Cross Layer Schemes for QoS Provisioning in Mobile Communication Networks

A Dissertation submitted to
the Department of Information and Communication Systems Engineering of
the University of the Aegean
for the Degree of the Doctor of Philosophy,

by

Saied M. Abd El-atty

UNIVERSITY OF THE AEGEAN

Greece, March 2008.

The Supervisory committee

Angelos Rouskas, Supervisor, Assistant Professor
University of the Aegean

Stefanos Gritzalis, Associate Professor
University of the Aegean

Ilias Maglogiannis, Lecturer
University of the Aegean

Approved by the examinations committee on March 11th, 2008.

Angelos Rouskas, Supervisor, Assistant Professor
University of the Aegean

Stefanos Gritzalis, Associate Professor
University of the Aegean

Ilias Maglogiannis, Lecturer
University of the Aegean

Michael Theologou, Professor
National Technical University of Athens

Miltiades Anagnostou, Professor
National Technical University of Athens

George Vouros, Professor
University of the Aegean

Ioannis Anagnostopoulos, Lecturer
University of the Aegean

Summary

Next generation wireless and mobile cellular networks are expected to support multimedia applications with different quality of service (QoS). User mobility and time-varying multipath fading of wireless channels make QoS provisioning a challenging issue in wireless and mobile networks. In order to provide service differentiation and QoS guarantee in such networks, many new technologies have been proposed. Wideband code division multiple access (WCDMA) is the major multiple access technology for the third generation (3G) and beyond mobile communication systems. One major challenge in multimedia services over WCDMA-based networks is QoS provisioning with efficient resource utilization. The allocation and management of radio resources are crucial for such networks, where the scarce wireless spectrum resources are shared by multiple users. On the other hand, in the standard layered networking architecture, each layer is designed and operated independently. However, wireless channels suffer from both interference and inherent instability; moreover, the statistical channel characteristics of different users are different. The inflexibility of this architecture results in inefficient resource utilization in wireless systems and efficient interaction design across different layers is essential. Motivated by this fact, in this thesis, we focus on employing cross-layer techniques for resource allocation in mobile wireless networks; and a main objective is to improve the system performance by incorporating information from the physical layer at MAC layer into the design of radio resource management.

Concluding, our study was mainly focused on WCDMA based 3G and beyond networks and how to develop cross-layer mechanisms for QoS provisioning. Firstly, we propose a cross-layer framework for efficient packet scheduling procedure in WCDMA-based networks. At the medium access control (MAC) layer, we perform a prediction of the future state of the wireless channel for each user by utilizing information from the physical layer. The prediction criterion is based on the signal-to-interference ratio (SIR) measurements of the received signal at each mobile user; we then employ a finite state Markov chain (FSMC) model for predicting the future state of the wireless channel. According to this prediction we define which users should be served by the packet scheduler during the next frame. The main idea of the scheduling scheme is to prioritize the connections not only according to their delay sensitivity, but also according to the predicted state of their wireless channel during the next frame.

Next we propose radio resource management scheme as an effective combination of a call admission control (CAC) mechanism together with a traffic scheduling algorithm. These two schemes are designed and operated in a complementary fashion in order to support QoS provisioning for handover (HO) calls in WCDMA systems. The CAC mechanism is based on a guard code scheme in WCDMA system which employs orthogonal variable spreading factor (OVSF) codes as channelization codes. The guard code scheme reserves some code capacity to favor the continuation of handover (HO) calls over the new calls. However, during the handover procedure the HO calls experience high delays as a number of packets have to be forwarded through the wireline infrastructure to the target cell. Thus, we additionally employ a delay-aware traffic scheduler namely delay driven scheduler (DDS) at the MAC layer which aims to prioritize calls which experience high delays such as HO calls. Furthermore, DDS is able to exploit information from the physical layer in order to avoid erroneous packet transmission and increases system performance.

Finally, we propose an improved channel quality information (CQI) reporting scheme which aims to reduce the required CQI signaling overhead in high speed packet access (HSPA) systems by exploiting a CQI prediction method. According to this scheme, Node B predicts all the intermediary CQI reports between two subsequent CQI reports by utilizing a FSMC model of the wireless channel. The main benefit of the proposed scheme is the decrease of the uplink interference and thus the improvement of the uplink reception quality at the negligible cost of slight reduction of the efficiency of Node B scheduling procedure.

[Saied M. Abd El-atty]

Department of Information and Communication Systems Engineering

UNIVERSITY OF THE AEGEAN

Acknowledgment

I would like to thank

- Prof. Angelos Rouskas for his guidance and support,
- Dr. Dimitrios Skoutas for his helpful comments and his suggestions,
- The department of Information and Communication Systems Engineering of the University of the Aegean for offering me the opportunity for this PhD study,
- The National Scholarship Foundation of Greece (IKY) for funding and my studies,
- My family, my wife, my daughter (Judy). My sisters and my uncle for their efforts to give and support me to have joyful life,
- All of my friends in Egypt and Greece.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

3G:	Third Generation
3G+:	Beyond 3G (3.5G)
3GPP:	Third Generation Partnership Project
4G:	Fourth Generation
AMC:	Adaptive Modulation and Coding
BER:	Bit Error Rate
BS:	Base Station or Node B
BSC:	Binary Symmetric Channel
CAC:	Call Admission Control
CBR:	Constant Bit Rate
CDGPS:	Code Division GPS
CDMA:	Code Division Multiple Access
CQI:	Channel Quality Indication
DCH:	Dedicated Transport Channel
DDS:	Delay Driven Scheduler
DFS:	Delay Fair Scheduler
DFS_CL:	DFS Based Cross-Layer
DFS_PRED:	DFS with Prediction
DFS_PRED_CL:	DFS with Prediction based Cross-Layer
DSCH:	Downlink Shared Channel
DSSS:	Direct Sequence Spread Spectrum
E-CQI:	Enhanced CQI Reporting
EPM:	Equal Probability Method
FDD:	Frequency Division Duplex
FSMC:	Finite State Markov Chain
GPRS:	General Packet Radio System
GPS:	Generalized Processing Scheduler
GPS:	Generalized Processor Sharing
GSM:	Global System for Mobile Communication
HO:	Handover (Handoff)
HSDPA:	High Speed Downlink Packet Access
HS-DSCH:	High Speed DSCH

HSPA:	High Speed Packet Access
HSUPA:	High Speed Uplink Packet Access
IMT:	International Mobile Telephony
LAN:	Local Area Network
MAC:	Medium Access Control
M-LWDF:	Modified-Largest Weighted Delay First
MSC:	Mobile Switching Center
Node B:	Base Station or Node B
OSI:	Open Systems Interconnection (OSI)
OVSF:	Orthogonal Variable Spreading Factor
P-CQI:	Prediction-based CQI Reporting Scheme
PF:	Proportional Fair
PHY:	Physical
QoS:	Quality of Service
QPSK:	Quadrature Phase Shift Key
RLC:	Radio Link Control
rms:	Root Mean Square
RNC:	Radio Network Controller
RRC:	Radio Resource Control
SINR:	Signal to Interference And Noise Ratio
SIR:	Signal to Interference Ratio
TCP:	Transport Control Protocol
TDD:	Time Division Duplex
TFCI:	Transport Format Combination Indicator
TFI:	Transport Format Indicator
TFRC:	Transport Format And Resource Combination
TPC:	Transmit Power Control
TTI:	Transmit Time Interval
UE:	User Equipment
UMTS:	Universal Mobile Telecommunication Systems
UTRAN:	UMTS Terrestrial Radio Access Network
VBR:	Variable Bit Rate
VoIP:	Voice over IP
WCDMA:	Wideband CDMA
WiMAX:	Worldwide Interoperability for Microwave Access

Chapter 1 - Introduction

1.1 Introduction

In traditional communication networks, a layered protocol architecture has been widely adopted, and the performance optimization is conducted largely within each individual protocol layer. Although the layered structure is still of great importance for next generation wireless networks, it is known that this conventional architecture may not work well in many wireless data applications. For instance, consider data traffic over wireless links. The scheduling disciplines at the medium access control (MAC) layer being used may have no knowledge of the channel condition at the physical layer and hence this leads to degradation of the system performance. However, if the higher layers can utilize some information from the physical layer, the system performance may be improved. Improving the system performance by cross-layer techniques can yield significant gains, and the goal of this thesis is to introduce such techniques in WCDMA based 3G and beyond mobile networks.

On the other hand, meeting Quality of Service (QoS) guarantees in wireless communication systems is not a trivial task at all. The growth of wireless packet applications, e.g., wireless web access, interactive mobile multimedia applications and interactive gaming drives the rapid evolution of mobile wireless networks. Again, the communication channels and traffic patterns in these networks are more unpredictable than traditional wired networks. Moving away from the classical layering approach of a static and layer independent protocol stack, efficient cross-layer techniques would be necessary to achieve transmission protocol performance meeting these QoS requirements. Since QoS in each layer is defined in different terms the coordination between layers is essential in order to preserve the desired QoS. For example at the physical layer (PHY), QoS is synonymous to acceptable bit error rate (BER) or signal to interference ratio (SIR), while at the Medium Access Control (MAC) layer, QoS is usually defined in terms of minimum rate or maximum delay guarantees based on the scheduling discipline. In addition, call admission control is necessary to maintain the delivered QoS to the different users already admitted to the system by allowing or denying access according to the traffic load (new or handover requests), and type of traffic service.

Cross-layer design in wireless networks has recently earned much attention. Recent research has shown that a well-designed cross-layer approach that supports multiple protocol layer adoption and optimization can yield significant performance gains [1]. According to the various combinations and interactions across multiple layers different cross-layer designs can be constructed. In addition these cross-layer techniques are emerged for different type of wireless systems such as wireless LAN, mobile ad hoc and 3G networks. We chose to concentrate on the study of 3G and beyond WCDMA systems which are based on 3GPP standards. Also, we focused our study on the lower layers in WCDMA protocol stack: physical layer and MAC layer.

Figure 1.1 illustrates the basic components in the protocol stack of WCDMA-based networks. The call admission control (CAC) and the reservation protocol mechanisms are generally treated as the RLC/MAC layer functionalities in the wireless transmission protocol stack. The scheduling discipline based on queue management and the multi-transmission rate based on SIR measurements are layer-2 (i.e., MAC) and layer-1 (i.e., physical) functionalities, respectively. The major wireless protocol components considered in this thesis (i.e., multi-rate transmission based on OVSF codes, scheduling, CAC and bandwidth reservation) are shown in the transmission protocol stack of Figure 1.1.

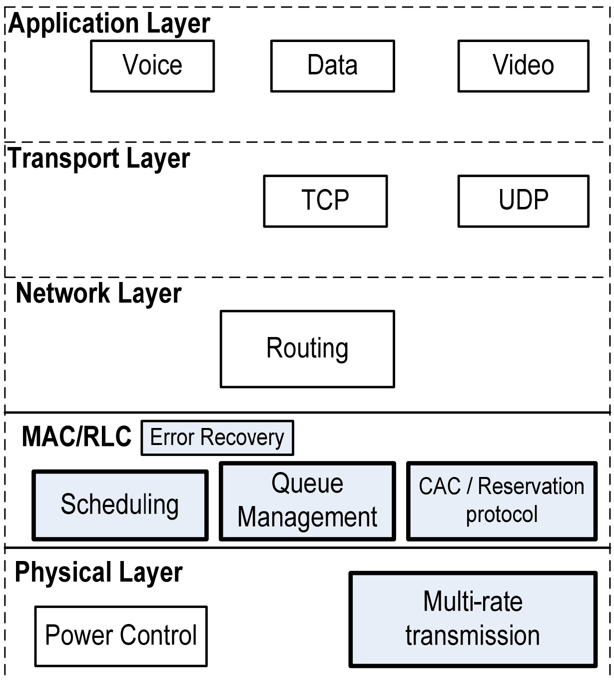

Figure 1-1: Standard protocol stacks and their components.

1.2 The Need for Cross-Layer Techniques

The design of modern wireless networks is in general based on a layered architecture. Each layer is associated with specific functions and communicates with other layers through a particular interface. The obvious advantage of this technique is that each layer can be redesigned without affecting the rest of the layers as long as it keeps the same interface. However the layered architecture can also redirect the network designer, leading him to isolate the functions of each layer from essential data coming from other layers. Over the past few years, many research works on wireless systems were concentrated on optimizing a single layer in the protocol stack [2]. The research focus has been to adapt existing algorithms and protocols independently and often have different objectives for multimedia transmission over wireless networks. However, an important aspect of mobile wireless networks is their dynamic behavior. The conventional protocol stack is inflexible as various protocol layers communicate in a strict manner [3]. In such a case the layers are designed to operate under the worst conditions as opposed to adapting to changing conditions. Furthermore, it is well known that the lack of correlation between network layers often causes the deterioration of the network throughput [4],[5].

Recent studies show that careful exploitation of some cross-layer protocol interactions can lead to more efficient performance of the transmission stack and hence to better application layer performances in different wireless networking scenarios [6]. One of several interpretations of cross-layer design is depicted at Figure 1.2. Cross-Layer Technique (CROT) concept means the interaction among the layers in the protocol stack as shown in Figure 1.2. As shown in the figure the technique consists of taking into account information available from different levels not necessarily adjacent [7], in order to create a system much more sensitive to its environment, load and usage. The information between the layers is communicated through upward and downward information. For example, the downward information carries the delay or loss constraints of the application in order to communicate to the link/MAC layer to enable the MAC to adapt its scheduling and error correction mechanisms. While in the upward, the information of transmission control power and packet loss are given to the application layer so that the user can adapt its data rate. In addition physical layer transmit power, BER information may be communicated to the link/MAC layer to make efficient packet scheduling. Clearly cross-layer techniques have the capability of adopting the system performance. Adaptation represents the ability of network protocols and applications to observe and respond to changing conditions. Finally, cross-layer interactions should be done in a controlled way, and preferably through a robust interaction among the physical layer and the higher layers.

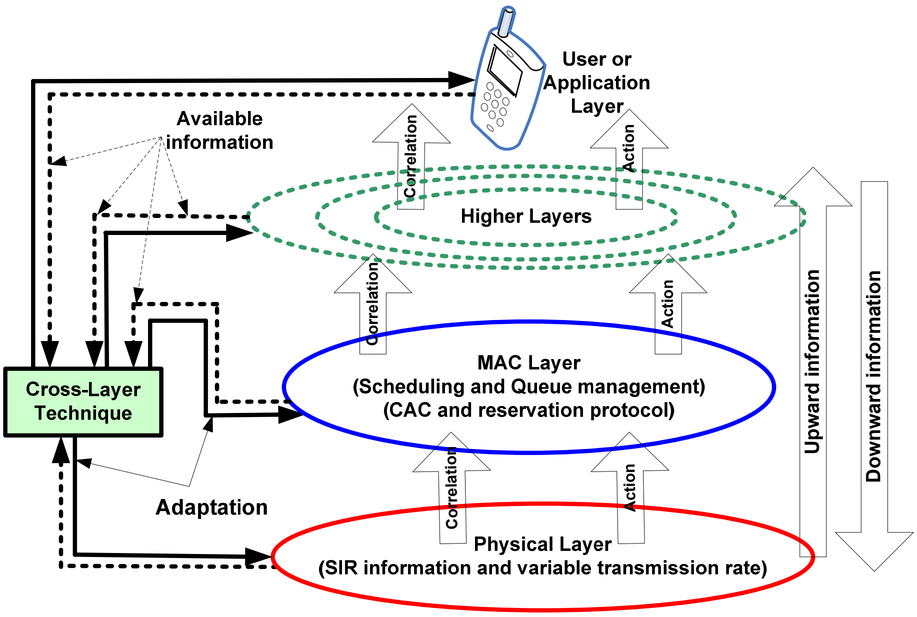

Figure 1-2: Cross-Layer Technique Concept.

Developing such robust cross-layer framework is of fundamental importance, since not only leads to improve multimedia performance over the existing mobile wireless networks such as 3G and 3.5G but also provides valuable insights into the design of the next-generation algorithms and protocols for heterogeneous wireless systems. Our proposed cross-layer framework does not necessarily require a redesign of existing protocols and can be performed by jointly optimizing the available information at the lower layer, such as SIR, channel state and the available strategies at the higher layer such as, scheduling, reservation bandwidth, admission control and resource management.

1.3 Objectives, Motivations, and Scope of the Thesis

The main objective of this thesis is to propose cross layer techniques for provisioning of services with different QoS in mobile wireless networks. The assumptions in the wireline networks are inadequate for mobile wireless networks, and the wireline scheduling algorithms are known to suffer from performance degradation in mobile wireless environments. This is because such environments are prone to packet losses due to both high bit error rates and wireless mobility in the cells. Before studying mechanisms by which services with different QoS can be supported over the wireless mobile networks, we briefly investigate the fundamental features of wireless system such as the wireless fading channels and wireless scheduling algorithms in Chapter 2. Generally, we consider each user's data are transmitted through the time varying, multi-path fading channel. The signal-to-interference ratio (SIR) value of the received signal is measured at the mobile user and transmitted back to the base station. Subsequently the SIR information is passed to the MAC layer. By representing the wireless fading channel by a finite state Markov chain (FSMC) model we manage to define a prediction criterion for the next frame transmission. Thus at the MAC layer, we perform a prediction of the future state of the wireless channel for each user by utilizing information from the physical layer. According to this prediction we define which users should be served by the packet scheduler during the next frame. Our cross-layer scheduler enjoys low-complexity implementation and provides QoS differentiation according to the delay sensitivity of each service flow.

Our main motivation was to introduce physical layer information to the MAC layer scheduling to improve the scheduler decisions. Our second motivation stemmed from the inherent requirement of mobile wireless networks to treat handover calls not in the same manner as new calls. We thus propose resource management algorithm as an effective combination between a call admission control (CAC) mechanism and a traffic scheduling algorithm. We propose a CAC based bandwidth reservation namely guard code scheme, which favors the handover (HO) calls over new calls at the call level in order to reduce the HO failure rate. The performance analysis of OVSF code tree is analyzed by a Markov chain. The code occupancy of the system is modeled by Markov chain and the differentiation between the HO and new call is performed at the code level by introducing a guard code scheme.

In 3G WCDMA system the bandwidth allocation to the connections is achieved through the use of the OVSF codes. Hence any guard scheme should be aware of the OVSF code tree characteristics. Therefore, although the scheme we propose belongs to the well known family of guard channel schemes, it differs from the other previously proposed schemes since it is adapted to the features of the OVSF code tree. Then we deploy this "guard code" scheme integrated with a traffic scheduling scheme at the MAC layer in order to provision the QoS requirement for handover calls. The scheduling scheme utilizes a priority criterion which prioritizes the connections not only according to their delay sensitivity but also according to their predicted error probability.

Finally we propose an improved CQI reporting scheme in high speed packet access (HSPA) networks. The scheme is based on a prediction method and aims to reducing the signaling overhead as well as increasing the efficiency of the adaptive modulation and coding (AMC) procedure in high speed downlink packet access (HSDPA) system.

1.4 Organization of The Thesis

The organization of this thesis is as follows:

- **Chapter 2** provides a background and survey of the most important components in wireless communication systems. The details of the wireless system architecture are investigated. The characteristics of the wireless channel and fading issues are explained. We also discuss the Markov model for wireless fading channels. The wireless scheduling algorithms are classified according to the QoS type and then we give a brief description of the most representative algorithms. Finally the most desirable characteristics of the wireless scheduling schemes are presented.

- **Chapter 3** describes the Wideband Code Division Multiple Access (WCDMA) system which is the basic system model in 3G mobile networks. In this Chapter we introduce necessary aspects that govern WCDMA based downlink transmissions. The details of WCDMA model are investigated in terms of the operation modes, spreading and modulation process, orthogonal variable spreading factor (OVSF) as channelization codes, WCDMA protocol layers and types of transport channels. Finally we discuss the QoS classes in 3G networks.

- **Chapter 4** presents a cross-layer scheduling framework for supporting data applications over WCDMA based networks. In this Chapter, we introduce a cross-layer framework which aims to make the packet scheduling procedure more efficient. By deploying a finite state Markov chain (FSMC) model, the future state of the wireless channel for each connection is predicted. Based on these predictions only connections with their predicted SIR above a predefined threshold are allowed to be served. The proposed cross-layer framework consists of the prediction criterion and the employed scheduling discipline. The main idea of the scheduling scheme is to prioritize the connections not only based on their delay requirements but also based on their predicted wireless channel state. Finally, simulation results are presented to measure the efficiency of the proposed cross-layer framework.

- **Chapter 5** presents handover traffic management with QoS provisioning in 3G WCDMA networks. This Chapter consists of two parts. In the first part we propose a "guard code" scheme for handover (HO) traffic management in WCDMA system employing OVSF codes. The scheme introduces a code reservation threshold, which allows HO calls to use the total capacity of OVSF code tree, but restricts the new calls to use only a part of the tree capacity. We model the code occupancy of the OVSF code tree by a Markov chain. We study the behavior of the system in terms of new call blocking, HO failure probability and code utilization under different traffic loads and different code reservation thresholds. In the second part we introduce a new design of radio resource management for handover provisioning in WCDMA networks. The algorithm consists of an effective combination between the CAC guard code scheme and traffic scheduling scheme and at the same time makes use of the cross-layer framework introduced in the previous Chapter. These two schemes are designed and operate in a complementary fashion in order to support QoS provisioning in WCDMA systems. We employed the guard reservation code at the call-level and a delay driven scheduler (DDS) at the packet-level in order to provide robust radio resource management in cellular WCDMA-based networks. Simulation results demonstrate the effectiveness of our proposal.

- **Chapter 6** introduces an improved CQI reporting scheme for HSDPA system. In this Chapter we propose a CQI reporting scheme, namely, prediction-based CQI scheme (P-CQI) which reduces the required CQI signaling by exploiting a CQI prediction technique. First we evaluate the gain in the uplink quality reception if the P-CQI is introduced. The accuracy of the adaptive modulation and coding (AMC) in HSDPA depends on CQI reporting scheme. Thus we introduce a packet scheduling algorithm based on the P-CQI reporting scheme and measure systems performance.

- **Chapter 7** concludes this thesis and outlines a few directions for future research.

Chapter 2 - Wireless Mobile Communication System- Overview

This Chapter provides an overview on the components of wireless networks. We investigate the basic elements in the wireless mobile communication system, such as the wireless link (the air interface between the mobile users and the base station) and the wireless scheduler at base station.

2.1 System Architecture

The wireless communication network typically consists of a fixed network backbone and a wireless system. The fixed network part, through a mobile switching center (MSC), which provides connections between radio-access stations, often called base stations (BSs). The wireless system has a cell structure, the service area is divided into cells and each cell has a BS [8]. Within a cell mobile stations (MSs) communicate via wireless links to the BS. The objective of the BS is to schedule these MSs in a timely manner in both downlink and uplink direction by employing a particular scheduler. In this thesis we restrict ourselves to just considering the downlink problem, where the direction of the data flow is from the BS to the MS. The architecture of simple wireless system with two-state channel model is shown in Figure 2.1.

