang, D. K. Sung, S. Jung, and J. S. Park, "QoS-Based AAL2/ATM multiplexing schemes in the UTRAN lub interface," presented at Personal, Indoor and Mobile Radio Communications, 2003. PIMRC 2003. 14th IEEE Proceedings on, 2003.

[44] MIWF, "IP in the RAN as a Transport Option in 3rd Generation Mobile System," Technical Report MTR-006, Rel: V2.0.0, June 18 2001.

[45] C. Y. Jung, J. W. Chong, H. Y. Hwang, D. K. Sung, and J. S. Park, "Performance comparison of ATM and IP based transmission schemes in the UTRAN," presented at Wireless Communications and Networking Conference, 2004. WCNC. 2004 IEEE, 2004.

[46] 3GPP, "UTRAN Iub/Iur interface user plane protocol for DCH data streams (Release 6)," TS 25.427 V6.1.0, 2004.

[47] G. Clark and Y. K. Ling, "Transport solutions for 3G cellular radio access network," presented at 3G Mobile Communication Technologies, 2002. Third International Conference on (Conf. Publ. No. 489), 2002.

[48] ITU-T, "Segmentation and Reassembly Service Specific Convergence Sublayer for the AAL type 2," I.366.1, June 1998.

[49] ITU-T, "B-ISDN ATM Adaptation Layer specification: Type 2 AAL," I.363.2, November, 2000.

[50] MEF, "User Network Interface (UNI) Requirements and Framework," November 2004.

[51] ITU-T, "GFP frame mapping into Plesiochronous Digital Hierarchy (PDH)," G.8040/Y.1340, September 2005.

[52] ETSI, "Universal Mobile Telecommunications System (UMTS); Selection procedures for the choice of radio transmission technologies of the UMTS," TR 101 112 V3.2.0, 1998.

[53] 3GPP, "Mandatory speech CODEC speech processing functions; AMR speech CODEC; General description (Release 6)," TS 26.071 V6.0.0, 2004.

[54] 3GPP, "Mandatory speech codec speech processing functions; Adaptive Multi-Rate (AMR) speech codec frame structure (Release 6)," TS 26.101 V6.0.0, 2004.

[55] 3GPP, "Transparent end-to-end Packet-switched Streaming Service (PSS); Protocols and codecs (Release 7)," TS 26.234 V7.3.0, 2007.

[56] 3GPP, "Transparent end-to-end Packet Switched Streaming Service (PSS); General description (Release 7)," TS 26.233 V7.0.0, 2007.

[57] *Network Dictionary*: Javvin Technologies, Inc., 2007.

[58] F. H. P. Fitzek and M. Reisslein, "MPEG-4 and H.263 video traces for network performance evaluation," *Network, IEEE*, vol. 15, pp. 40-54, 2001.

[59] P. Seeling, M. Reisslein, and B. Kulapala, "Network Performance Evaluation Using Frame Size and Quality Traces of Single-Layer and Two-Layer Video: A Tutorial," *IEEE Communications Surveys and Tutorials*, vol. 6, pp. 58-78, 2004.

[60] "E1 Physical Interface Specification," ATM Forum af-phy-0064.000, 1996.

[61] M. C. Necker and A. Weber, "Parameter Selection for HSDPA Iub Flow Control," presented at Wireless Communication Systems, 2005. 2nd International Symposium on, 2005.

[62] A. Toskala, H. Holma, E. Metsala, K. I. Pedersen, and D. Steele, "Iub efficiency analysis for high speed downlink packet access in WCDMA," presented at WPMC, Aalborg, Denmark, 2005.

[63] 3GPP, "Requirements for Evolved UTRA(E-UTRA) & Evolved UTRAN (E-UTRAN)," TR 25.913, V7.3.0, 2006.

[64] A. Toskala, H. Holma, K. Pajukoski, and E. Tiirola, "Utran Long Term Evolution in 3GPP," presented at Personal, Indoor and Mobile Radio Communications, 2006 IEEE 17th International Symposium on, 2006.

[65] C. Liu, S. Munir, R. Jain, and S. A.-D. Dixit, S., "Packing density of voice trunking using AAL2," presented at Global Telecommunications Conference, 1999. GLOBECOM '99, 1999.

[66] S. M. Ross, Introduction to Probability Models (Sixth Edition): Academic Press, 1997.

[67] Wikimedia: http://en.wikipedia.org/wiki/Binomial_distribution, 2007.

[68] E. Weisstein. http://mathworld.wolfram.com/BinomialDistribution.html: Wolfram MathWorld, 2007.

[69] J. Derksen, R. Jansen, M. Maijala, and E. Westerberg, "HSDPA performance and evolution," *Ericsson Review*, vol. No.3, pp. 117-120, 2006.

[70] M. C. Necker, "A comparison of scheduling mechanisms for service class differentiation in HSDPA networks," *AEU - International Journal of Electronics and Communications*, vol. 60, pp. 136-141, 2006.

[71] L. Bajzik, L. Korossy, K. Veijalainen, and C. Vulkan, "Cross-Layer Backpressure to Improve HSDPA Performance," presented at Personal, Indoor and Mobile Radio Communications, 2006 IEEE 17th International Symposium on, 2006.

[72] H. Fattah and C. Leung, "An overview of scheduling algorithms in wireless multimedia networks," *Wireless Communications, IEEE [see also IEEE Personal Communications]*, vol. 9, pp. 76-83, 2002.

[73] X. Liu, E. K. P. Chong, and N. B. Shroff, "Opportunistic transmission scheduling with resource-sharing constraints in wireless networks," *Selected Areas in Communications, IEEE Journal on*, vol. 19, pp. 2053-2064, 2001.

[74] M. Assaad and D. Zeghlache, "Opportunistic Scheduling for Streaming Services in HSDPA," presented at Personal, Indoor and Mobile Radio Communications, 2006 IEEE 17th International Symposium on, 2006.

[75] A. Jalali, R. Padovani, and R. Pankaj, "Data throughput of CDMA-HDR a high efficiency-high data rate personal communication wireless system," presented at Vehicular Technology Conference Proceedings, 2000. VTC 2000-Spring Tokyo. 2000 IEEE 51st, 2000.

[76] T. E. Kolding, "Link and system performance aspects of proportional fair scheduling in WCDMA/HSDPA," presented at Vehicular Technology Conference, 2003. VTC 2003-Fall. 2003 IEEE 58th, 2003.

[77] "Inverse Multiplexing for ATM (IMA) Specification Version 1.1," ATM Forum af-phy-0086-001, 1999.

[78] J. A. Gubner, Probability and Random Processes for Electrical and Computer Engineers: Cambridge University Press, 2006.

[79] "Global Mobile Broadband: Market potential for 3G LTE," UMTS Forum White Paper, 2008.

[80] S. Chia, "Mobile Network Evolution Beyond 3G," presented at Radio and Wireless Symposium, 2007 IEEE, 2007.

[81] H. Ekstrom, A. Furuskar, J. Karlsson, M. Meyer, S. Parkvall, J. Torsner, and M. Wahlqvist, "Technical solutions for the 3G long-term evolution," *Communications Magazine, IEEE*, vol. 44, pp. 38-45, 2006.

[82] H. Holma, A. Toskala, K. Ranta-aho, and J. A.-P. Pirskanen, J., "High-Speed Packet Access Evolution in 3GPP Release 7," *Communications Magazine, IEEE*, vol. 45, pp. 29-35, 2007.