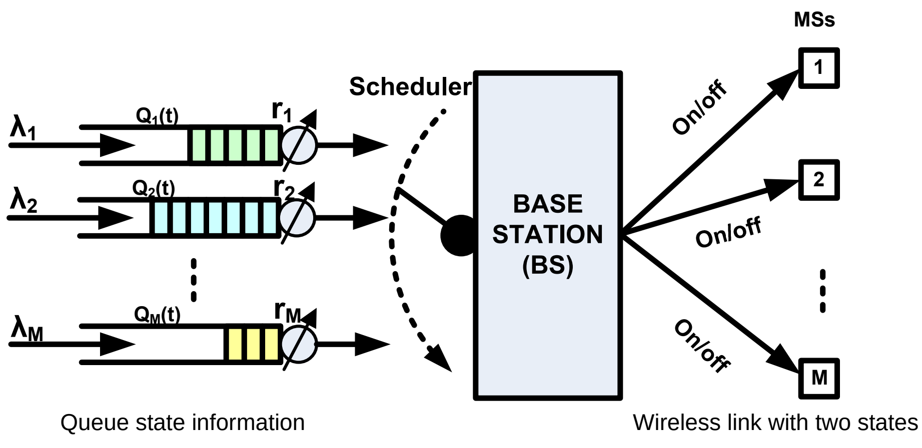

Figure 2-1: Wireless system architecture.

The BS simultaneously serves, M users; each of which has a queue (Q) to receive its incoming packets with data rate, r, in every frame, T_s. The wireless link between a BS and each of the MSs are independent of each other. In order to make scheduling decisions at BS certain information such as, the number of the connections, their reserved rates, and channel states and the status of the connection queues is needed by the scheduler. Wireless links are subject to bursty errors. These errors are caused either due to mobility of the MSs or due to changes in the surrounding physical environment. One of the most popular methods to model the wireless link is a two-state (good/bad) channel model, i.e., the channel is either one of the two states: good (or error free) or bad (or error) state [9], [10]. If the wireless channel is in a bad state, packets transmitted on the link will be corrupted with very high error probability. Therefore, in order to reduce bandwidth wastage, the conventional scheduler defers transmissions of users whose links are in the error state and compensates those users when their links recover. However, in reality the capacity of a wireless link does not jump between zero capacity and full capacity [11]. To model the links more accurately, more states should be used. In Chapter 4, we use a channel model with multiple states based on a finite state Markov chain (FSMC). In the following sections, we briefly overview each part in the wireless system architecture.

2.2 Wireless Channel Architecture

A wireless system experiences special problems that do not exist in wireline systems, such as time-varying channel capacity and location-dependant errors [12]. Packet transmissions on wireline networks have very low error rate due to the high quality of the media. However, wireless channels are more error-prone and suffer from interference, fading and shadowing. As a result, the capacity of a wireless link has very high variability [13]. Therefore, the characteristic of the wireless communication poses special problems such as: 1) high error rate and bursty errors; 2) location dependent and time varying wireless link capacity; 3) scarce bandwidth; 4) user mobility; 5) power constraint of mobile users. Therefore, the function of the scheduler in wireless networks is not only assigned resources (as time slots) as in wireline networks but also assigned powers, data rates, channels, or combination of them when packets are transmitted [12].

2.2.1 Mobile Radio Channels

The shared wireless channel is not only time-dependent but also location-dependent. This is due to the shared wireless transmissions are shared among multiple users, and hence effective channel capacity is time-dependent. In addition, due to interference, fades, multipath effects, channel errors are also location-dependent. In mobile cellular networks, the received signal at the mobile receiver consists of several scattered signals. As shown in Figure 2.2, the signal travels from the BS to the MS and vice verse by many different paths with different strengths. This type of fading called multipath Rayleigh fading when the scattered signals strength are independent identical distributed (iid) Gaussian random variables [8]. The envelope of the received signal at the mobile station in Figure 2.2 is considered to be Rayleigh distributed given by [14]:

$$p(x) = \frac{x}{\beta^2}\exp\left(-\frac{x^2}{2\beta^2}\right) \qquad 0 \le x \le \infty \tag{2.1}$$

where β is the rms value of the received signal before envelope detection, and β^2 is the time-average power of the received signal before envelope detection. One of the most important methods for modeling the multipath Rayleigh fading is Jakes model [15]. In this method the movement of the mobile changes the received scatters signals at the mobile receiver and then changes the Doppler shift or Doppler frequency, the maximum Doppler frequency is $f_d = v.f_c / c$ due to mobile velocity v. According to the 3GPP standard, radio mobile environment consists of three different radio environments: Vehicular, Pedestrian and Indoor which are related to the design of cell structure macrocell, microcell and picocell respectively [16]. In our system model we consider a macrocell environment.

The different propagation path lengths of the multipath signal will lead to different propagation time delays. For example, the delay profile extends typically from 1 to 2μs in urban and suburban areas, although in some cases delays as long as 20μs or more with significant signal power have been observed in hilly areas. In WCDMA systems, the chip duration at 3.84Mcps is 0.26μs. Therefore, if the time difference of the multipath components is at least 0.26μs, the receiver can separate those multipath components by using a RAKE receiver and combine them coherently to obtain multipath diversity [17].

2.3 Markov Model for Wireless channel

The Markov model is the most widely method to represent the wireless channel. There are several reasons for the popularity of Markov models. These models are analytically tractable, their theory is well established in the statistical literature, and they have been applied successfully to a variety of important communication problems.

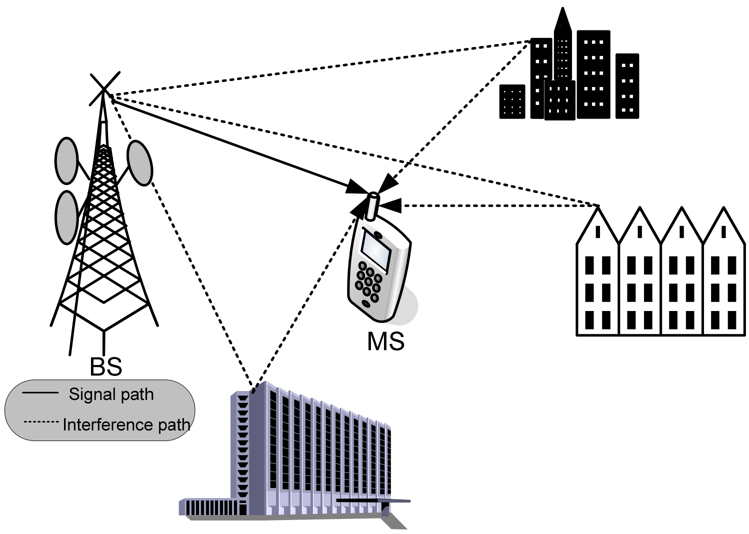

Figure 2-2: Multipath propagation in urban area.

Markov modeling techniques are directly applicable to the modeling and analysis of digital communication channels [18]. Also, Markov models can be used to evaluate the capacity of a discrete channel and for the design of optimal error control coding techniques. Markov models have been successfully used to characterize fading channels in wireless communication systems [19], [20], [21], [22] and [23]. More importantly, computationally efficient techniques are available for estimating the parameters of Markov sequences from simulated or measured error patterns. One of the most well known channel models used by the researchers in the scheduling algorithms is the two-state Markov model Gilbert-Elliott model [9], [10].

2.3.1 Two-state Markov model

In order to evaluate the two-state Markov model, we consider a fading channel in which the received signal strength is above an acceptable threshold part of the time, and below the threshold during a deep fade. If we have interest only in the "above threshold" or "below threshold" conditions, we can model the channel to be in either one of the two states as shown in Figure 2.3. We have a good state, g , in which the system performance is acceptable (for example, probability of error < threshold value, $P_E < P_{th}$) and a bad state, b , in which the received signal level is so low the probability of error is unacceptably high ($P_E > P_{th}$). Therefore, in the simple two-state Markov model shown in the figure, the states can be represented by the set, $S = \{g, b\}$.

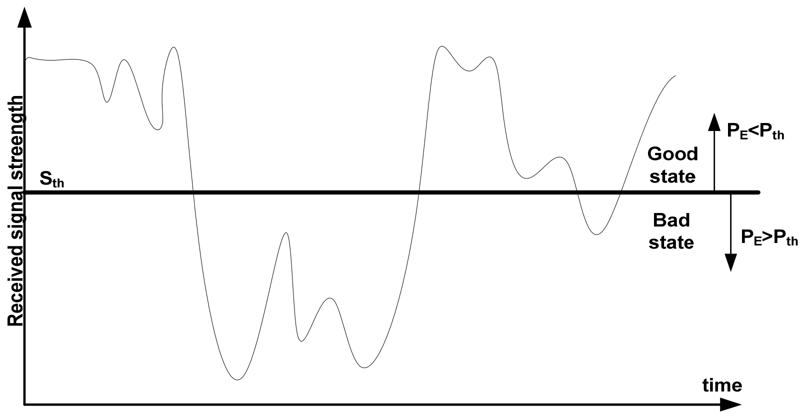

Figure 2-3: Received signal level with two-state model.

2.3.2 Gilbert-Elliott's Model

The Gilbert-Elliott model [9], [10] and [24] is a time varying binary symmetric channel. The transition probabilities of the channel are determined by the current state of a discrete time stationary binary Markov process. The source has two states: g (for good or no errors) and b (for bad or burst errors). This constitutes the Markov chain with the state transition matrix, $A(t)$,

$$A(t) = \begin{bmatrix} Q & p \\ P & q \end{bmatrix} \tag{2.2}$$

The probability of remaining in good state, given the model is in good state, is Q and the probability going to bad state is p. The probability of remaining in bad state and the probability of going from bad state to good state is q and P respectively. Note that each row of the state transition matrix must sum to 1. The state transition diagram of the two-state model is shown at Figure 2.4.

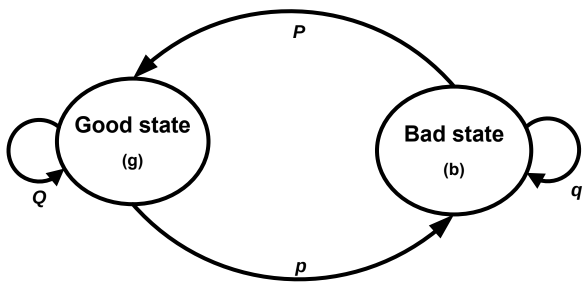

Figure 2-4: Gilbert-Elliott's channel model.

2.4 QoS Scheduling Disciplines

The packet scheduling algorithm is one of the most important components for QoS provisioning mechanism [25], [26]. The design of a scheduling scheme for wireless mobile communication networks that enables sharing the air interface among multimedia applications with different QoS requirements is not a trivial task. Compared with the scheduler design for the wireline networks, the design of scheduling for wireless networks with QoS guarantees is particularly challenging. This is due to the fact that the wireless channels have low reliability, and time varying signal strength, which may cause severe QoS violations [13]. On the other hand, a unique feature of wireless scheduling with QoS guarantees is its channel state dependency [27], i.e., how to schedule the resources depends on the channel state. This is necessary since without the knowledge about the channel state, it is impossible to guarantee QoS. In addition, the capacity of a wireless channel is severely limited, making efficient bandwidth utilization a main concern. Therefore, the task of wireless scheduling algorithm is scheduling in terms of resources in time slots (code), powers, data rates, channels, or a combination of those. An overview of scheduling techniques for wireless networks can be found in [27], where the major issues in wireless scheduling have been summarized. In the next sections, we summarize the QoS scheduling disciplines in wireless networks.

2.4.1 Fair Queueing

Several important scheduling algorithms, capable of provisioning certain guaranteed QoS, have been developed for wireline networks. However, these existing scheduling disciplines such as fair queueing scheduling [28], virtual clock [29], and earliest due date (EDD) [30] are not directly applicable in wireless networks because they do not consider the varying in wireless link capacity and the location dependent channel state. Scheduling algorithms for wireless networks were first designed to follow the rules of the fair queueing in wireline systems. Packet fair queuing (PFQ) [28] scheduling algorithms are based on generalized processor sharing (GPS) discipline [31] and can be applied in wireless networks. In addition, idealized wireless fair queuing (WFQ) scheduling algorithm based on wireless fluid fair procedure is proposed at [32]. Although these PFQ algorithms have taken into consideration both bursty channel errors and location dependent errors, they mainly focused on guaranteeing fair sharing of the channel resources ignoring the guarantees of the QoS requirements such as throughput and delay.

2.4.2 BER Scheduling: WISPER

A novel MAC protocol called wireless multimedia access control protocol with BER scheduling (WISPER) for CDMA system was proposed in [33]. The WISPER was developed to take advantage of power control characteristics of the IS-95 standard. This protocol was designed so that the bit error rate (BER) of the transmission channel is maintained below a given BER threshold value.

In this algorithm, each active user has a batch of packets which represents the traffic arrival, and then each batch has a priority value proportional to the packet number in the batch and inversely proportional to the remaining time before the batch expires. Packets with the same or similar BER requirements are transmitted with the same power level and in the same time slot whenever possible. This provides significant performance advantages in the presence of power control imperfection [34]. However, the maximum number of packets accommodated in a slot is determined by the most stringent BER requirements of the allocated packet, thus resulting in the underutilization of a time slot. To overcome this shortcoming, different power levels should be assigned in each slot based on the BER requirements of the allocated MSs. Under this power allocation strategy, resource underutilization in each slot is avoided and better system performance is achieved [35] at the cost of more complex power allocation. In our scheduling scheme, we make use of the fast power control in WCDMA systems to achieve the BER requirements of each user.

2.4.3 Multiuser Diversity Gain: The Proportional Fair Scheduler

The Proportional Fair (PF) scheduler was the first algorithm that achieves multiuser diversity in high speed data networks. The priority criterion, $m(t)$ that is employed in this scheduler is defined as follows [36]:

$$m(t) = \frac{R(t)}{th(t)} \tag{2.6}$$

where $th(t)$ is an estimate of the user's average throughput in some window of time prior to the current slot and $R(t)$ is the rate requested by the user in a given slot. If

i^* denotes the user with the highest priority, then $th(t)$ can be obtained by averaging the i^{th} user's throughput over a scheduling time scale t_c i.e.,

$$th_i(t+1) = \begin{cases} \left(1-\dfrac{1}{t_c}\right)\times th_i(t) + \dfrac{1}{t_c}R_i(t) & i = i^* \\ \left(1-\dfrac{1}{t_c}\right)\times th_i(t) & i \neq i^* \end{cases} \tag{2.7}$$

Summarizing PF algorithm increases the cell throughput at the cost of increased service delay time.

2.4.4 Provisioning QoS: Throughput and Delay Guarantees

In recent years, improving wireless technologies have increased the demand for wireless data services [37]. Real-time applications such as voice, video conferencing and games share network resources with non-real-time traffic such as file transfers and messaging. QoS supporting for wireless data is therefore a natural consequence of the integration of packet switched wireless networks with the Internet, placing new demands on scheduling algorithms.

The authors at [38] have proposed an efficient scheme for provisioning QoS of different users sharing the capacity of the wireless channel. They named it as Modified-Largest Weighted Delay First (M-LWDF) and it aims at achieving optimal throughput for each user in the serving base station. Furthermore, they have extended this scheme in order to provide optimal throughput together with delay requirements of each user [39]. Many works have been based on M-LWDF idea, the authors at [40] presented an efficient scheduling scheme for the mixture of real and non-real time applications in high speed data networks. This scheme used exponential (EXP) method for averaging the throughput and delay requirements of the users sharing the capacity of wireless channel. The efficiency of this scheme was the capability of achieving the QoS provisioning like M-LWDF for mixture applications in high speed data networks.

In [41], a fair scheduling and admission control framework is proposed for service differentiation in packet switched networks. They provided an analytical model at connection and packet level as well as service differentiation according to the traffic type.

A framework provides QoS guarantee has two important elements. The first is an admission control policy that determines the level to which a user can utilize system resources. The second is a mechanism to evaluate the QoS obtained by an appropriate scheduling algorithm in terms of its control parameters. In fact, this mechanism determines whether the QoS requirements of an admitted user can be supported. In Chapter 5, we propose a new effective combination between a CAC mechanism and a traffic scheduling algorithm which is based on a cross layer framework for QoS provisioning in WCDMA –based networks.

Summarizing the wireless scheduling algorithms should achieve more or less the following desirable requirements as illustrated in Table 2.1 [42] and [43]:

Efficient link utilization	The algorithm must utilize the channel efficiently. This implies that the scheduler should not assign a transmission slot (code) to a session with a currently bad link since the transmission will simply be wasted.
Delay bound	The algorithm must be able to provide delay bound guarantees for individual sessions in order to support delay-sensitive applications.
Fairness	The algorithm should redistribute available resources fairly across sessions. It should provide fairness among error-free sessions (short-term fairness) and error-prone sessions (long-term fairness).
Throughput	The algorithm should provide guaranteed short-term throughputs for error-free sessions and guaranteed long-term throughputs for all sessions.
Implementation complexity	A low-complexity algorithm is a necessity in high-speed networks in which scheduling decisions have to be made very rapidly.
Energy consumption	The algorithm should take into account the need to prolong the MS battery life.
Delay/bandwidth decoupling	For most schedulers, the delay is tightly coupled with the reserved rate; i.e., a higher reserved rate provides a lower delay. However, some high-bandwidth applications, such as Web browsing, can tolerate relatively large delays.
Scalability	The algorithm should operate efficiently as the number of sessions sharing the channel increases.

Table 2-1: Desirable requirement in wireless scheduler.

Chapter 3 - WCDMA System Model

3.1 Introduction

Unlike 2G mobile systems like global system for mobile communication (GSM), third generation (3G) wireless mobile systems like Universal Mobile Telecommunication System (UMTS) provide a wide variety of packet data services [17,44]. The radio resource management module allocates the available resources to the active users of the network. These resources could be radio resources: they are time-slot in the case of a 2G like GSM or 2.5G networks like general packet radio system (GPRS), but in a 3G networks like UMTS using WCDMA (wideband code division multiple access) technology, they are spreading codes and power [45]. WCDMA is the main third generation air interface in the world and deployed in Europe and Asia, including Japan and Korea, in the same frequency band, around 2GHz [46]. The purpose of this Chapter is to present the principles of the WCDMA-UMTS system model as specified in the 3GPP standards.

3.2 Characteristics of WCDMA

Third Generation Partnership Project (3GPP) produced the first full version of the WCDMA standard at the end of 1999. This release, called Release'99, contains all the necessary elements to meet the requirements for international mobile telephony (IMT-2000) technologies, including 2Mbps up to 10Mbps data rate with variable bit rate capability, support of multi-service, QoS differentiation and efficient packet data.

WCDMA supports variable user data rates, Bandwidth on Demand (BoD), where each user is allocated frames of 10ms duration during which the user data rate is kept constant [45]. However, the data capacity among the users can change from frame to frame. This radio-capacity allocation is controlled and coordinated by radio-capacity resource management functions in the network to achieve optimum throughput for packet data services and to ensure sufficient QoS [47]. Furthermore, UMTS-based WCDMA supports QoS to provide services from low-delay conversational class to flexible background class. These QoS improve the efficiency of the air interface as the radio network can optimize the resource allocations according to the QoS of each service, thus avoiding overdimensioning of the network. The variable bit-rate capability and service multiplexing in WCDMA system are illustrated in Figure 3.1.

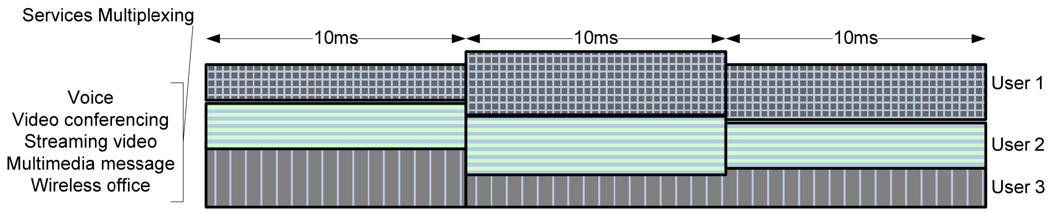

Figure 3-1: Variable bit rate and service multiplexing in WCDMA system.

3.2.1 WCDMA Operation Modes

The WCDMA standard has two modes for the duplex method: A Frequency Division Duplex (FDD) and Time Division Duplex (TDD). The frequency bands allocated for UTRA are shown in Figure 3.2. In UTRA there is one paired frequency band in the range 1920–1980MHz and 2110–2170MHz to be used for UTRA FDD. There are two unpaired bands from 1900–1920MHz and 2010–2025MHz used for the operation of UTRA TDD [48]. In this thesis we are only interested of the FDD mode. Table 3.1 lists the most important parameters of the UTRA FDD [49].

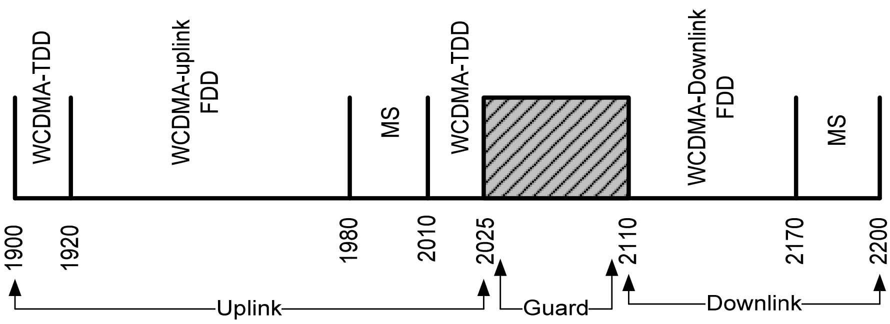

Figure 3-2: The frequency spectrum in UTRA.

The goals of 3G networks based on WCDMA system is to provide multimedia and high-speed data services at rates up to 2Mb/s, the nominal channel bandwidth is 5MHz and the chip rate is 3.84Mcps [50] and [51] as can be seen in Table 3.1. A service provider may, however, adjust the channel bandwidth if necessary to optimize the spectrum utilization. The center frequency must be an integer multiple of 200kHz. The following specifications apply to the radio transmission and reception in the UMTS FDD mode. The maximum transmitter power of the user equipment (UE) is in the range of 21 to 33dBm (i.e., 125mW to 2W). The receiver sensitivity, which is nominally defined as the minimum receiver input power at the antenna port such that the bit error ratio (BER) is 0.001 or less, is -117dBm for the UE and -121dBm for a base station. With transmit power control (TPC) commands in steps of 1dB or more, the UE and base station can adjust their transmitter power output.

Multiple Access scheme	DS-CDMA
Channel Multiplexing in downlink	Data and control channels time multiplexed
Multirate/Variable rate scheme	Variable spreading factor and multi-code
Chip Rate	3.84 Mcps
Channel bandwidth	5 MHz/200kHz carrier
Spreading Factors	4-256 (uplink) and 4-512 (downlink)
Frame Length	10 ms
Inter BS synchronization	No accurate synchronization needed
Data Modulation	QPSK (downlink)
Channel Coding Scheme	Convolution code (rate ½ and 1/3) and Turbo Code
Handover	Soft and Inter-frequency handover
Power Control	Open and fast closed loop (1.5kHz)

Table 3-1: Technical features in WCDMA-FDD system.

3.2.2 WCDMA Spreading Operation

WCDMA spreading procedure consists of two operations. The first operation is the channelization operation where the spreading code is applied to every symbol in the transmitted data. Thus the bandwidth of the data signal is increased. In this channelization operation, the number of chips per data symbol is called the spreading factor (*SF*). The second spreading operation is the scrambling operation, where a scrambling code is applied to the already spreaded signal. Both of the spreading operations are applied to the so called In-phase (I) and Quadrature phase (Q) branches of the data signal. In the channelization operation, the Orthogonal Variable Spreading Factor (OVSF) codes are independently applied to the I and Q branches [50] and [52].

3.2.3 WCDMA Coding and Modulation

For the channel coding in WCDMA system, the standard suggests three methods of coding for different Quality-of-Services (QoS) requirements [53]. The three coding schemes are: Convolutional coding, Turbo coding, or no coding. The selection of one of the three options is done by the upper layers. In addition, bit interleaving is used to improve the Bit Error Rate (BER). The modulation scheme selected in 3GPP WCDMA standard is QPSK for downlink and is BPSK for uplink [50].

Figure 3.3 illustrates a general structure of the spreading and modulation processes. As can be observed, the coded bits that have been mapped onto the physical channel are initially multiplied by the channelization code C_{ch}. The multiplication assumes that the values {0, 1} of the bits are mapped to levels {+1, -1}, respectively. The code sequence has a length of *SF* chips and, as a result, *SF* chips are obtained from each bit. Then, the chip rate *W* after the channelization procedure is *SF* times the original bit rate. After the channelization procedure, the obtained sequence is multiplied by the scrambling code C_{scr}. This procedure does not modify the bandwidth of the signal, since the chip rate of the scrambling code is equal to the chip rate of the sequence after channelization. Then the chips resulting from the scrambling procedure are modulated and transmitted to the antenna at the corresponding carrier frequency. The final bandwidth of the signal at antenna *BW* will depend on the chip rate *W* and the modulation scheme.

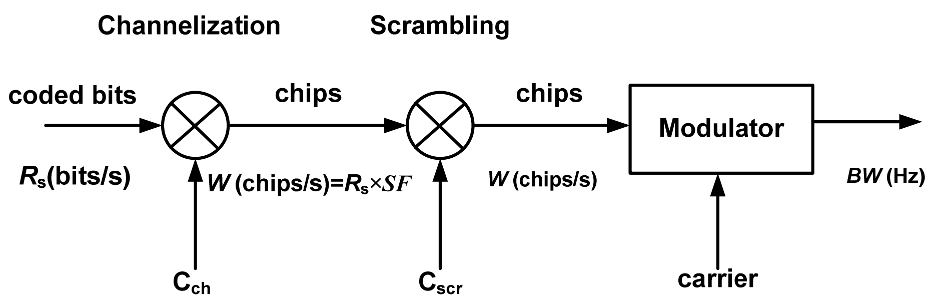

Figure 3-3: WCDMA spreading and modulation procedures.

3.3 OVSF Channelization Codes

The channelization codes are used to spread the data to the chip rate, preserving orthogonality between physical channels with different rates and spreading factors. So-called orthogonal variable spreading factor (OVSF) codes are used for the channelization [17,50]. The OVSF codes are organized according to a code tree structure, as depicted in Figure 3.4, each level in the code tree corresponds to a certain spreading factor. The channelization codes are uniquely defined by $C_{SF,k}$, where *SF* is the spreading factor of the code and *k* is the code number $0 \leq k \leq SF-1$. The tree generation rule states that, for each code {x} with spreading factor *SF*, two descendant codes with spreading factor 2×SF are obtained: the first one, in the upper branch, is generated simply by repeating the predecessor code (i.e. {x, x}) and the second one, in the lower branch, is generated by repeating and inverting the predecessor code (i.e. {x, -x}). Although this generation procedure can be extended indefinitely, in practice the allowed values of spreading factor for UTRAN FDD range from 4 to 256 in the uplink and from 4 to 512 in the downlink direction. The downlink orthogonal codes within each base station are managed by the radio network controller (RNC) in the 3G network.

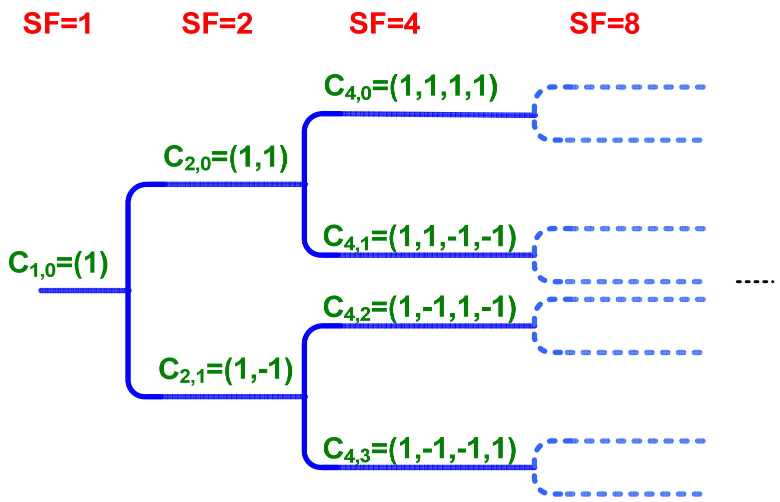

Figure 3-4: Generation of channelization codes by OVSF tree.

In WCDMA system the signal transmitted from the base station (BS) to the mobile has a symbol rate of *W*/*SF*, where *W* is the chip rate and *SF* is the spreading factor. Since the chip rate is fixed to 3.84Mchips, the service rate is determined by the spreading factor. If we denote with $R = W/512$ the lowest service rate, then each connection may have a service rate R_S that is always a multiple of a power of two of the lowest available rate *R*.

3.4 WCDMA Protocol Layers

The UMTS utilizes a layered protocol architecture at different interface points, each layer performing a set of specific functions. These architectures are usually described in terms of the control plane and user plane protocols [54,55]. The control plane protocols are concerned with the signaling and control required to establish a connection between a UE and the network, or to request specific services or resources from the network. The user plane protocols, on the other hand, specify how the user data is to be transferred across an interface after a connection has been established between the UE and the network [17]. Figure 3.5 shows the protocol architecture of the UTRAN. It is also the lower-layer protocols on UE. On the UTRAN side, the physical layer, which is responsible for carrying the information bits, is provided by the BS while the other layers reside in the RNC. The Radio Resource Control (RRC) is a layer-3 protocol in the control plane that interfaces with the radio link control (RLC) sublayer of layer-2 and terminates in the UTRAN [45]. The RRC layer simply passes the messages to the higher layers or to the RLC layer below to the physical layer.

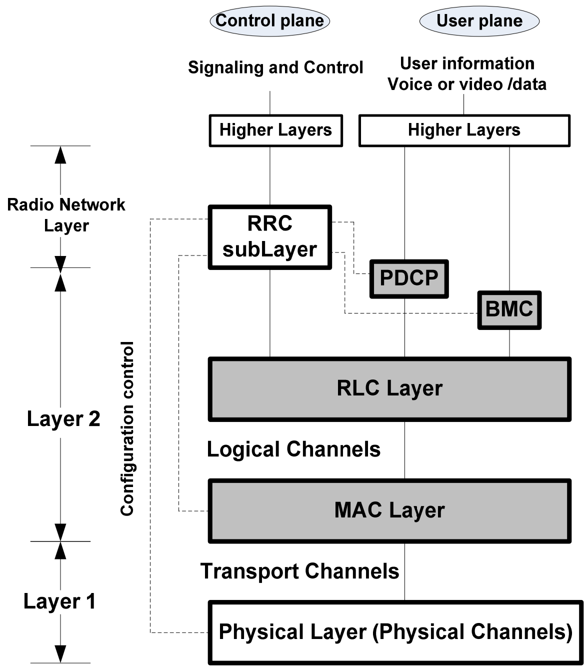

Figure 3-5: The Layered Protocols at UTRAN.

3.5 WCDMA Transport Channels

In WCDMA system the data generated at higher layers is carried over the air with transport channels, which are mapped in the physical layer to different physical channels [17,52]. The corresponding period to transmit a transport block is known as the transmit time interval (TTI). In Release 99, the TTI can take the values of 10, 20, 40, or 80ms. For voice services, the TTI is fixed at 10ms whereas for data services it changes according to the service used. The physical layer is required to support variable bit rate transport channels to offer bandwidth-on-demand services, and to be able to multiplex several services to one connection [17]. WCDMA supports several transport channels for flexible packet data operation [56]. Each transport channel is accompanied by the transport format indicator (TFI) at each frame interval at which data is expected to arrive for the specific transport channel from the higher layers. Here, we study the most important transport channels in WCDMA-based networks.

3.5.1 Dedicated Transport Channel (DCH)

The dedicated transport channel (DCH) carries all the information intended for the given connection coming from layers above the physical layer, including data for the actual service as well as higher layer control information [17,52]. The content of the information carried on the DCH is not visible to the physical layer, thus higher layer control information and user data are treated in the same way. The dedicated transport channel is characterized by some fundamentals features such as fast power control, fast data rate change on a frame-by-frame basis, and the possibility of transmission to a certain part of the cell or sector with varying antenna weights with adaptive antenna systems. In addition, the dedicated channel supports soft handover.

3.5.2 Downlink Shared Channel (DSCH)

The downlink shared channel (DSCH) is a downlink transport channel, which is carried by a physical downlink shared channel (PDSCH), and is modulated with normal QPSK [17,52]. A downlink dedicated channel (Downlink DCH) is always operating together with each DSCH. The DSCH is a transport channel intended to carry dedicated user data and/or control information; it can be shared by several users. The transport format combination indicator (TFCI), which is carried on a frame-by-frame basis on the associated DCH, allows the DSCH to dynamically vary its spreading factor at each frame. By definition, the DSCH capacity can be shared among several users on a frame-by-frame basis, where each frame has duration of 10ms and consists of 15 time slots. The transmission of high peak rate and low duty cycle data in a DSCH channel, does not lead to low utilization as in the case of DCH channels, since a user may be allocated a different rate every 10ms. That allows several users to be active in parallel using lower rates or one user to be active on high data rate.

3.5.3 Frame Structure of Transport Channels

In the downlink, the dedicated physical data channel (DPDCH) and dedicated physical control channel (DPCCH) are time multiplexed within each radio frame. The downlink DPDCH contains layer-2 data, while the DPCCH carries pilot bits, transport-power-control (TPC) commands, and an optional TFI where the TFI informs the receiver side what TF is used in the current data frame in order to simplify detection, decoding, and demultiplexing [17,45]. The downlink frame structure of DPDCH is shown at Figure 3.6, each frame of length 10ms is divided into 15 slots of length 0.625ms, each corresponding to one power-control period.

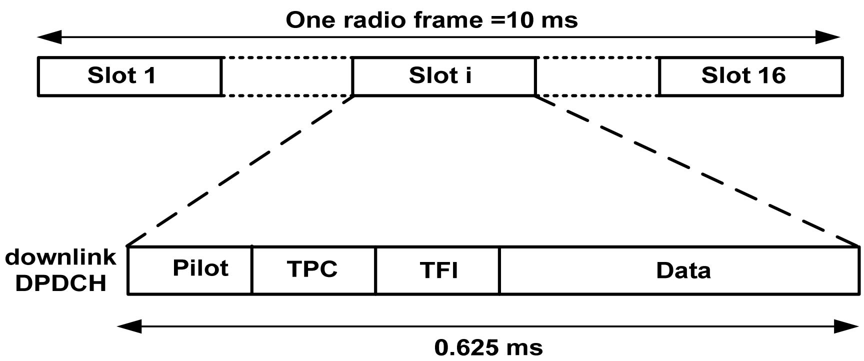

Figure 3-6: Frame structure of downlink DPDCH in WCDMA system.

3.6 WCDMA QoS

The QoS provisioning for the wireless network must be robust and capable of providing reasonable QoS resolution [57]. With the growing deployment of wireless data services, it is highly expected that future wireless networks will provide services for different services of traffic. These networks should satisfy services QoS requirements such as bit error rate, bounded packet delay and required data rate. In order to provide service differentiation and QoS guarantee in wireless systems, many technologies have been proposed. For example, the QoS in WCDMA-based UMTS network must take into account the restrictions and limitations of the air interface where all users use the same bandwidth at the same time and therefore, users interfere with one another. Because of the propagation mechanism, the signal received by the base station from a user close to the base station will be stronger than the signal received from another user located at the cell boundary. Hence, the close users will dominate the distant users. This situation becomes much worse if the close user happens to be a high speed data user [58]. To achieve considerable capacity and QoS, all signals should arrive at the base station with the same mean power, irrespective of distance. A solution to this problem is perfect power control, which attempts to achieve constant mean power for each user. The UTRA/WCDMA air interface uses power control on both uplink and downlink at 1500 Hz [59].

3.6.1 QoS Classes

In WCDMA based networks such as 3G and beyond, a great number of services such as video telephony, streaming multimedia, web browsing and e-mail are supported. These applications and services can be divided in four different QoS classes [59]: conversational, streaming, interactive, and background. The main distinguishing factor between these QoS classes is how delay sensitive the traffic is: conversational class is meant for traffic which is very delay sensitive, while background is the most delay insensitive class. In order to meet the specified QoS of each service, a scheduling algorithm should be deployed in order to prioritize connections according to their delay requirements. In Chapter 4, we propose a cross-layer scheduler which provides services differentiation and guarantee QoS requirements.

The conversational real-time services, such as video telephony and speech are the most delay sensitive applications. They require data streams to be carried in conversational class. However, the interactive class and background class are mainly used for traditional applications such as www, e-mail, Telnet, FTP, and news. Due to less stringent delay requirements compared to the conversational and streaming classes, both classes provide a better error rate by means of channel coding and retransmission. The main difference between the interactive and background class is that the interactive class is mainly used for interactive applications (e.g., e-mail or interactive web browsing), and the background class is meant for background traffic (e.g., background of e-mails or to background the downloading). Separating interactive and background applications ensure response for the interactive applications and background applications. Traffic in the interactive class has a higher priority in scheduling than the background class traffic. The background applications use transmission resources only when interactive applications do not need them. The required QoS features for UMTS services are summarized at Table 3.2.

Class	Traffic Feature
Conversational	Low delay / loss rate
Streaming	Less sensitive to delay / high data rate
Interactive	Bursty, moderate delay / loss rate
Background	Highly tolerant to delay / loss rate

Table 3-2: QoS requirements for UMTS services.

3.6.2 Technical QoS Categories

QoS can be classified according to its technical implementation in WCDMA systems. This classification can have a hierarchy structure of three different levels: bit level, packet level, and call level [60]. This classification is referred to the technical QoS categories. In Chapter 5 our proposed scheme takes into consideration the following QoS levels:

- Bit-level QoS: To ensure integrity of transmission rate, the target bit error rate (BER) for each user is required. Achieving the required SIR guarantees target BER for that user.

- Packet-level QoS: In the case of real time applications such as voice and video conferencing which are delay sensitive, each packet should be transmitted within a packet delay threshold. On the other hand, data applications are usually tolerant and throughput is the most stringent QoS parameter. The scheduling procedure can achieve the QoS of different services where the transmission parameters are determined according to the QoS class and channel state.

- Call-level QoS: In the cellular WCDMA system, the admission of a call into the system depends on the available capacity and the type of call (new or handover). A call admission control mechanism is essential in order to provide the required call-level QoS.

3.7 HSDPA-WCDMA Evolution

High-speed downlink packet access (HSDPA) and High-speed uplink packet access (HSUPA) are standardized as part of 3GPP Release 5 and Release 6 respectively both of them referred to HSPA 'high-speed packet access'. HSDPA enhances the WCDMA bit rate capabilities by means of fast physical layer retransmission and transmission combining and channel quality information (CQI) reporting schemes, as well as fast link adaptation controlled by the Node B. With HSDPA, two of the most fundamental features of WCDMA, variable SF and fast power control, are disabled and replaced by methods of adaptive modulation and coding (AMC), multi-code operation and a fast retransmission strategy [61]. The expected data rate evolution of HSPA is illustrated in Figure 6.1. The upgrade from WCDMA to HSPA requires new software package and potentially some pieces of hardware in the Node B and in RNC to support high data rate and capacity. The scheduling discipline in HSDPA has a capability of scheduling such that, most of the cell capacity may be allocated to one user for a very short time (2ms), when channel conditions are favourable. HSDPA uses a number of new channels to carry the signaling and the data to UEs. Transport channel carrying the user data with HSDPA operation is denoted as the High-speed Downlink Shared Channel (HS-DSCH). The HS-DSCH of HSDPA operation uses two different modulation schemes QPSK and 16QAM depending on the number of codes which will be allocated to different UEs. Furthermore, HSPA capacity naturally suited for supporting not only symmetric services but also asymmetric services with higher data rates and volumes in downlink. HSPA was initially designed to support high bit rate non-real time services. However, some simulation results have shown that HSPA can also provide attractive capacity for low bit rate-low latency applications like voice over IP (VoIP).

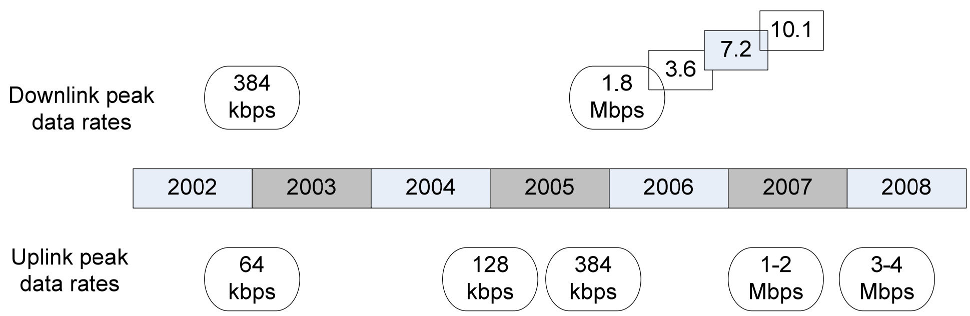

Figure 3-7: Data rate evolution in WCDMA and HSPA.

Chapter 4 - Cross Layer Scheduling Framework for Supporting Bursty Data Applications in WCDMA Networks

4.1 Introduction

In future mobile wireless networks, such as 3G and beyond WCDMA systems, it is expected that data traffic will grow and finally dominate the overall traffic volume. Dedicated channels have a fixed transmission rate and should not be used for bursty data applications since that would result in decreased capacity utilization. On the other hand, shared channels are able to periodically modify their transmission rate and as a result are suitable for the transmission of bursty data.

Furthermore, a scheduling algorithm should also be employed in order to efficiently share the available capacity. As we mentioned previously, for wireline networks, a large number of traffic scheduling schemes, most of them with high computational complexity, have been proposed such as Generalized Processor Sharing [31], Packet-by-Packet GPS (WFQ), or Virtual Clock [29]. However, these algorithms can not be applied to wireless networks due to the error characteristics of the air interface. A wireless traffic scheduler is essential to have some knowledge of the state of the wireless channel during the next frame. If the signal to interference ratio (SIR) of a wireless connection is below a predefined threshold SIR_T then, for as long as this situation remains unchanged, the scheduler should not allocate bandwidth to that connection. Therefore, the idea of making use of information about channel conditions in scheduling strategies is, in fact, a cross-layer technique [2].

Recently, many researchers focused on the cross-layer design in wireless networks. Cross-layer design is necessary in order to improve the performance of a wireless network, efficiently utilize the scarce radio resources and achieve the overall quality of service (QoS) requirement [26,62,63]. For example, in traditional layering architecture, the link layer has a statistical knowledge of the lower physical layer, such as the average wireless channel capacity. However, it is better for the packet scheduler at the link/MAC layer, in order to exploit the time-varying wireless channel, to have knowledge of the channel state at each time instant. Most of the previous works are often considering the optimization of a specific layer without at the same time considering the available information from the other layers. However, in each layer the QoS is defined in different terms the coordination between layers is essential in order to preserve the desired QoS. For example at the physical layer, QoS is synonymous to the acceptable bit error rate (BER) or to the signal to interference ratio (SIR), while at the Medium Access Control (MAC) layer, QoS is usually defined in terms of minimum rate or maximum delay guarantees.

In this Chapter, we introduce a cross-layer framework which aims at making the packet scheduling procedure more efficient. Deploying a finite state Markov chain (FSMC) model, this framework, allows the prediction of the future state of the wireless channel for each connection. Thus, only connections with their predicted SIR above a predefined threshold are allowed to be served. Subsequently, these connections are prioritized and served according to the employed scheduling scheme. The traffic scheduler which we propose is adapted to the cross-layer framework and leads to further enhancement of the system's performance. The principles of the proposed framework can be applied to any wireless network which supports multirate services and QoS differentiation between connections. However we base the present study on the 3GPP specifications for 3G systems.

The rest of this Chapter is organized as follows. In Section 4.2 we describe the system model based on WCDMA downlink transmissions. In Section 4.3 we investigate some related work based scheduling in WCDMA system. In Section 4.4 we present the cross-layer framework and the prediction criterion. Section 4.5 presents the traffic scheduler based the proposed cross layer framework. In Section 4.6 we discuss the simulation model and the numerical results. Finally, Section 4.7 contains some concluding remarks of this Chapter.

4.2 System Model

The forward link for third generation and beyond wireless networks will be a bottleneck since the new services and applications introduced in these networks typically download much more data than they send "in this thesis our discussion focus on the forward link". Therefore, we consider WCDMA system which supports multi-rate transmission by employing Orthogonal Variable Spreading Factor (OVSF) codes. The total capacity of the OVSF code tree at the BS is shared among the users in the cell.

4.2.1 System Model Description

We consider single-cell system architecture as shown in Figure 4.1. The BS provides the required transmission data rate of each user throughout the downlink shared channel (DSCH). Time is partitioned into frames of constant duration, T_s. The BS simultaneously serves, M_u users, each of which has a queue (Q) to receive its incoming packets with data rate, r in every frame, the scheduler makes resource assignment once at every frame interval, T_s. The total transmission power available to the BS is shared among the active users in the cell. The mobile users communicate to BS through wireless link. The BS models the wireless channel of each user by employing a finite state Markov chain (FSMC). The incoming call is admitted to system through a call admission control (CAC). The CAC scheme is based on the available capacity at the OVSF code tree.

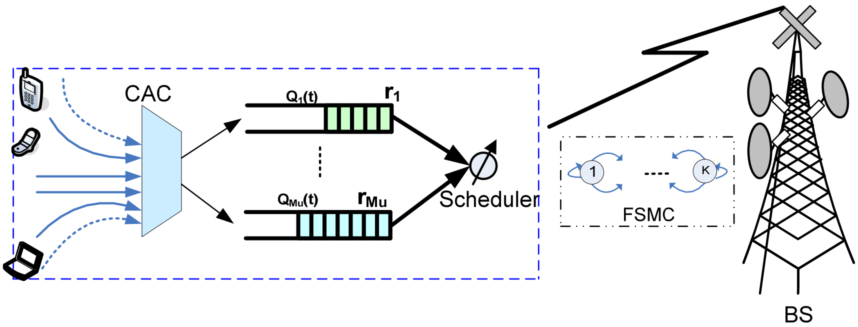

Figure 4-1: System model architecture.

4.2.2 Signal-to-Interference- Ratio

In the downlink, the signal-to-interference-ratio at the receiver can be expressed as [17]:

$$SIR = \frac{G_p \times P_r}{\alpha \cdot P_{\text{intra}} + P_{\text{inter}} + P_N} \tag{4.1}$$

where G_p is the processing gain, P_r is the received power, P_{intra} is the interference generated by other users which are connected to the same cell, P_{inter} is the interference from other cells, α is the orthogonality factor and P_N is the thermal noise power [64]. A fast power control scheme which operates at a frequency of 1.5 kHz [17] adjusts the transmission power devoted to each connection in order to keep the SIR above a target threshold of SIR_T.

4.3 Related Work

Several scheduling schemes have been proposed in the literature for WCDMA systems. At [65] and [66] two credit based scheduling algorithms are proposed. The credit, C, is a priority variable given by:

$$C = total\_time \times GBW - total\_number\_of\_received\_packets\_so\_far$$

GBW is the guaranteed bandwidth for each connection. Connections with more credits are scheduled to receive more packets. This type of credit-based prioritization does not provide low packet delays and QoS differentiation. At [67] the proposed credit-based scheme provides QoS differentiation between users by introducing a variation of the credit variable, the normalized credit. The normalized credit is defined as the ratio C_i / a_i where a_i is the guaranteed rate of user *i*. We can say that the main objective of the credit based schemes is to provide a guaranteed service rate to each user. However, they do not take into account the delay sensitivity of the services and therefore they cannot prioritize the connections according to their delay requirements.

At [68], a class of schemes based on the GPS approach [31] under the generic name Code Division GPS (CDGPS) is proposed to support multimedia traffic in WCDMA systems. The basic principle of CDGPS schemes is to assign a fixed positive real number (weight) ϕ_i to each flow, and dynamically allocate bandwidth for all flows according to their weights and the traffic load. The remaining capacity is distributed proportionally to users who expect more than their guaranteed service rate. The CDGPS scheme provides weighted fairness to heterogeneous traffic, as well as a delay bound for each flow that is properly shaped. A credit based CDGPS scheme is also proposed at [68] to further improve the utilization of the soft capacity. The main drawback of CDGPS is that it requires extensive computations, as it is a GPS based scheme.

Finally, the scheduling scheme presented at [69] is able to dynamically assign priority to each connection providing QoS differentiation among different service classes while takes into account the spectral efficiency for each user. The disadvantage of this scheme is that the employed controlling parameters at the calculation of the priority are predefined for each traffic class. However, within each traffic class the services do not have exactly the same characteristics and QoS requirements (e.g. Web browsing and network computer games in Interactive class).

4.4 The Cross-Layer Framework

The proposed cross layer architecture is shown in Figure 4.2. At the MAC layer the SIR information gathered from the physical layer is utilized by the packet scheduling procedure. At the start of each frame we predict the error probability of each connection during the next frame. The prediction is based on SIR measurements from the physical layer during the previous frame. Subsequently, the connections which have predicted bit error rate (BER) below a predefined threshold are excluded from service and the rest of the connections are served by the packet scheduler.

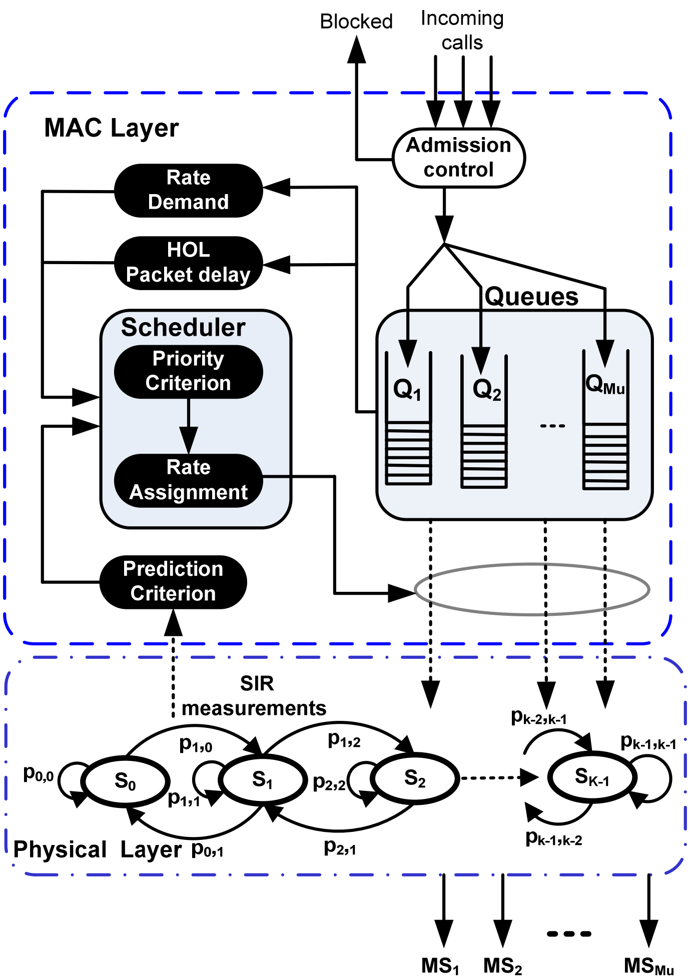

Figure 4-2: The cross layer framework.

4.4.1 The Prediction Criterion

Each connection i has its own QoS requirements in Bit Error Rate and therefore a BER threshold, BER_{T_i}, is defined. The BER is related to the SIR of the received signal at the mobile. For each connection i two SIR thresholds are also defined: $SIR_{\min_i}$ and SIR_{T_i}. The SIR_i at the receiver is measured at each frame and according to its value we can distinguish three cases:

1) If $SIR_i \geq SIR_{T_i}$ then $BER_i \leq BER_{T_i}$: the BER requirement is satisfied and the user will be subsequently served at the current frame according to the employed scheduling discipline.

2) If $SIR_i < SIR_{\min_i}$ then $BER_i > BER_{T_i}$: the BER requirement is not satisfied and the user is excluded from service for the current frame.

3) Finally the last case occurs if $SIR_{\min_i} \leq SIR_i < SIR_{T_i}$. This case needs further examinations with the prediction criterion in order to determine if the required BER can be achieved.

For the last case a Finite State Markov Chain (FSMC) is employed. The FSMC model, shown in Figure 4.3, represents the time varying behavior of the Rayleigh fading channel. By using this model, which is described in detail in the following section, we are able to make a more accurate prediction of the future state of the wireless channel. Based on this prediction we decide which users will be served by the packet scheduler.

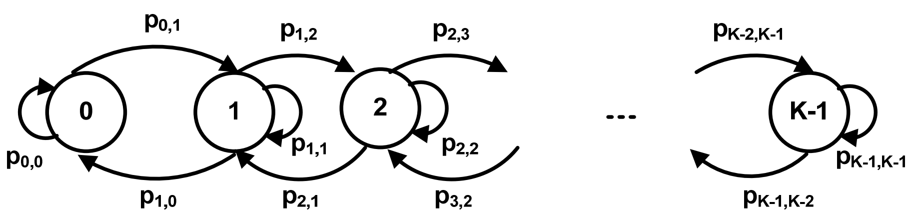

Figure 4-3: FSMC for wireless Rayleigh fading channel.

The prediction criterion consists of three parts as follows:

a) At the first part we define the parameters of the FSMC model of the user according to the SIR measurements at the previous frame.

b) Subsequently, we define the most probable state of the wireless channel for the next frame.

c) Finally, for the predicted state of the wireless channel we calculate the respective error probability and we decide if the user should be served by the scheduler according to his QoS requirements.

4.4.1.1 Parameters of the FSMC model

The parameters of FSMC model are obtained by partitioning the received SIR_i of user *i* into *K* states. Let ψ_i be the instantaneously received SIR_i of user *i* in the multipath fading environment, which is proportional to the square of the signal envelop. The probability density function (*pdf*) of ψ_i can be expressed as in [70].

$$p(\psi_i) = \frac{1}{\psi_0} e^{-\psi_i/\psi_0}, \qquad \psi_i \geq 0 \tag{4.2}$$

where ψ_i is the instantaneous received SIR_i of user *i* in the multipath fading environment and $\psi_0 = \mathrm{E}\{\psi_i\}$ is mean value by averaging the SIR of all active users in the cell as follows:

$$\psi_0 = \frac{\sum_{j \in M_u} \psi_j}{M_u} \tag{4.3}$$

where M_u is the number of the active users in the cell.

In order to construct the prediction criterion for the wireless channel by using the FSMC model, we assume that the channel is slow fading and thus the transition probability between in-adjacent states is very small. By partitioning the range of the received SIR into a finite number of intervals, a finite state channel model for Rayleigh fading channel can be built as shown in Figure 4.3. Suppose that $\Psi_0 = 0 < \Psi_1 < \Psi_2 < \cdots < \Psi_{K-1} = \infty$ are the thresholds of the received ψ_i. The channel is considered to be in state *k* if the ψ_i lies in the interval $\{\Psi_k, \Psi_{k+1}\}$. Associated with each state there is a Binary Symmetric Channel (BSC) with error probability $P_{em}(k)$, which depends on the modulation scheme. Assuming that the channel fades slowly with respect to the frame period T_s, and the maximum Doppler frequency is $f_d = v.f_c / c$, the Markov steady state probabilities of state *k* are [71]:

$$\pi_k = \int_{\Psi_k}^{\Psi_{k+1}} p(\psi)\, d\psi = e^{-\frac{\Psi_k}{\psi_0}} - e^{-\frac{\Psi_{k+1}}{\psi_0}} \tag{4.4}$$

The choice of the SIR thresholds is flexible and the partitioning can be done in many ways. Equal partitioning method considered the inter-partitioning gap between the thresholds interval is the same, this method is very simple and with low complexity design. However, it is considered as an inaccurate modeling approach. Another method of Lloyed-Max quantizer, which uses the quantization thresholds of the optimum Minimum Mean-Square Error (MMSE) since the SIR is considered to be a random variable follows the exponential distribution. However, this method does not take into account the dynamics of fading process [71]. Finally, Equal Probability Method (EPM) [72] is used such that; given the number of states (*K*) for the FSMC model, the SIR thresholds can be determined as follows:

$$\pi_0 = \pi_1 = \cdots = \pi_{K-1} = \left[\frac{1}{K}\right] \tag{4.5}$$

We chose this method to determine the number of the SIR thresholds which is computationally simple and fairly accurate for our simulation model. On the other hand, the increasing of the number of states above four states is not achieved high improvement in system performance and the modeling accuracy achieved by increasing the number of states beyond four is negligible as investigated at [73]. Therefore, in our simulation model the BS considers the Markov channel of each user with four states Markov chain model. Thus, the SIR thresholds are $\Psi_0 = 0, \Psi_1, \Psi_2, \Psi_3, \text{and}\, \Psi_4 = \infty$ can be determined by using the EPM method. At BS, based on the SIR measurements and the thresholds estimation, when the received SIR is obtained, it will be compared with the thresholds levels (Ψ_1, Ψ_2 and Ψ_3) to determine the present state as shown at Figure 4.4. Then from the Markov channel model (transition and steady state probabilities), the state of the channel during the next frame transmission is predicted and sent to the scheduler in order to prioritize the users based on their channel state.

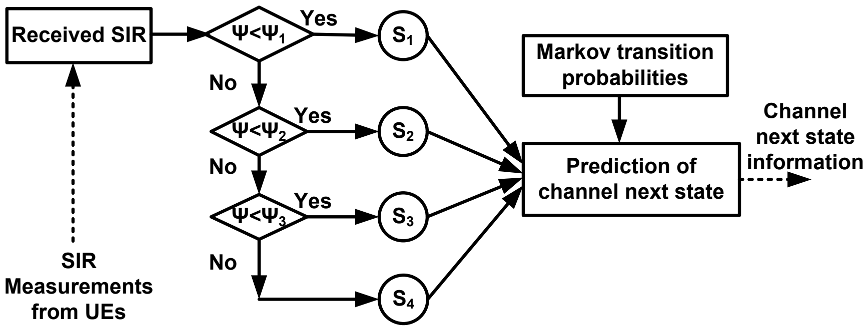

Figure 4-4: Parameters of FSMC model.

4.4.1.2 Prediction of the Next State

According to the random distribution of user mobility which provides the maximum Doppler frequency, f_d, we make use of the level crossing rate in order to determine Markov transition probabilities. The level crossing rate of a predefined threshold (Ψ) for SIR levels in only one direction (positive or negative) can be expressed as follows [72],

$$\mathrm{N}(\Psi) = \sqrt{\frac{2\pi\Psi}{\psi_0}} f_d \, e^{-\frac{\Psi}{\psi_0}} \tag{4.6}$$

The rate assignment in the WCDMA systems can be changed on frame by frame basis. Therefore the transition probability between the adjacent states can be written as the following equations:

$$p_{k,k+1} = \frac{\mathrm{N}(\Psi_{k+1})T_s}{\pi_k} \quad \text{and} \quad p_{k,k-1} = \frac{\mathrm{N}(\Psi_k)T_s}{\pi_k} \tag{4.7}$$

Therefore, the Markov transition probabilities can be approximated as follows:

$$p_{k,k+1} \cong \frac{1}{\pi_k} f_d T_s \sqrt{\frac{2\pi\,\Psi_{k+1}}{\psi_0}} \, e^{-\frac{\Psi_{k+1}}{\psi_0}} \tag{4.8}$$

$$p_{k,k-1} \cong \frac{1}{\pi_k} f_d T_s \sqrt{\frac{2\pi \Psi_k}{\psi_0}} e^{-\frac{\Psi_k}{\psi_0}} \tag{4.9}$$

$$p_{k,k} = 1 - (p_{k,k+1} + p_{k,k-1}) \tag{4.10}$$

In practical wireless system, the probability of error is changed on frame by frame basis. The transition probability between in-adjacent states is very small, so it can be assumed that transitions happen only between adjacent states. So the wireless channel state is one step prediction. Therefore, the prediction procedure is rather simple and is based on the previous channel state information, namely the state *k* where the channel was found, and the Markov transition probabilities of the FSMC channel model. Thus, given the current state, the next state *m* is the one in which the Markov chain will transit to with the highest state transition probability. In other words:

$$p_{k,m} = \max(p_{k,k}, p_{k,k+1}, p_{k,k-1}) \tag{4.11}$$

The state transition probabilities can be computed over a scheduling interval and updated periodically at the base station based on SIR levels received from multiple mobile users.

4.4.1.3 Error Probability

The error probability $P_{em}(k)$ at each state of the FSMC model is a function of SIR for each user *i* and it can be expressed as [71] :

$$P_{em}(k) = \frac{1}{\pi_k} \int_{\Psi_k}^{\Psi_{k+1}} p(\psi_i)\, p_m(\psi_i) d\psi_i \tag{4.12}$$

where $p_m(\psi_i)$ for WCDMA systems with QPSK modulation under additive Gaussian noise is given by [70]:

$$p_m(\psi_i) = 0.5 erfc(\sqrt{\psi_i}) \tag{4.13}$$

Finally, connections are served by the packet scheduler only if their predicted error probability $P_{em}(k)$ is less than or equal to their respective error probability threshold.

4.5 The Traffic Scheduler

The users which pass the prediction criterion are served by the employed scheduling discipline. We study the performance of the cross-layer framework together with two scheduling schemes, namely delay fair scheduler (DFS) and delay fair scheduler with prediction (DFS_PRED).

4.5.1 The Delay Fair Scheduler (DFS)

The DFS [74], scheme has low computational complexity and provides fair distribution of the available shared capacity to the connections according to their delay requirements. According to DFS each connection *i* at the start of each frame *n*, is characterized by its priority, P_i:

$$P_i = \frac{D_i(nT_s)}{T_i} \geq 0 \quad n = 0, 1, 2, \cdots \tag{4.14}$$

where T_i is a threshold for the acceptable data packets delays, defined during connection setup, D_i is the head-of-line packet delay for queue *i* and T_s=10ms is the scheduling period of the DFS scheme. Subsequently, the connections are sorted and served in decreasing order of their priorities.

4.5.2 The Delay Fair Scheduler with prediction

The main idea of delay fair scheduler with prediction (DFS_PRED) is to prioritize the connections not only according to their delay requirements but also according to their predicted error probability during the next frame. As with DFS each connection *i* at the start of each frame *n*, is characterized by its priority P_i:

$$P_i = \frac{D_i(nT_s)}{T_i}(1 - P_{em}(i)) \geq 0 \quad n = 0, 1, 2, \cdots \tag{4.15}$$

However, at this case the priority of a connection is proportional to its probability of successful transmission $\left(1 - P_{em}(i)\right)$. Consequently, between two connections with the same head-of-line packet delay and delay threshold, the one with the higher probability for successful transmission will also have higher priority. Thus, DFS_PRED is able to encounter the variable capacity of the wireless interface better than DFS.

DFS_PRED is performed in two steps:

1) At the first step, P_i is recalculated for all the connections which have non empty queues. P_i is used as a criterion to determine if the service of a connection should take priority over another. Then the connections are sorted in decreasing order of their priorities.

2) In the second step, based on the OVSF channelization codes, DFS_PRED allocates the available bandwidth to the connections in the sorted list beginning from the higher priority connections. The rate assignment process concludes:

    i. When all the connections of the list are capable of transmitting all their respective queued packets during the next frame or

    ii. When all the available capacity has been allocated.

4.6 Numerical Results and Discussion

The performance of the proposed cross layer framework and scheduling schemes is evaluated in terms of average packet delay, average queue length and channel utilization.

4.6.1 Simulation Model

The performance of the proposed cross layer framework and scheduling schemes is evaluated through event driven simulation on a 7-layer OVSF code tree and assuming a cell with a radius of 2000m and a BS located at the center. The initial location of each user is randomly distributed within the cell while the velocity, v, of each user is uniformly distributed in the interval {0, 3}km/h. The call arrival process is modeled by a Poisson distribution, while the call duration is exponentially distributed with equal mean. The traffic load is increased by increasing the average arrival rate (call arrivals/s) while the average call duration remains constant. Finally, all the connections are assumed to be served by DSCH channels.

The macrocell propagation model proposed in [64] is adopted for path loss. The attenuation L_p of the transmitted signal for a BS antenna height of 15m and a 2000MHz carrier frequency is defined by the following equation:

$$L_p(d_i) = 128.1 + 37.6\log(d_i) \tag{4.16}$$

where d_i is the distance between the BS and the mobile in km. For all services the target SIR_T is equal to 7.9dB while the $SIR_{\min}$ is equal to 3dB. The power assignment to the active users is based on 3GPP standards [75]. The maximum transmission power of a downlink traffic channel is 30dBm and the total transmission power of the BS is 43dBm. Assuming that the cell is isolated and therefore there is no interference from other cells, therefore *SIR* can be calculated by Equation (4.1). The FSMC model used for the partitioning of the wireless channel has four states. The equal probability method (EPM) is used to determine the steady state probabilities and thereby the transition probabilities.

4.6.2 Bursty Traffic Model

For each connection the traffic is assumed to arrive according to an "ON-OFF" model. As long as the connection is in the "OFF" state, it has no arrivals. While in the "ON" state a batch of *N* packets arrives per timeslot. *N* is uniformly distributed between N_L and N_H, where N_L, $N_H \in R$ are positive real numbers. A packet is defined as the amount of bits that can be received during one timeslot at the lowest available rate R. During connection set-up, the mobile user negotiates with the network management module for a guaranteed delay threshold T. The probability P_{on} of being in the ON state, as well as the N_L, N_H and a packet delay threshold T are predefined for each connection. Therefore, the mean service rate, R_S, for each connection can be determined in terms of the N_L, N_H and P_{on} as follows:

$$R_S = \frac{N_H + N_L}{2} \times P_{on} \tag{4.17}$$

We use a simple call admission control (CAC) scheme where an incoming call i which requests a mean rate of $R_{S,i}$ is accepted as long as

$$\sum_{i=1}^{M_u} R_{S,i} \leq C \tag{4.18}$$

where C is the available capacity and M_u is the current active number of users in the cell. A relative system load L is used to represent how heavy the system is loaded. $L=1$ implies that the maximum number of connections is reached, according to Equation (4.18).

4.6.3 The Efficiency of the Cross Layer Framework

In this section, we study the effect of the proposed cross-layer framework at the packet scheduling procedure. Thus, we compare the performance of the DFS scheduler with and without the proposed framework and we denote the two schemes by DFS_CL and DFS respectively.

We assume a simulation scenario where all the connections have the same traffic characteristics and QoS requirements. Specifically, the delay threshold is set to T = 1s, the target BER is set to $BER_T = 10^{-4}$ while, $N_H = 6$, $N_L = 4$ and $P_{on} = 0.8$.

As shown at Figure 4.5 DFS_CL outperforms DFS in terms of average packet delay at all traffic loads. DFS_CL through the use of the cross layer framework is able to serve connections which have SIR in the interval $SIR_{\min} \leq SIR < SIR_T$ while in contrast DFS serves only the connections with $SIR \geq SIR_T$. Therefore, the use of the cross layer framework increases the number of users that can be supported with the required QoS.

At Figure 4.6 the average queue lengths for both DFS_CL and DFS schemes are shown. At each frame DFS_CL utilizes the cross layer framework and thus serves more connections per frame than DFS. Consequently, DFS_CL achieves decreased average queue length at all traffic loads. Finally the efficiency of the framework can also be verified at Figure 4.7 where the bandwidth utilization for DFS_CL and DFS schemes is shown respectively. DFS_CL is able to serve connections which would be excluded from service by DFS. Thus DFS_CL utilizes the wireless bandwidth better than DFS at all traffic loads.

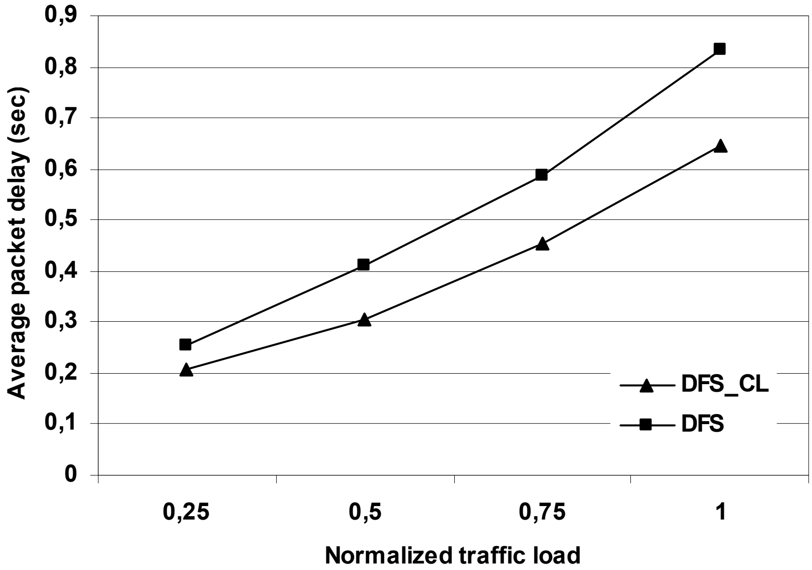

Figure 4-5: Average packet delay comparison for DFS_CL and DFS schemes.

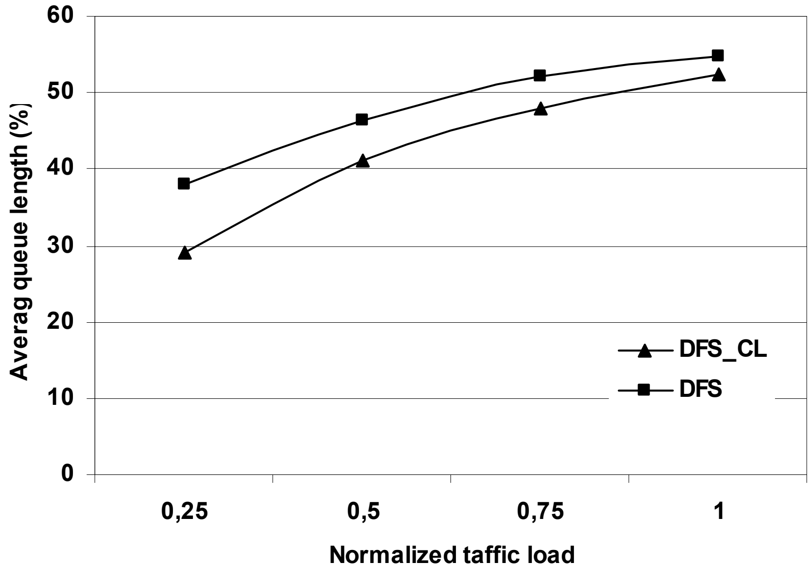

Figure 4-6: Average queue length comparison for DFS_CL and DFS schemes.

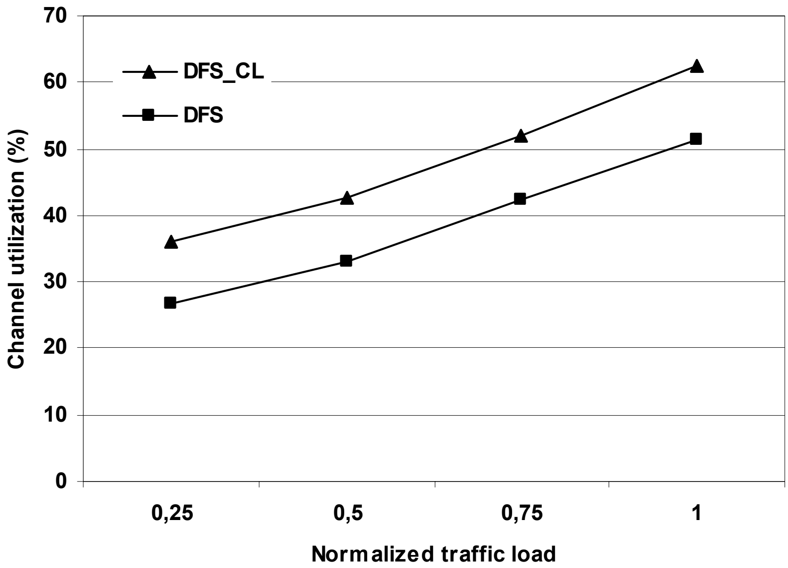

Figure 4-7: Bandwidth utilization for DFS_CL and DFS schemes.

4.6.4 Performance Evaluation of DFS_PRED

In this simulation scenario we evaluate the performance of DFS and DFS_PRED under the cross layer framework. Therefore we will denote in the following the two schemes as DFS_CL and DFS_PRED_CL respectively. We assume that all mobile connections have the same QoS requirements in terms of delay threshold, the delay threshold is set to $T=$ 1s, $BER_T = 10^{-4}$ while, $N_H = 6$, $N_L = 4$ and $P_{on} = 0.8$.

The priority criterion of DFS_PRED is proportional to the successful transmission probability of each connection. Consequently, DFS_PRED does not only consider the packet delay and the respective delay threshold of each connection, as DFS does, but also takes into account the state of the wireless channel. Thus, when DFS_PRED is used, we have more successful transmissions and therefore better channel utilization is achieved as it is shown at Figure 4.8.

Furthermore, the increased channel utilization of DFS_PRED leads to lower average packet delay compared to the DFS scheme. Figure 4.9 shows the average packet delay for DFS and DFS_PRED schemes. As we can see DFS_PRED outperforms DFS at all traffic loads. Consequently, DFS_PRED increases the number of users that can be supported with the required QoS.

On the other hand, by employing DFS_PRED the average queue length at the BS is significantly reduced as shown at Figure 4.10. This is due to the DFS_PRED scheme efficiently has the capability of assigning the required bandwidth for the users with higher probability of successful transmission compared to DFS scheme. In fact, this will lead to reducing the number of the packets which waiting at the BS for transmission related to the users experienced bad channel quality.

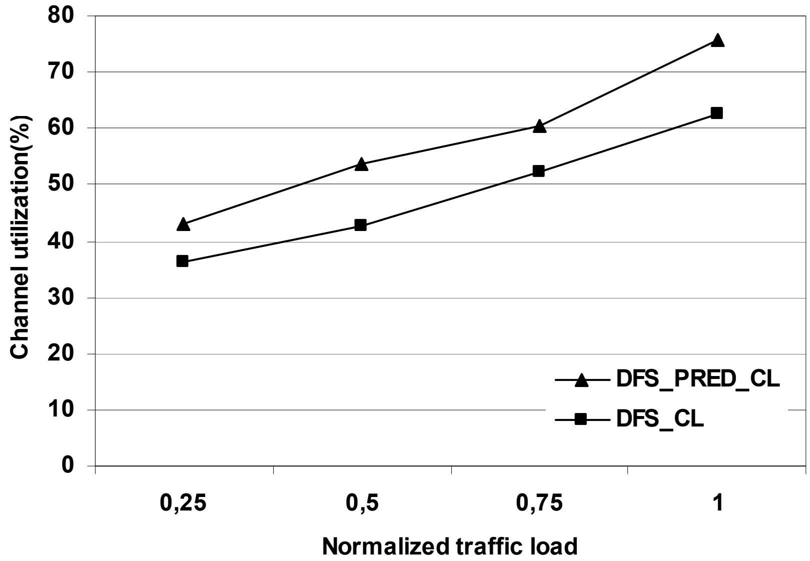

Figure 4-8: Bandwidth utilization comparison for DFS_PRED_CL and DFS_CL schemes.

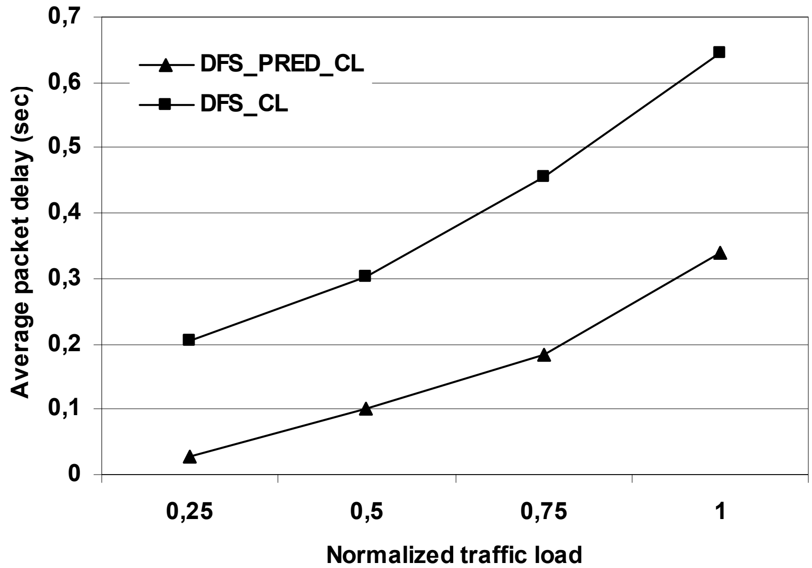

Figure 4-9: Average packet delay comparison for DFS_PRED_CL and DFS_CL schemes.

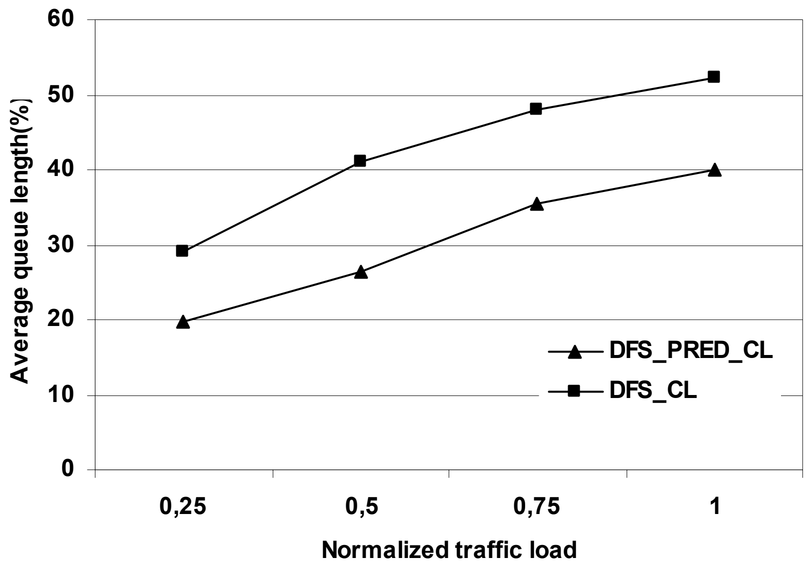

Figure 4-10: Average queue length comparison for DFS_PRED_CL and DFS_CL schemes.

4.6.5 Performance Evaluation of Wireless Scheduling Characteristics

In this simulation scenario we want to measure the major scheduling characteristics of the proposed scheduler based cross-layer technique such as: fairness and the QoS differentiation. According to QoS classes in 3G systems, we assume that four services with different QoS requirements and then we measure the fairness and the QoS differentiation properties in terms of average packet delay among the users with different type of traffic service.

4.6.5.1 Fairness Property

In this study scenario we want to observe how services with different traffic characteristics but with the same delay sensitivity are served by the proposed scheme. We assume four services with the same delay threshold while the traffic characteristics are different. Table 4.1 illustrates the four services characteristics. Figure 4.11 shows the fairness property of DFS_CL is not affected by the proposed cross layer framework. As it is illustrated in this figure the four services approximately have the same average packet delay despite of their different traffic features.

No # Services	P_{on}	N_H	N_L	T(sec)
Serv1	0.8	6	4	1
Serv2	0.5	12	4	1
Serv3	0.34	16	8	1
Serv4	0.2	24	16	1

Table 4-1: Four services with different traffic characteristics: Fairness Property.

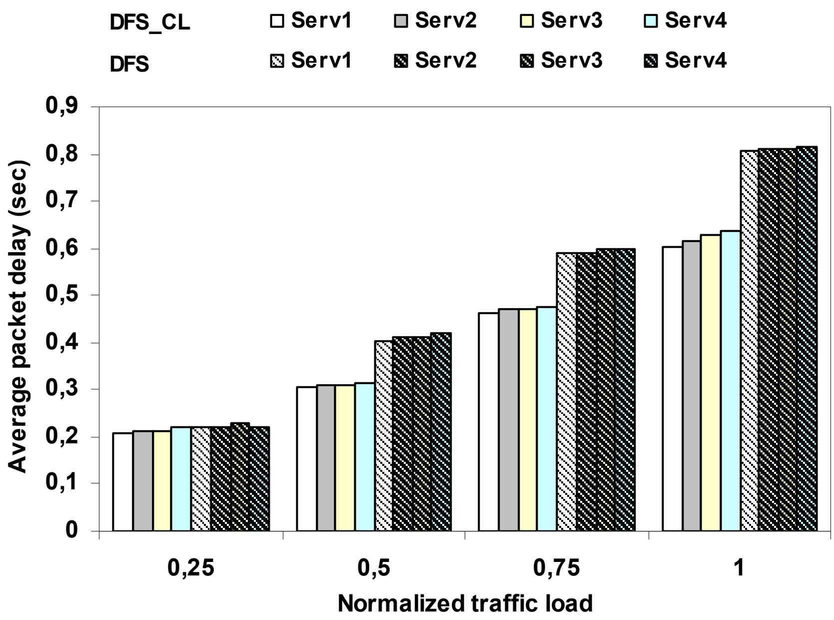

Figure 4-11: Fairness: Average delay per service for DFS_CL and DFS schemes.

4.6.5.2 QoS differentiation

In this scenario we assume the four services with the same traffic characteristics but with different delay requirements. The traffic features and delay sensitivities of the four services are illustrated in Table 4.2. Figure 4.12 shows the QoS differentiation of the DFS and DFS_CL disciplines. As we can see in this figure, although all the services have the same traffic characteristics, the packet scheduler (DFS_CL or DFS) takes into account the respective delay thresholds and gives higher service priority to the most delay sensitive connections.

No # Services	P_{on}	N_H	N_L	T(sec)
Serv1	0.5	12	4	1s
Serv2	0.5	12	4	0.6s
Serv3	0.5	12	4	0.4s
Serv4	0.5	12	4	0.2s

Table 4-2: Four services with different delay requirements: QoS differentiation.

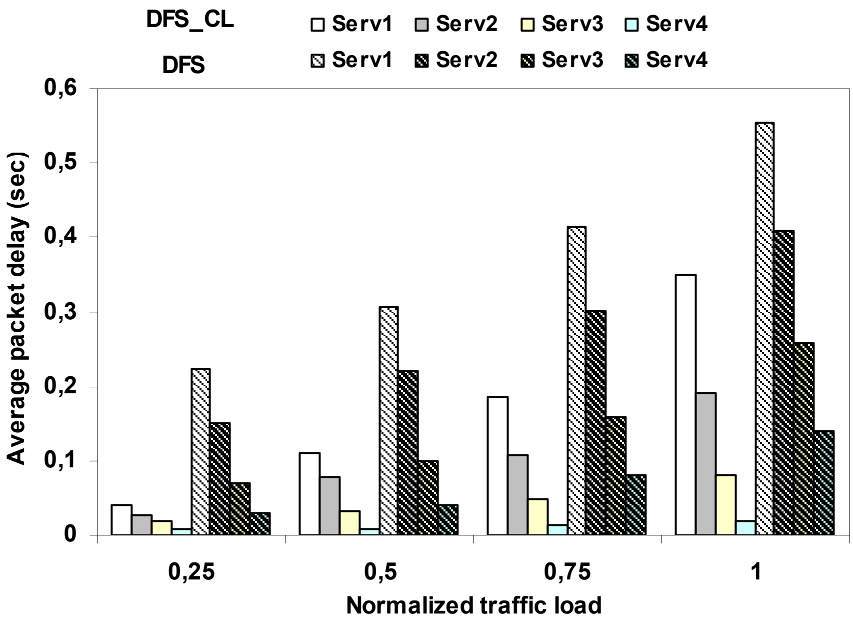

Figure 4-12: QoS differentiation: average delay per service for DFS_CL and DFS schemes.

4.7 Concluding Remarks

In this Chapter we have proposed a cross-layer framework which is able to accurately identify the connections which can be served with the required QoS. In particular with the introduction of the cross-layer framework smaller average packet delays, shorter queue lengths and increased bandwidth utilization are achieved as indicated in our results. Therefore, the proposed framework significantly improves the efficiency of the packet scheduling procedure in WCDMA based networks. However, the principles of the proposed framework can be applied to any wireless network which supports multirate services and QoS differentiation between connections.

Furthermore, we have proposed a new scheduling algorithm, which serves the connections not only according to their delay sensitivity, but also according to the predicted state of their wireless channel. The simulation results demonstrate the efficiency of the proposed scheme in terms of average packet delay, average queue length and bandwidth utilization**.** Also, we measure the major issues for wireless scheduling algorithms such as fairness and QoS differentiation.

Our results indicate that, in terms of both bandwidth utilization and average packet delay, a significant improvement is possible for any wireless network deploying the proposed cross-layer framework. Such technique of the cross-layer framework is considered a first step in WCDMA system towards deploying cross-layer design in the future mobile wireless networks. Finally, we have shown in our results, the major issues of wireless scheduling strategies such as fairness and QoS differentiation could not effect by deploying the proposed cross layer framework.

Chapter 5 - Radio Resource Management for Handover Provisioning in 3G WCDMA Networks

5.1 Introduction

One major challenge in multimedia services over third generation mobile communication networks is QoS provisioning with efficient resource utilization. In such a network, the cross layer design principle should be applied to capture the call-level and packet-level QoS performance [76]. Therefore, efficient radio resource management based on cross-layer technique is necessary to support the provisioning of 3G mobile multimedia services. Call admission control (CAC) mechanism at the core of the radio resource management strategies is not only the responsible for assuring the prescribed QoS in the cellular WCDMA networks but also the traffic scheduler provides and supports QoS provisioning. Therefore, the major components of the radio resource management are scheduling discipline and call admission control mechanism. One of the most important functions of a CAC in 3G WCDMA networks is code management and handover management procedures [66]. Code management is devoted to manage the orthogonal variable spreading factor (OVSF) codes tree among different mobile connections. Handover management is the mechanism that allowing the mobile users communicated with the network and still guaranteeing their QoS. On the other hand, the traffic scheduler at the radio resource management strategies should be QoS-aware in order to be able to meet the QoS requirements of the handover calls.

In fact, the mission of this Chapter is two folded. In the first Part, we propose a guard code scheme for HO traffic management in WCDMA system employing OVSF codes. This scheme prioritizes HO calls over new ones and thus improves HO failure rate due to capacity shortage compared to the new call blocking rate. The scheme introduces a code reservation threshold, which allows HO calls to use the total capacity of the OVSF code tree, but restricts the new calls to use only a part of the tree capacity. We model the code occupancy of the OVSF code tree by a Markov chain, where the arrival rates at each state depend on the existing code reservation threshold, and numerically solve the flow balance equations and compute the steady state distribution of certain problem instances. We study the behavior of the system in terms of the famous QoS performance metrics such as new call blocking, HO failure probability and code utilization under different traffic loads and different code reservation thresholds.

In the second Part, we introduce a new effective combination between the a CAC mechanism and a traffic scheduler algorithm for efficient radio resource management strategy. These two schemes are designed for WCDMA systems and operate in a complementary fashion in order to support handover provisioning. The CAC mechanism belongs to the well known family of guard channel schemes and reserves some code capacity to favor the continuation of handover (HO) calls over the new calls. However, during the handover procedure the HO calls experience high delays as a number of packets have to be forwarded through the wireline infrastructure to the target cell. Thus, we employ a delay driven traffic scheduler (DDS) which aims to prioritize calls which experience high delays such as HO calls. Furthermore, DDS based on cross-layer technique, is able to exploit information from the physical layer in order to avoid erroneous packets transmissions and increases system performance.

The rest of this Chapter is organized as follows. In Section 5.2 we investigate the handover procedures such as soft and hard handover in 3G WCDMA networks. Then in Section 5.3, Part 1 introduces the analytical model of the guard code scheme, we build our system model throughout some basic concepts in WCDMA system such as analyzing OVSF code tree, code capacity, code blocking concepts and some related work. Then we present the performance analysis of the guard code scheme in Subsection 5.3.6, we study the system state space by using Markov chain model and then we measure the performance metrics such as new call blocking and handover failure probabilities. Finally Subsection 5.3.6 presents the numerical results and performances evaluation for instants problems.

Part 2 at Section 5.4 presents the combination between the CAC mechanism and the traffic scheduling algorithm. In Subsection 5.4.1 we describe the system model. Then we calculate the code reservation and the adaptation of reservation capacity to the OVSF code tree in Subsection 5.4.2. The avoiding of erroneous packet transmission and the prioritization and rate allocation investigate the proposed traffic scheduling algorithm in Subsection 5.4.3. Then Section 5.5 presents some related work in QoS provisioning schemes. Section 5.6 discusses the simulation model and the numerical results. Finally, in Section 5.7 we give some concluding remarks of that Chapter.

5.2 Handover procedures in 3G WCDMA Networks

In 3G networks based on WCDMA systems, OVSF codes are chosen as the channelization codes to separate streams of data among the mobile users. OVSF codes allow WCDMA systems to dynamically allocate a variable transmission rate to each user or connection by assigning one (single-code operation) or multiple codes (multi-code operation) with different spreading factors. However, this rate allocation dynamically changed due to the link quality and the interference level which caused by the mobility of users [17,77]. Therefore, sometimes some particular user changes its serving base station. This process is known as handover.

In such networks both soft and hard handovers are possible depending on the type of channel used by the application. Conversational applications use dedicated channels where soft handover is possible, while streaming interactive and background services use the shared channel where only hard handover is possible. Hard handover is the handover procedure in which all the old radio links of a mobile are released before the new radio links are established. Therefore, a hard handover failure will lead to a temporary connection outage and additional delay induced in the packets waiting to be transmitted in the new cell. In soft handover procedure a mobile simultaneously communicates with two or more cells belonging to different BSs. As a result, soft handover ensures transparent user communications and better communication quality during handover. Hence, a soft handover failure is also undesirable due to the signaling overhead induced during the re-establishment of the connection. As both cases lead to poor network performance and should be avoided, it is important to favor HO requests over new call request during the OVSF code allocation process [78]. A simplified scenario of soft and hard handover procedures in 3G WCDMA networks is shown in Figure 5.1. In the following sections we propose a guard code scheme which favors the hard handover calls which are supported by the DSCH channels.

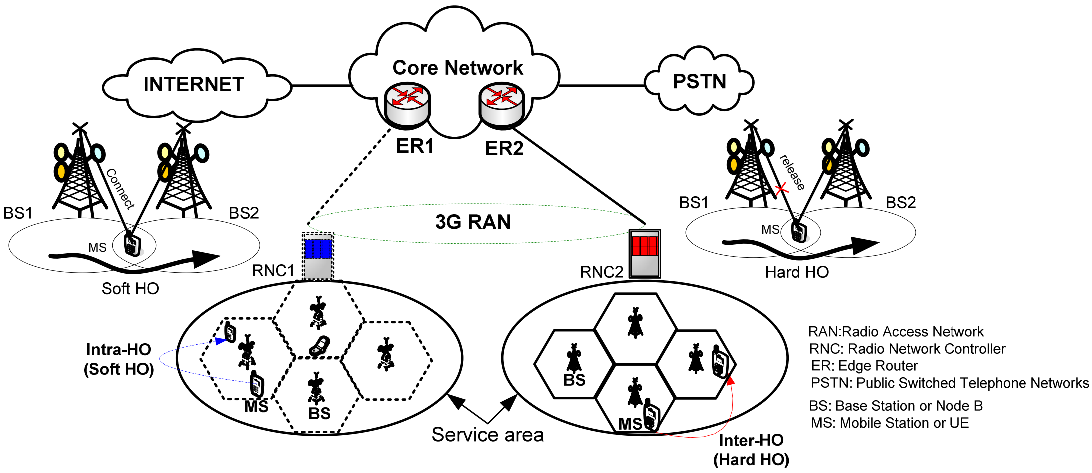

Figure 5-1: A Simplified scenario of handover procedures.

5.3 Part I: A Guard Code Scheme for Handover Traffic Management in WCDMA Systems

In this Part, we propose a prioritization of handover (HO) calls over new calls in WCDMA systems employing orthogonal variable spreading factor (OVSF) codes as channelization codes. The code occupancy of the system is modeled by a Markov chain and the differentiation between HO and new call is performed at the code level by introducing a "guard code" scheme. The scheme belongs to the well known family of guard channel schemes and reserves some code capacity to favor the continuation of HO calls over the new calls. As the management of the general case is intractable, we solve certain numerical instances of the problem and manage to calculate several popular performance metrics like new call blocking and HO failure probabilities and code utilization.

5.3.1 Basic Concepts

In WCDMA systems, such as 3G systems, OVSF codes are employed as channelization codes in order to preserve the orthogonality among different physical channels [17]. At the transmitter, each bit of a traffic flow is multiplied with an OVSF code. Thus, the spectrum of each signal is on purpose widened in order to occupy all the available bandwidth. As a result, the transmitted signal has lower power density and a noise-like spectrum. At the receiver the despreading is performed by multiplying with the same OVSF code sequence initially used at the spreading procedure.

5.3.2 Analysis of OVSF Codes

The method that is used for the generation of the OVSF codes is described in detail in [50], [79]. OVSF codes are generated by a binary tree with L layers. Each node represents a channelization code $C_{SF,i}$, where SF is the spreading factor of the code and i is the code number, $1 \leq i \leq SF$. The higher the spreading factor the lower the transmission rate supported by a code. The spreading factors are doubled between consecutive layers, as we move from the higher (root) layer to the lower (leaf) layer. Leaf codes have the maximum spreading factor ($SF_{\max}$) and therefore the minimum data rate, which is denoted by R. The transmission rate supported by an OVSF code with spreading factor SF is always a multiple of a power of two of the lowest available rate.

A difficulty in the assignment of an OVSF code is the orthogonality constraint. According to this constraint the assignment of a code is possible, if only if none of its ancestor codes and none of its descendant are already occupied. Once a code is assigned, all of its ancestors, as well as all of its descendants are blocked and can be used after the code is released. Hence, at the 4-layer OVSF code tree shown at Figure 5.2, the assignment of code $C_{4,1}$ immediately blocks its ancestor codes $C_{2,1}$, $C_{1,1}$, and its descendant codes $C_{8,1}$, and $C_{8,2}$. Therefore, during the code assignment procedure in OVSF code tree we should take into account not only code blocking but also capacity blocking constraints.

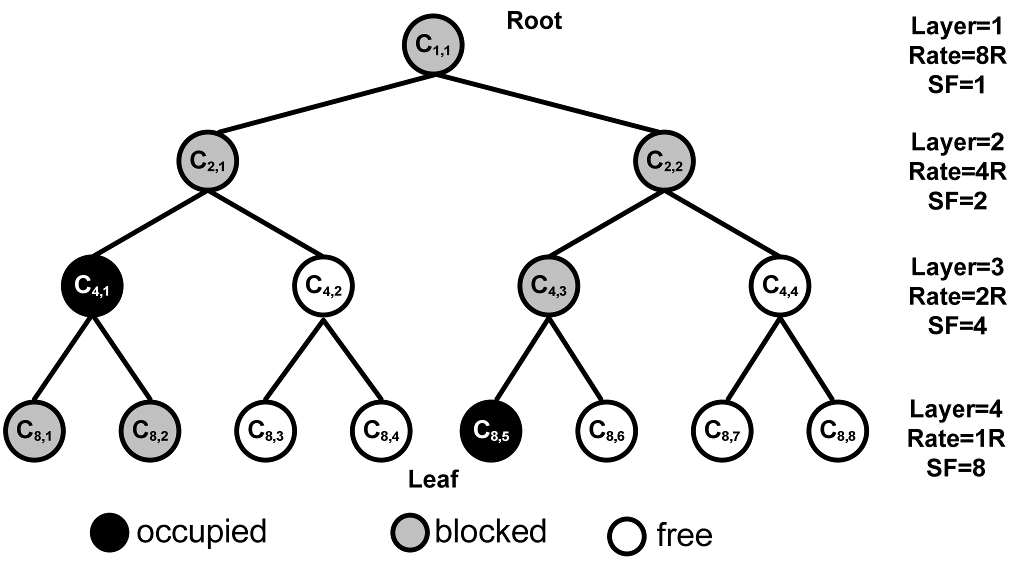

Figure 5-2: The OVSF Code tree.

5.3.3 Code Blocking and Capacity Blocking

The blocking of an incoming call can be characterized either as capacity blocking or as code blocking. If the remaining OVSF code tree capacity is lower than the rate requirement of an incoming call then the latter will be blocked and we will refer to this case as capacity blocking. For example at the OVSF code tree of Figure 5.2 the available capacity is $5R$ since the free codes $C_{8,6}$, $C_{4,2}$ and $C_{4,4}$ support transmission rates of $1R$, $2R$ and $2R$ respectively. Therefore, if an incoming call requests a rate of $8R$ then the call will be blocked due to the lack of capacity.

However, capacity shortage is not the only cause of call blocking. As a result of the statistical nature of the arrival and departure processes the occupied OVSF codes are randomly spread across the code tree during system operation. Consequently, the capacity of the code tree becomes fragmented and that in turn causes code blocking. Code blocking is defined as the condition that a new call cannot be supported although the system has enough capacity to support the rate requirement of the call [80]. For example, as we previously mentioned, the available capacity of the OVSF code tree at Figure 5.2 is $5R$. However, if an incoming call requests a rate of $4R$ will be blocked as the codes $C_{2,1}$ and $C_{2,2}$ which support rate $4R$ are both blocked by lower layer occupied codes.

A first possible countermeasure for the alleviation of code blocking is the clever selection among possible candidate codes during the assignment procedure. Nevertheless, as a consequence of the statistical nature of the departure process, the complete elimination of code blocking is accomplished only if a code reassignment procedure is employed [81], [82]. A reassignment procedure reallocates ongoing calls to other codes so that a new call can always be supported if the system has adequate free capacity to support the requested rate.

5.3.4 System Model

We consider a WCDMA system which supports multi-rate transmission rate by employing an OVSF code tree with L layers. The codes at the leaf layer have the maximum SF ($SF_{\max}$) and consequently the lowest possible transmission rate R, while codes at some layer with spreading factor SF have transmission rate $\frac{SF_{\max}}{SF} \times R$. We denote the code capacity occupancy state (COS) by c, $0 \le c \le C_T$, where C_T is the total capacity of the OVSF code tree and is equal to the number of leaf layer codes. We assume that two types of calls, namely new and HO calls, share the total system capacity C_T and that code blocking is completely eliminated by some code reassignment procedure so that an incoming request, either new or HO, is blocked only due to capacity shortage. A call is assigned a code of rate kR, where $k = \frac{SF_{\max}}{SF}$ =1, 2, 4, 8,… is a power of two as above and SF is the requested spreading factor.

Finally, without loss of generality, we assume that the only possible requested rates are $8R$, $4R$, $2R$ and $1R$. The HO load is assumed to be known in advance by the measurements collected in the Node B. For simplicity of the model analysis, we assume that the distribution of the arrival traffic is Poisson and denote by $\lambda_{n,k}$ and $\lambda_{h,k}$ the arrival rates of new and HO calls, respectively, requesting a rate of kR. We also assume that the mean service time and the cell residence time of a call follow negative exponential distributions and therefore the average code holding time for a call with transmission rate kR is also exponentially distributed with mean of $1/\mu_k$.

5.3.5 Related Work

Many reservation schemes have been proposed in the literature in order to alleviate the HO dropping rate and favor HO calls over new ones. At [83], the authors proposed a Dynamic Multiple Threshold Bandwidth Reservation (DMTBR) scheme, which employs multiple bandwidth reservation thresholds in order to provide QoS provisioning in wireless multimedia networks. At [84] a virtual guard channel (VGC) scheme for handoff calls is evaluated in comparison with the conventional guard channel (GC) scheme. At [85] the performance of a handoff scheme with guard channels in ATM-based mobile networks is considered. The handoff guard channels model together with DOVE (Delay Of Voice End-user) algorithm is proposed at [86]. The aim of this scheme is to preserve the QoS of handoff calls while at the same time to guarantee the QoS of new calls. A multi-level dynamic guard channels (MLDGC) scheme which aims to efficiently enhance the admission priority for handoff calls is proposed at [87]. Finally, at [88] a scheme based on the fractional guard channel policy is proposed and its performance is evaluated via simulation.

Concluding, the main idea behind all the guard channel schemes is the reservation of additional bandwidth for the HO calls from the total available capacity. This is achieved by prohibiting the acceptance of new calls if the channel utilization has superseded some threshold value. In 3G WCDMA systems the bandwidth allocation to the connections is achieved through the use of the OVSF codes. Hence, any guard scheme proposed for 3G systems should be aware of the OVSF code tree characteristics. Therefore, although the scheme which we propose belongs to the well known family of guard channel schemes it differs from other previously proposed schemes since it is adapted to the features of the OVSF code tree.

5.3.6 The Guard Code Reservation Scheme

As we already mentioned, the failure of an incoming HO call yields poor network performance in the case of soft HO, and deterioration of the QoS perceived by the user, in the case of hard HO. Thus, the failure of a HO attempt due to code capacity shortage is less acceptable compared to the blocking of a new call due to the same reason. As we previously mentioned the most well known methods which are employed in order to alleviate the HO dropping rate are the guard channel reservation schemes which are easy to implement and manage.

In our analysis, we consider OVSF codes as the precious resource at the downlink of WCDMA systems and apply a guard code reservation scheme to favor incoming HO calls over new calls. As long as the available code capacity is not lower than a threshold value, C_H , the BS allocates codes to either new or HO calls. If, however, the available capacity falls below, C_H , only HO calls are accepted and new calls are blocked. Assuming that the HO load is known by the measurements collected at the base station, the base station can dynamically adapt the threshold value in the guard code reservation scheme.

5.3.6.1 System State Space

We denote the system code occupancy state by ($x_8\ x_4\ x_2\ x_1$), where x_8, x_4, x_2, x_1 are integers denoting the number of occupied codes of rate $8R$, $4R$, $2R$ and $1R$, respectively, such that $8x_8 + 4x_4 + 2x_2 + x_1 = c$. For example the system state (0012) corresponds to code occupancy $c=4$ with no codes of rate $8R$ and $4R$, one code of rate $2R$ and two codes of rate $1R$ occupied. Depending on the possible transmission rates in the system, we can obtain all the feasible occupancy states in the OVSF code tree by using a generation function as in [81]. For example, all the possible states in an OVSF code tree with $C_T=8$ are 36 as shown in Figure 5.3. When the OVSF code tree size increases the total number of states will increase dramatically. For example, at $C_T=16$, total number of states=201 states, while at $C_T=32$, total number of states=1625 states [81]. (See Appendix I which provides the generation function and the possible states when $C_T=16$).

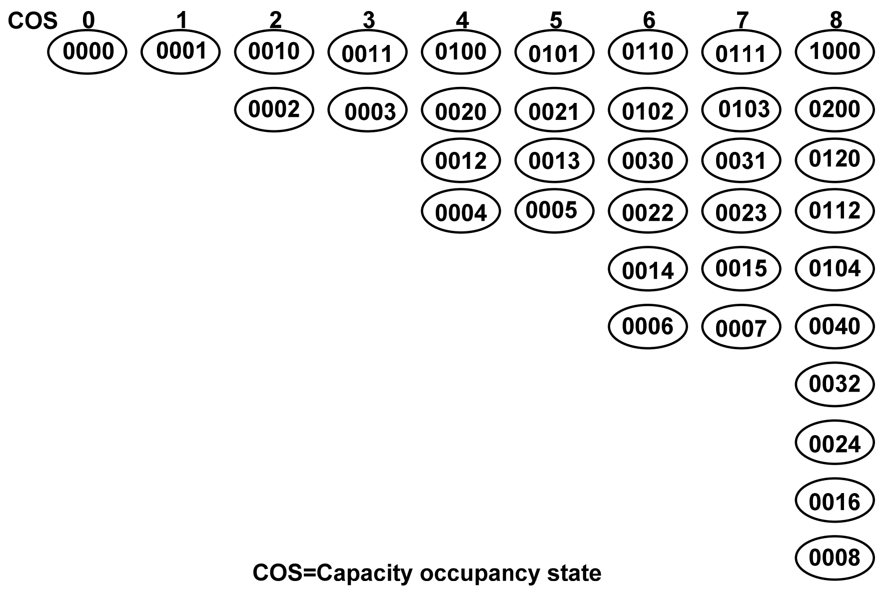

Figure 5-3: Code occupancy state space when $C_T=8$.

5.3.6.2 Performance Analysis

The code occupancy of the OVSF code tree can be modeled by a Markov chain with multiple transitions among the feasible states. Figure 5.4 shows only a small part of the Markov chain, illustrating all the feasible transitions into and out of state (0112), when $C_T=16$ and $C_H=6$. As the used code capacity is $c=6$, the available capacity is $C_T-c=10$ and the threshold is $C_H=6$, no new calls of rate $8R$ and $4R$ are allowed in the system and the corresponding transitions from this state with arrival rates $\lambda_{n,8}$ and $\lambda_{n,4}$ are absent from the transition diagram. The flow balance equation for the same state is given in the following:

$$P_{0112}\left(2\mu_1+\mu_2+\mu_4+\lambda_{n,1}+\lambda_{h,1}+\lambda_{n,2}+\lambda_{h,2}+\lambda_{h,4}+\lambda_{h,8}\right)=P_{0113}\left(3\mu_1\right)+P_{0122}\left(2\mu_2\right)+ \tag{5.1}$$
$$P_{0212}\left(2\mu_4\right)+P_{1112}\left(\mu_8\right)+P_{0111}\left(\lambda_{n,1}+\lambda_{h,1}\right)+P_{0102}\left(\lambda_{n,2}+\lambda_{h,2}\right)+P_{0012}\left(\lambda_{n,4}+\lambda_{h,4}\right)$$

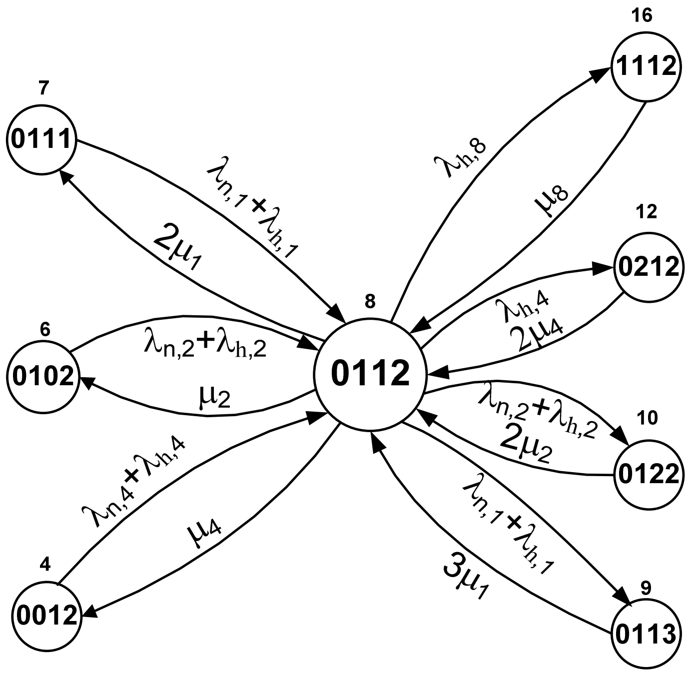

Figure 5-4: Markov state transitions to and out of state (0112) when $C_T = 16$ and $C_H = 6$.

For constructing the general form of the flow balance equations the following notations are necessary:

- P_j : the steady state probability of the code capacity state $j = (x_8\ x_4\ x_2\ x_1)$.
- C_j : code capacity occupancy in state j.
- $n_k(j)$: the number of calls with transmission rate kR in state j.
- $j \oplus k$: the new state after accepting a call of rate kR into the system while in state j.
- $j \ominus k$: the new state after completion of a call of rate kR, while in state j.
- $I_n(j,k)$ and $I_h(j,k)$: functions that indicate the possibility of accepting a new or HO call of rate kR while in state j, such that:

$$I_n(j,k)=\begin{cases}1 & \text{if } C_{j\oplus k}\le C_T-C_H,\\ 0 & \text{otherwise.}\end{cases} \quad \text{and} \quad I_h(j,k)=\begin{cases}1 & \text{if } C_{j\oplus k}\le C_T,\\ 0 & \text{otherwise.}\end{cases}$$

Now we can derive the general form of the flow balance equations for any state j, as follows:

$$P_j \cdot \sum_{k=1,2,4,8} \left(n_k(j)\cdot\mu_k\right) + P_j \cdot \sum_{k=1,2,4,8} \left(I_n(j,k)\cdot\lambda_{n,k} + I_h(j,k)\cdot\lambda_{h,k}\right) = \tag{5.2}$$
$$\sum_{k=1,2,4,8} \left(P_{j\oplus k}\cdot I_h(j,k)\cdot n_k(j\oplus k)\cdot\mu_k\right) + \sum_{\substack{k=1,2,4,8 \\ C_{j\ominus k}\le C_T - C_H,\, n_k(j)>0}} \left(P_{j\ominus k}\cdot\left(I_n(j\ominus k,k)\cdot\lambda_{n,k} + I_h(j\ominus k,k)\cdot\lambda_{h,k}\right)\right).$$

The complete set of equations includes the sum of steady-state probabilities of all code occupancy states which is equal to unity, $\sum_j P_j = 1$. The solution of the above set of equations yields the steady state probability P_j for each state j.

The probability $P_{COS}(c)$ of having a capacity occupancy c is defined as:

$$P_{COS}(c) = \sum_{\forall j: C_j = c} P_j \qquad 0 \le c \le C_T. \tag{5.3}$$

The new call blocking probability $P_{nb}(kR)$, of the calls requesting code rate kR (k=1, 2, 4 and 8) is given by:

$$P_{nb}(kR) = \sum_{c=C_T-C_H-(k-1)}^{C_T} P_{COS}(c). \tag{5.4}$$

The HO failure probability $P_{hf}(kR)$, of the calls requesting rate kR is given by:

$$P_{hf}(kR) = \sum_{c=C_T-(k-1)}^{C_T} P_{COS}(c). \tag{5.5}$$

Another important performance measure is code utilization U_T, which is defined as the percentage of the total code capacity being utilized and can be expressed as:

$$U_T = \frac{\sum_{c=1}^{C_T} c \times P_{COS}(c)}{C_T}. \tag{5.6}$$

5.3.6.3 Numerical Results

The performance of the proposed guard code reservation scheme is evaluated in an OVSF code tree with total capacity $C_T = 16$. All the possible code occupancy states of the system, as well as the flow balance equations were generated by a C++ program.

Different input parameters, such as call arrival and call service rates, as well as capacity threshold values produced different instances of the problem. The equations were then fed to Mathematica to calculate the steady states probabilities equations and metrics of interest, i.e. the new call blocking and handover failure probabilities as well as code utilization.

We used several thresholds values $C_H = 4$, 6, and 8, different traffic intensities and different traffic mix of new and HO calls. Furthermore, in our study we assume two different distributions for the traffic patterns of both new and handover calls. The first distribution is ($8R:4R:2R:1R$) = ($10:20:30:40$), where the lower rate calls dominate the overall traffic volume in the system while the second distribution is uniform ($8R:4R:2R:1R$) = ($25:25:25:25$). We will refer to these two traffic distributions as A and B respectively.

5.3.6.3.1 New Call Blocking and Handover Failure Probabilities

Figures 5.5 to 5.8 show the new call blocking and handover failure probabilities (P_{nb} and P_{hf}) as the offered traffic load increases following the traffic patterns A and B respectively. In this particular problem instance, we assumed that the HO traffic load is only a low percentage (20%) of the total offered load and $C_H = 4$. As illustrated in Figures 5.5 and 5.7 for both distributions (A, B) the higher rate calls suffer higher blocking and dropping than the lower rate calls. However, due to the guard code scheme the HO failure rate of every call rate (Figs. 5.6 and 5.8) is significantly lower than the new call blocking probability of the corresponding call rate.

We repeat this scenario to investigate the effect of the HO traffic load in the cell. We also assume that the same code reservation $C_H = 4$, while the HO traffic load is high percentage (40%) of the total offered load. As illustrated in the Figures 5.9 to 5.12, the effect of the heavy load of the HO traffic in both the blocking and the dropping probabilities. As we can see in these figures, the new call blocking probability P_{nb} is slowly increased compared with low HO traffic load. However due to the heavy load of HO traffic, the HO failure rate, P_{hf} is increased dramatically compared with low HO traffic load as shown in Figs. 5.10 and 5.12.

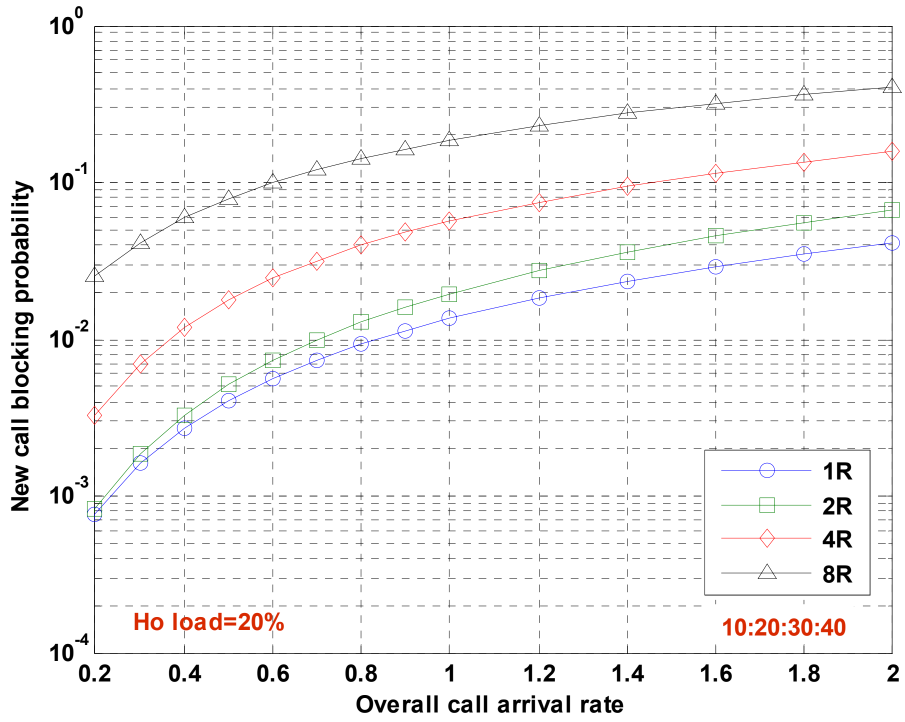

Figure 5-5: New call blocking probability versus overall call arrival rate (λ) at traffic pattern A and HO load 20%.

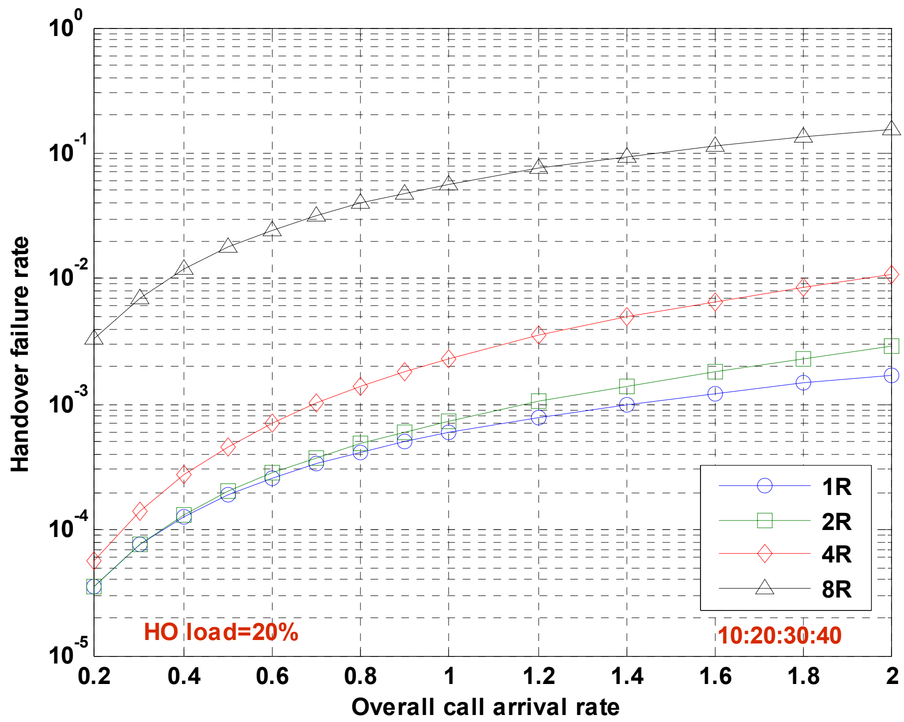

Figure 5-6: Handover failure rate versus overall call arrival rate (λ) at traffic pattern A and HO load 20%.

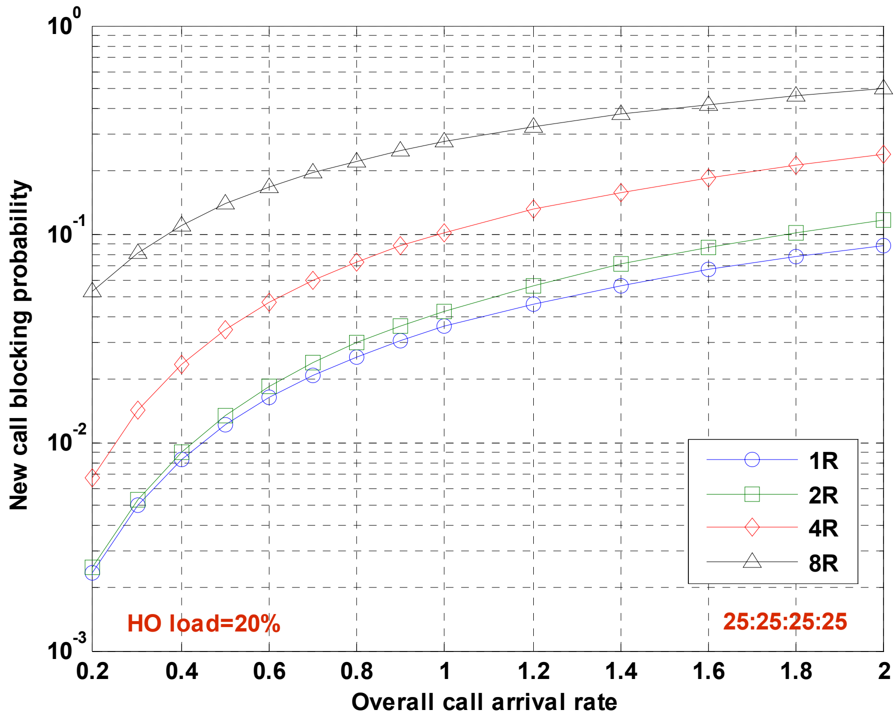

Figure 5-7: New call blocking probability versus overall call arrival rate (λ) at traffic pattern B and HO load 20%.

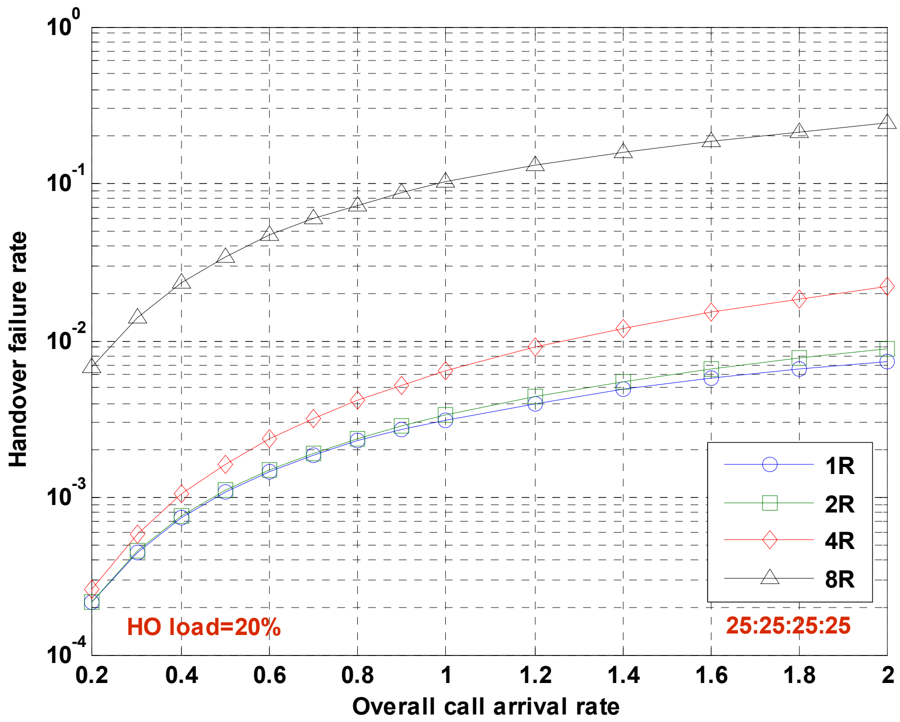

Figure 5-8: Handover failure rate versus overall call arrival rate (λ) at traffic pattern B and HO load 20%.

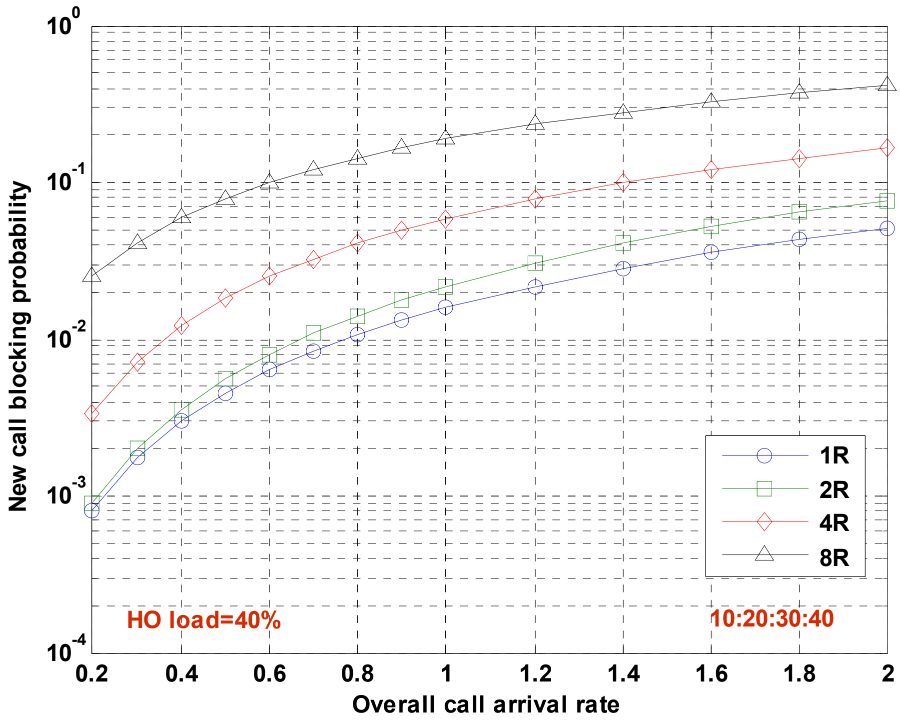

Figure 5-9: New call blocking probability versus overall call arrival rate (λ) at traffic pattern A and HO load 40%.

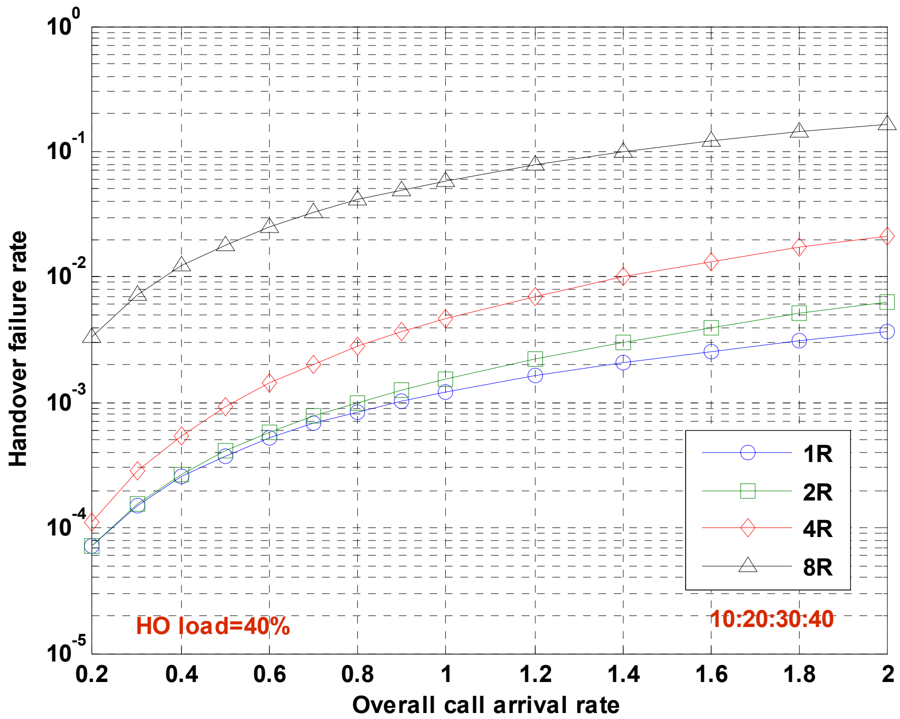

Figure 5-10: Handover failure rate versus overall call arrival rate (λ) at traffic pattern A and HO load 40%.

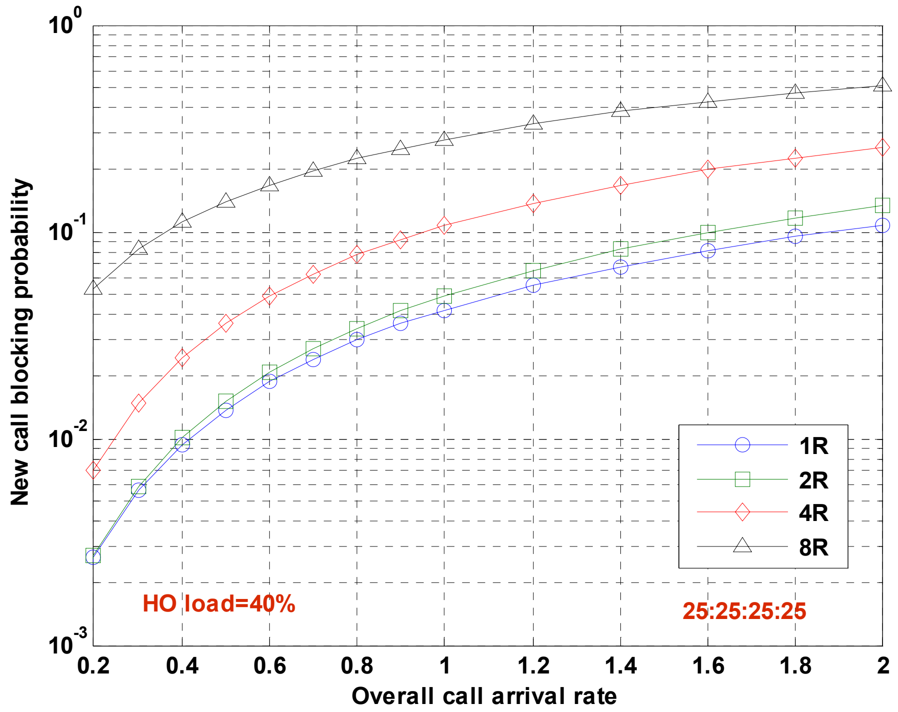

Figure 5-11: New call blocking probability versus overall call arrival rate (λ) at traffic pattern B and HO load 40%.

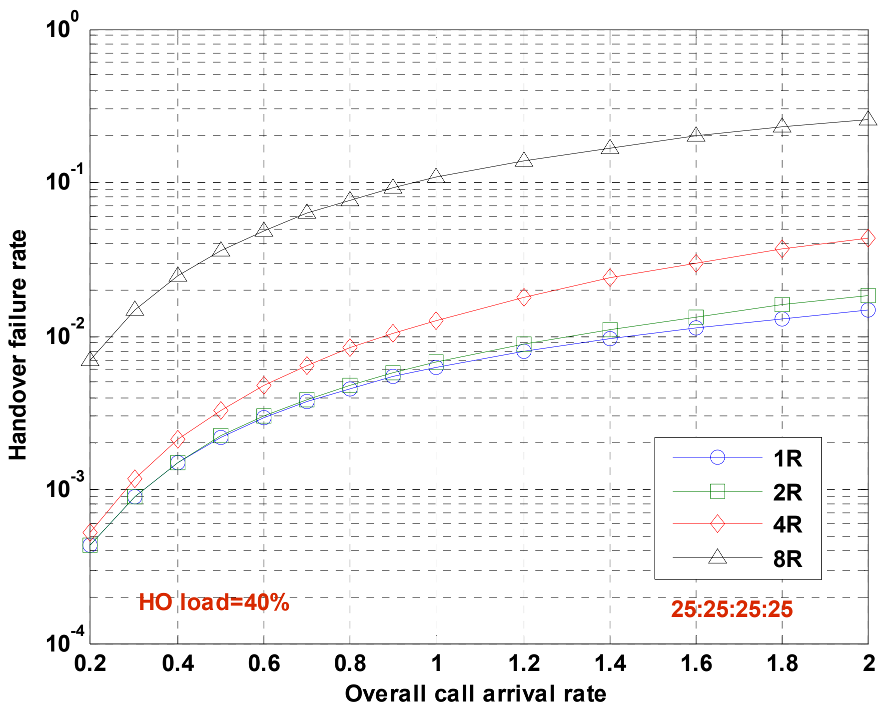

Figure 5-12: Handover failure rate versus overall call arrival rate (λ) at traffic pattern B and HO load 40%.

5.3.6.4 The Capacity Threshold

The value of the code reservation threshold plays a significant role at the utilization of the available bandwidth as well as to the overall system performance. In this scenario we employ traffic pattern A and we compare in Figure 5.13 the code bandwidth utilization U_T at different code reservation thresholds (C_H=0, 4, 6, and 8). The case C_H=0 corresponds to the complete sharing scheme where no code capacity reservation for HO calls exists.

As shown in the figure, U_T increases nearly linearly as the offered load increases and it achieves the highest values when there is no capacity reservation, as expected. Of course, this benefit comes with the cost of increased HO failure rates. Also as indicated in the figure, U_T at large value of code reservation C_H, has low value. Therefore, better performance will be achieved at lower code reservation.

Figure 5.14 shows the performance of P_{nb} and P_{hf} probabilities of calls requesting rates $1R$ and $4R$ at different code capacity thresholds when the overall arrival rate ($\lambda = 1$). It seems that for this certain problem instance the increase of threshold above $C_H = 4$, does not yield any significant improvement at the HO failure rate, while causes unnecessary deterioration of the new call blocking.

Therefore, we can conclude that from Figs. 5.13 and 5.14, better system performance in terms of bandwidth utilization and QoS provisioning for HO calls can be achieved at lower values of code reservation capacity threshold (C_H).

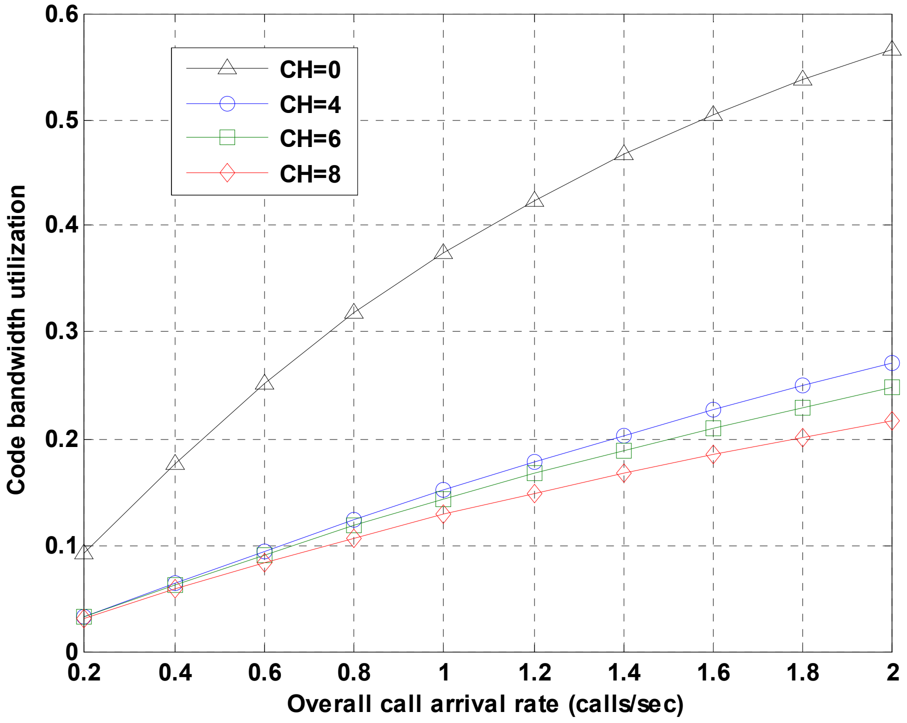

Figure 5-13: Code capacity utilization versus overall call arrival rate (λ) at traffic pattern A and HO load 20%.

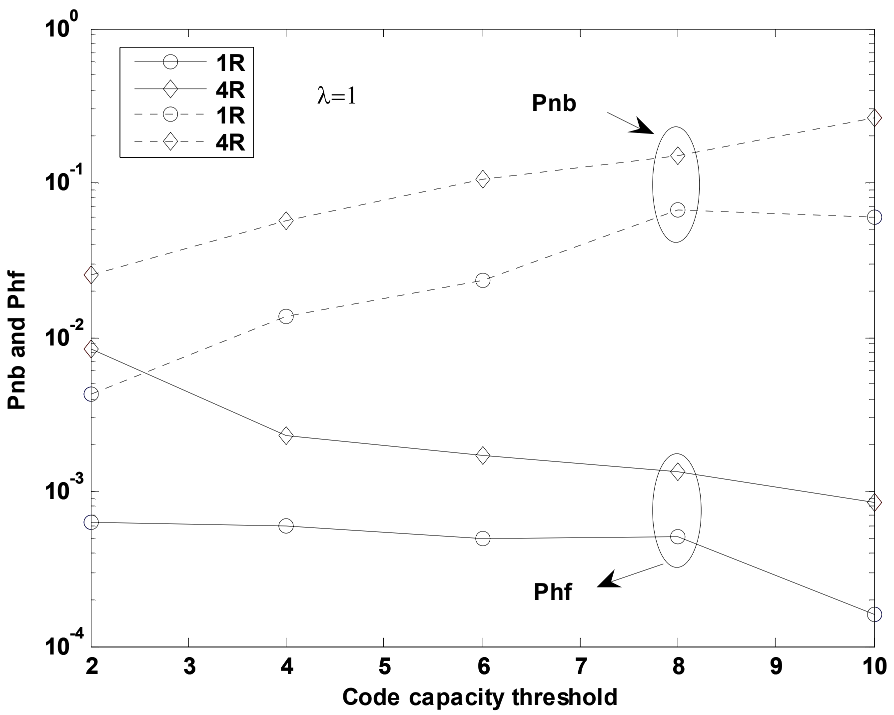

Figure 5-14: New call blocking versus HO failure rate at different code reservation thresholds at traffic pattern A and HO load 20%.

5.4 Part 2: Effective Combination of CAC Scheme and Traffic Scheduling Algorithm for Handover Provisioning in WCDMA Systems

At Part 1 in this Chapter, we introduced a guard code scheme for 3G WCDMA networks. The guard code scheme favors HO calls over new calls and thus improves the HO failure rate. Therefore, we will employ that scheme at the call admission control in the radio resource management algorithm in order to accept the requests of the HO traffic according to the code reservation threshold. Nevertheless, the accepted HO calls experience high packet delays since a number of their queued packets will be forwarded to the target cell through the wireline infrastructure. Hence, the employed traffic scheduler at MAC layer should be delay-aware in order to be able to meet the QoS requirements of the handover calls. Thus, we employ a low complexity delay driven traffic scheduler (DDS) which is able to exploit information from the physical layer. Throughout a cross-layer framework the scheduler is able to utilize SIR measurements from the UE's in order to estimate the state of the wireless channel during the next frame. The priority criterion employed by the traffic scheduler takes into account the maximum packet delay of each connection, the respective delay sensitivity as well as the state of the wireless channel.

5.4.1 System Model

We consider a WCDMA system which supports multi-rate transmission by employing OVSF codes. The total capacity of the OVSF code tree is shared among M_u users which may have a new or HO request in the cell. Therefore, connection $i \in \{1, 2, \cdots, M_u\}$ may have a service rate $R_{S,i}$ that is always a multiple of a power of two of the lowest available rate R:

$$R_{S,i} = 2^k R, \quad k \in \mathbb{Z}^+ \ and \ \ i \forall M_u \tag{5.7}$$

We assume that a code reassignment procedure completely eliminates code blocking and thus an incoming request, either new or HO, is blocked only due to capacity shortage. The use of a common set of orthogonal codes at the downlink of a WCDMA system should theoretically eliminate intra-cell interference. However, due to multipath propagation the downlink signals are not perfectly orthogonal. Consequently, the Signal-to-Interference-Ratio (SIR) perceived by user i, from a set of M_u active users, can be expressed as [64]:

$$SIR_i = \frac{SF_i \times P_{r,i}}{\alpha \cdot P_{\text{intra},i} + P_{\text{inter},i} + P_N}, \qquad i = 1, 2, \cdots, M_u \tag{5.8}$$

where α is the orthogonality factor SF_i is the spreading factor of the assigned OVSF code, $P_{r,i}$, is the received power of user i, $P_{\text{intra},i}$ is the interference generated by the received signals of the other $M_u - 1$ users which are connected to the same cell, $P_{\text{inter},i}$ is the interference received from other neighboring cells and P_N is the thermal noise power.

During connection setup, each connection, i, negotiates with the network management module its SIR threshold $SIR_{T,i}$ and delay threshold $D_{T,i}$. The UE is able to measure and report the respective SIR to the Node B at each frame. Based on a finite state Markov chain (FSMC) model the next state of the wireless channel can be predicted according to the measured value of SIR as previously investigated in Chapter 4 at Section 4.4. Therefore, for each connection, i, the predicted error probability $P_{e,i}$ can be calculated.

5.4.2 Admission Control and Bandwidth Reservation

The failure of a HO attempt due to code capacity shortage is less acceptable compared to the blocking of a new call due to the same reason. Therefore, the proposed CAC scheme introduces a code reservation threshold, C_T, which allows HO calls to use the total capacity of the OVSF code tree, but restricts the new calls to use only a part of the tree capacity. As long as the available capacity, C_A is not lower than the threshold value C_T, the BS allocates codes to either new or HO calls. If, however, the available capacity falls below, C_T, only HO calls are accepted and new calls are blocked.

5.4.2.1 Calculation of the reservation threshold

The reservation threshold is calculated based on measurements of the handover traffic load at the physical layer and it is adjusted to the characteristics of the OVSF code tree. Assuming that the arrival rate, $\lambda_{h,k}$, for each class k of handover calls requesting an average rate of , r_k , is known by measurements collected at the base station, the handover reservation threshold , C_T , can be expressed as follows:

$$C_T = \sum_{k=1}^{K} \left(\lambda_{h,k} \times r_k \times T_{S,k} \right) \times \phi_k \tag{5.9}$$

where ϕ_k and $T_{S,k}$ are the reservation factor and the average service time respectively of class k calls. As the reserved bandwidth will be jointly used by all the handover calls the reservation factors can be summed up to a total reservation factor, Φ . Thus, Equation (5.9) can be simplified as:

$$C_T = \Phi \times \sum_{k=1}^{K} \left(\lambda_{h,k} \times r_k \times T_{S,k} \right), \quad \Phi \in (0,1] \tag{5.10}$$

Ideally, if $\Phi = 1$ then the reserved bandwidth would be adequate to serve the entire HO traffic. However, that would lead to high blocking rate for new calls. Thus, the reservation factor is used in order to tune the performance of the proposed CAC. Furthermore, based on the measured handoff traffic load, the threshold value, C_T , can be periodically updated accordingly.

5.4.2.2 Adaptation to the OVSF code tree

Finally, the reserved capacity at the OVSF code tree should be a multiple of the lowest available rate, R. Therefore, the threshold value, C_T, can be defined as the integer $C_T \in \mathbb{Z}^+$ for which the following Equation holds:

$$\left\lfloor \Phi \times \sum_{k=1}^{K} \left(\lambda_{h,k} \times r_k \times T_{S,k} \right) \right\rfloor \le C_T < \left\lfloor \Phi \times \sum_{k=1}^{K} \left(\lambda_{h,k} \times r_k \times T_{S,k} \right) \right\rfloor + 1 \tag{5.11}$$

If the reserved capacity should be a single branch of the OVSF code tree then, as denoted by Equation (5.7), the reservation threshold should be defined as a power of two of the lowest available rate, R :

$$C_T = 2^k R, \quad k \in \mathbb{Z}^+ \tag{5.12}$$

In this case the reservation threshold is defined as the minimum value of, C_T, as defined by Equation (5.12), for which the following equation holds:

$$C_T = 2^k R \geq \left[\Phi \times \sum_{k=1}^{K} \left(\lambda_{h,k} \times r_k \times T_{S,k} \right) \right] \quad (5.13)$$

Therefore, the use of a reservation factor makes possible to fine tune the CAC performance and dynamically adjust it to the traffic load variations. Consequently, the use of the proposed CAC improves the HO failure rate due to capacity shortage compared to the new call blocking rate.

5.4.3 Traffic Delay Driven Scheduler (DDS) Algorithm

Although the HO calls will be admitted by the proposed CAC with higher priority than new calls at the same time it is expected that they will experience high packet delays. That is due to the inevitable procedure of forwarding packets to the target cell through the wireline infrastructure. Therefore, we propose a delay driven scheduling algorithm which at the same time exploits information from the physical layer in order to increase its efficiency.

5.4.3.1 Avoiding erroneous packet transmissions

Due to the error characteristics of the air interface it is essential that the QoS mechanisms provided in a wireless network be robust and with low computational complexity. The effect of frame error rate and retransmission on the scheduler performance studied in [17]. The optimal frame error rate is function of SIR, propagation model and cell capacity. In particular, based on SIR measurements which are performed by the UE during the previous frame for each connection, i, we can estimate from Equation (4.12) at Chapter 4 the predicted bit error rate ($P_{e,i}$) at the next frame. If the estimated value is higher than the target value of BER the connection is excluded from service for the current frame. Therefore, the scheduler has the ability to avoid erroneous packet transmission for each connection during the rate allocation.

5.4.3.2 Prioritization and rate allocation

For the rest of the connections the priority criterion is calculated. The priority of each connection i with head of line (HOL) packet delay, d_i, delay threshold $D_{T,i}$ and predicted bit error probability $P_{e,i}$, is defined by (5.14)

$$P_i = \frac{d_i\left(1-P_{e,i}\right)}{D_{T,i}} \qquad i = 1,2,\cdots,M_u \tag{5.14}$$

Subsequently, the connections are sorted in decreasing order of their priorities and the scheduler shares the available capacity among the users with the higher priority. In particular, the connection with higher priority is able to use all the available residual capacity for the current frame. However, if the highest priority connection has not adequate backlogged traffic then the assigned rate is scaled down accordingly. The rest of the capacity, if any, is offered to the second user in the priority list and so forth. The detailed description of the rate assignment procedure follows:

Let C be the available system capacity. Starting from the highest priority connection and for each connection i in the sorted list:

1. **IF** connection i has not been assigned a rate r, and $C-R \geq 0$, set $r \leftarrow R$

2. **WHILE** connection i has more queued packets than the packets that can be transmitted during the next frame at the assigned rate r and $C-2\cdot r \geq 0$,

    i) Double the assigned rate, $\left(r \leftarrow 2\cdot r\right)$

    ii) Decrease C by the assigned rate $\left(C \leftarrow C-2\cdot r\right)$

The rate assignment process concludes when $C=0$ or when all the assigned rates to the connections of the list are capable of transmitting all their respective queued packets during the next frame.

Finally we can say that, the basic idea of the proposed traffic scheduler is to prioritize connections based on their head of line (HOL) packet delay as well as their delay sensitivity. Thus, the incoming HO calls will gain higher priorities compared to the ongoing calls and the handover procedure becomes more transparent to the user. Furthermore, by utilizing SIR measurements from the physical layer the scheduler is able to avoid erroneous packet transmissions and thus increases the bandwidth utilization.

5.4.4 Related Work

The provisioning of services with different QoS requirements poses many challenges in the radio resource management algorithms. Recently, many researchers have focused on radio resource management design, because the 3G and beyond networks provided high-speed multimedia services. Most of the works have been done in that area has achieved the QoS provisioning at the CAC level. Due to the adaptive multimedia services, the authors in [89] have proposed a new scheme for QoS provisioning based on bandwidth degradation. They used two new QoS parameters namely, degradation ratio (DR) and degradation degree (DD) at the call level and hence the QoS provisioning based on these parameters. However, some multimedia applications have stringent QoS requirements and can not accept any type of violations on their services. At [90] the authors introduced a QoS-aware radio resource management scheme based on dynamic interference guard margin (IGM). They utilized two concepts to derive the IGM value at the CAC level: guard channel (GC) and load curve (LC). Although the proposed scheme reduced the failure rate of the handover calls, the scheme used a complex reservation scheme which is based on IGM and also induced service degradation. Again at the CAC level, the authors at [91] proposed an efficient QoS provisioning scheme for adaptive multimedia services. This scheme is based on Q-learning methods to solve the QoS provisioning for adaptive multimedia applications. However the Q-learning utility may increase the complexity of the radio resource management algorithm, although it contains the traffic scheduler which is able to adapt the bandwidth of the applications based on the information collected at the physical layer. As illustrated from the previous works, all the authors have solved the QoS provisioning problems at only a CAC level. In this Chapter we presented an effective combination between a CAC mechanism and a traffic scheduling algorithm for efficient radio resource management by taking into account the features of 3G WCDMA networks.

5.4.5 Numerical Results and Discussion

The performance of the proposed radio resource algorithm based cross-layer framework is evaluated throughout event driven simulation. We assume a single cell scenario where the base station (BS) is located at the center of a hexagonal cell with a radius of R_{cell} =1km. We also assume a 7-layer OVSF code tree and therefore a total capacity of $64R$. The UE's are uniformly distributed in the cell and each class k connection has an average velocity of, v_k. Thus, the boundary crossing rate $\lambda_{h,k}$ for the class k connections is [92]:

$$\lambda_{h,k} = \frac{D_k v_k L_{cell}}{\pi} \tag{5.15}$$

where L_{cell} is the length of the perimeter of the cell and D_k is the density of the UE's in the cell. The initial value of D_k is 50 connections per square kilometers. The rest of the most simulation parameters can be found in Chapter 4 at Section 6.6, while in this simulation scenario we employ two service classes which have the characteristics presented at Table 5.1.

Traffic Parameters	**Class1**	**Class2**
N_L, N_H	6, 4	12, 4
P_{on}	0.4	0.5
Average service time	180 *sec*	180 *sec*
Average rate	2R	4R
QoS Requirements	**Class1**	**Class2**
Delay Threshold	0.5 *sec*	1 *sec*
Target BER	10^{-4}	10^{-4}

Table 5-1: Traffic and QoS parameters.

5.4.5.1 Performance of the Guard Code CAC scheme

At the first scenario we want to study the performance of the proposed CAC under different reservation factors. The reservation threshold is defined by Equation (5.13) while the reservation factor is increased from 0 up to 30% with a step of 10%.

Figures 5.15 and 5.16 illustrate the average handover failure rate and new call blocking rate respectively for increasing traffic load. At $\Phi = 0$ there is no OVSF code reservation and therefore both new and handover calls have the same probability to fail. However, as the reservation factor increases the corresponding plots in Figure 5.15 are shifted downwards compared to the plot at $\Phi = 0$, while on the other hand the plots at Figure 5.16 are shifted upwards. As it is expected, the increase of the reservation factor yields a decrease of the handover failure rate and a respective increase of the new blocking rate. However, the proposed CAC gives the capability to dynamically adjust the reservation factor in order to take into account the variation of the UE's mobility as well as the characteristics of the total offered load.

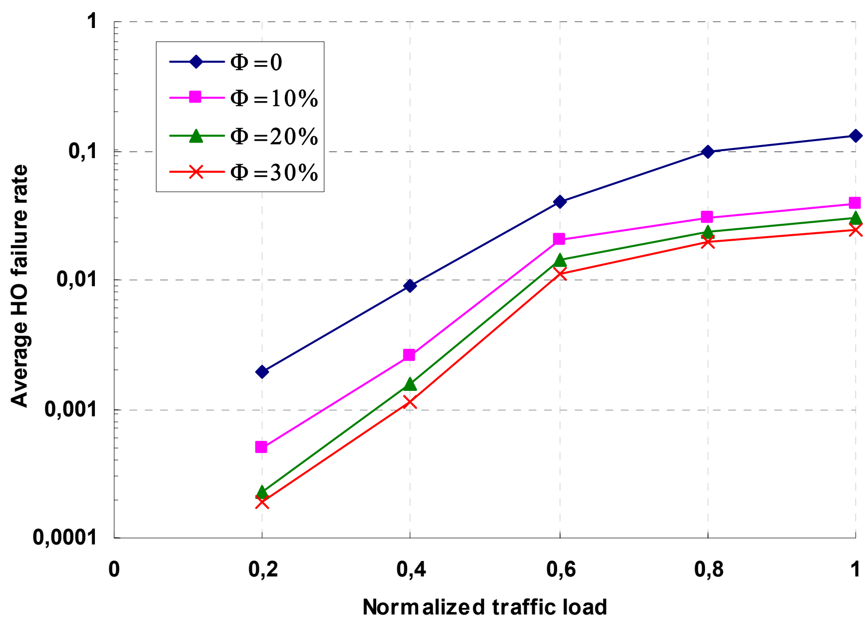

Figure 5-15: Average handover failure rate at different reservation factors.

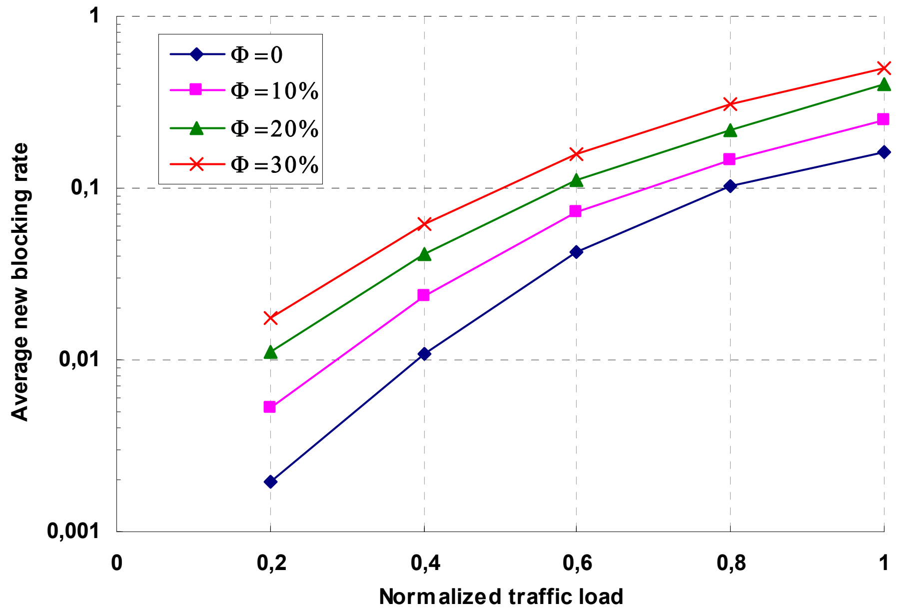

Figure 5-16: Average blocking rate of new calls at different reservation factors.

5.4.5.2 Performance of the traffic scheduler

In this scenario we want to study the performance of the proposed scheduler. Thus, we assume a reservation factor of $\Phi = 0$, while the classic maximum C/I scheduler is also evaluated for comparative purposes. The maximum C/I scheduler prioritizes the connections based on the respective carrier to interference ratio and thus it favours connections with better channel quality.

5.4.5.2.1 Capacity Utilization

Figure 5.17 illustrates the utilization of the available capacity under the DDS and Max C/I scheduling disciplines for increasing traffic load. As expected Max C/I achieves highest utilization at all traffic loads compared to DDS. By always favouring the connection with the highest channel quality, Max C/I achieves the least possible erroneous packet transmissions and consequently the highest possible utilization of the available capacity. On the other hand, due to the use of the BER prediction criterion DDS is able to exploit efficiently the information gathered from the physical layer. Thus, even though throughput maximization is not the main goal of our proposed scheduler, DDS is also able to achieve high capacity utilization.

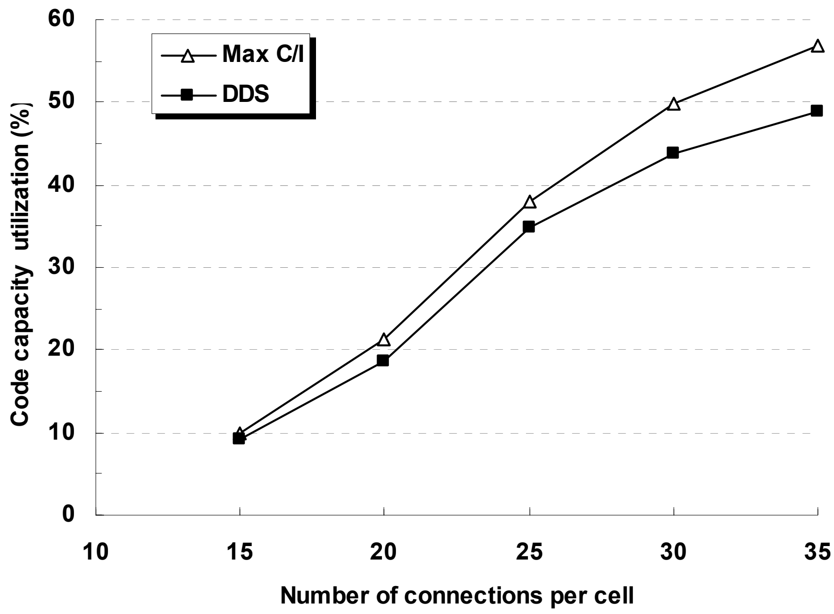

Figure 5-17: Capacity Utilization under Max C/I and DDS schedulers.

5.4.5.2.2 Delay sensitivity

Figures 5.18 and 5.19 show the average packet delay for handover and new calls of classes 1 and 2 respectively. Looking at the overall performance of the two schedulers we can observe that DDS outperforms Max C/I at all traffic loads.

Additionally, DDS as a delay driven scheduler gives priority to calls which experience high packet delays such as handover calls. Thus, it manages to achieve nearly the same packet delay for calls of the same class irrespective of their origin as New or HO calls. On the other hand, under the Max C/I scheduling discipline the handover calls experience higher packet delays than the new calls especially at high traffic loads. This is due to the fact that Max C/I takes into account only the state of the wireless channel of each connection ignoring the respective delay sensitivity and actual delay experienced by the connections.

This conclusion can also be verified by Figure 5.20 where the average queue size of all the ongoing connections is depicted. As the traffic load increases the average queue size of all ongoing connections is also increased. However, under the DDS scheduling discipline, the connections which experience high packet delays and therefore have more queued packets they also gain higher service priority. As a result, DDS achieves to reduce the average queue size at all traffic loads compared to Max C/I.

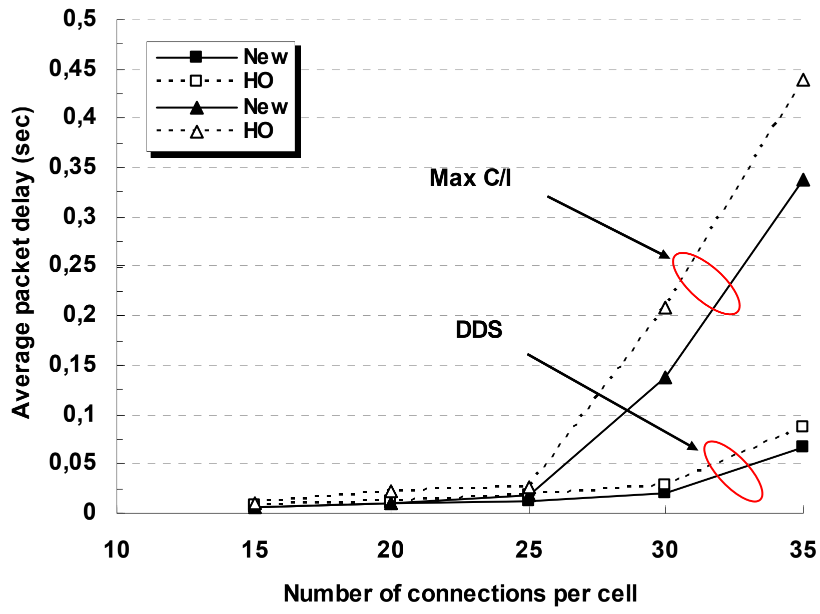

Figure 5-18: Average packet delay for Class 1 HO and New calls under Max C/I and DDS schedulers.

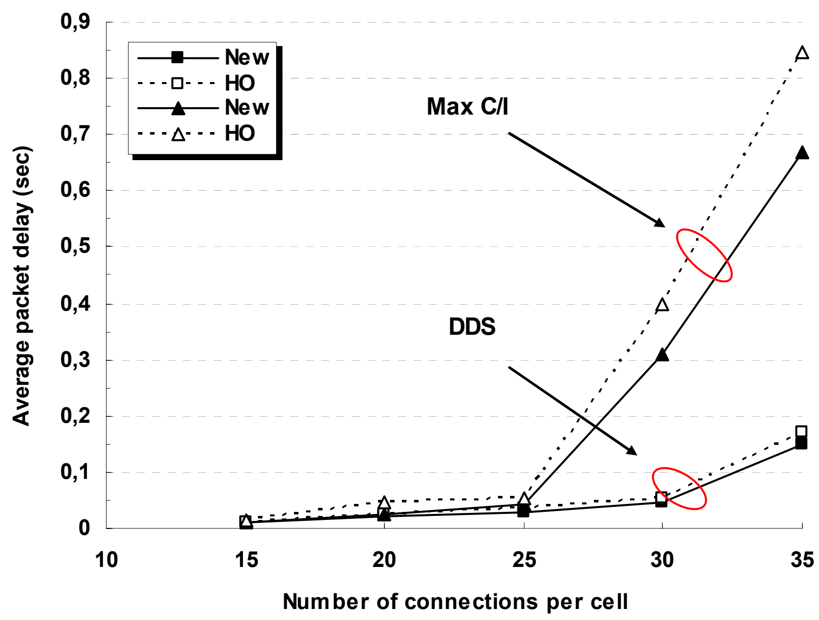

Figure 5-19: Average packet delay for Class 2 HO and New calls under Max C/I and DDS schedulers.

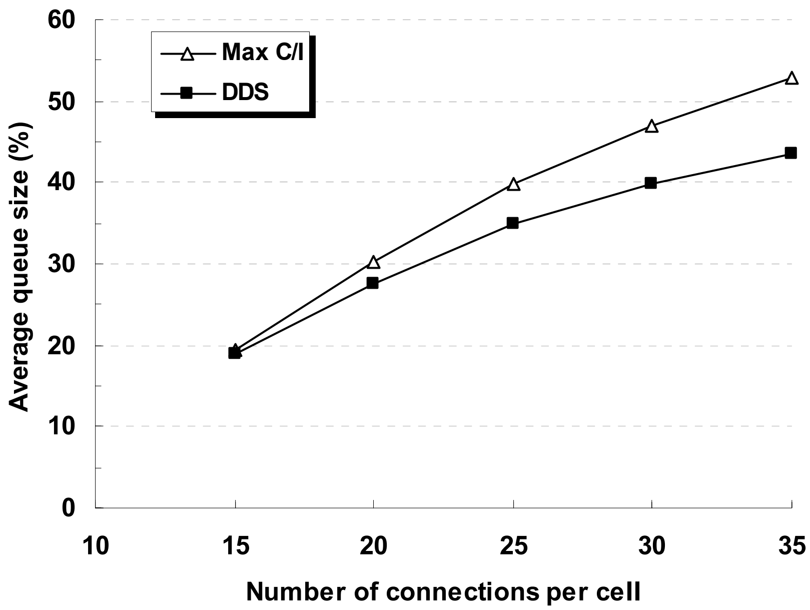

Figure 5-20: Average queue size of all ongoing connections under Max C/I and DDS schedulers.

5.5 Concluding Remarks

In this Chapter, we have proposed radio resource management design for handover provisioning in 3G WCDMA networks. The radio resource management algorithm is an effective integration between a CAC mechanism at the call level and a traffic scheduling algorithm at the packet level. These two schemes are deigned and operate in complementary fashion in order to support QoS requirements of handover calls and ensure transparent user communication during the handover procedure.

In Part 1 at this Chapter, we proposed a guard code scheme which favors the HO calls over new calls in WCDMA system employing OVSF codes as channelization codes. The performance analysis of the proposed scheme is analyzed by using Markov chain with multiple transitions. The scheme belongs to the well known family of guard channel schemes and reserves some code capacity to prioritize the HO call. To the best of our knowledge, the existing literature in OVSF code management focuses only on the code blocking and code assignment and reassignment issues, without distinguishing between new and HO connection requests. The performance evaluation of the proposed scheme is evaluated by using different code thresholds reservation, different HO traffic load and two different traffic patterns. The performance metrics such as new call blocking probability, HO failure rate, and the code utilization are discussed.

In Part 2 a new design for effective combination between a CAC mechanism and a scheduling procedure at MAC layer of the WCDMA 3G networks is proposed. This design is based on the cross layer technique. We introduced a CAC scheme at the call level which is based on a guard code reservation in the OVSF code tree. The calculation of the code reservation threshold and adaptation to the OVSF codes are driven. Then we introduced a delay driven scheduling (DDS) algorithm which exploits information from the physical layer in order to avoid erroneous packet transmission and increases system performance. The scheduling procedure is not only based on the delay sensitivity but also based on the predicted error probability of each user in the cell.

Compared to the existing literature in handover provisioning schemes which focuses only on provisioning QoS requirements for HO calls by using a CAC mechanism at call level, our proposed scheme has the capability of utilizing the available radio resources and also supporting the QoS provisioning for HO calls at both call and packet levels. The proposed scheme introduced an efficient integration between a CAC scheme and a traffic scheduling algorithm at the MAC layer for provisioning QoS requirements of handover calls. We discussed the performance of a CAC scheme at different code reservation thresholds. However, when the total available capacity is shared among the users, the performance evaluation of the scheduler is evaluated in terms of the average packet delay, average queue size and the code utilization.

Chapter 6 - Minimizing CQI Signaling Overhead in HSPA

6.1 Introduction

High Speed Packet Access (HSPA) aims to advance the performance of the existing UMTS networks by improving the level of Quality of Service (QoS) and increasing the supported peak data rates at both the forward and the reverse link. HSPA consists of the High Speed Downlink Packet Access (HSDPA) and High Speed Uplink Packet Access (HSUPA) standards. Different technologies are considered for enhancement WCDMA system performance in 3GPP specifications in order to deploy HSDPA for next generation of mobile communication systems. Enhanced channel quality information (CQI) reporting scheme extends the Release 5 regular feedback CQI scheme is introduced to improve the radio link performance of HSDPA systems, [93], [94] and [95]. The reported CQI value estimates the transport high speed downlink shared channel (HS-DSCH) block size, number of codes, and modulation which could be received with a given 10% error probability. Therefore, the CQI reporting scheme is directly related to the accuracy of selection the suitable transmission rate and the efficiency of adaptive modulation and coding (AMC) procedure.

AMC procedure of HSDPA makes possible the adaptation of the employed Transport Format and Resource Combination (TFRC) to the wireless channel variations as well as to the varying rate requirements of the UE. However, to perform efficiently AMC requires frequent CQI reports by the UE to Node B. As the number of CQI reports increases, uplink interference also increases, thus reducing uplink reception quality. In this Chapter, we propose an improved CQI reporting scheme which reduces the required CQI signaling by exploiting a CQI prediction method based on a Finite State Markov Chain (FSMC) model of the wireless channel.

The rest of this Chapter is organized as follows. In Section 6.2 we overview the HSDPA system, and describe the network architecture based HSDPA system. In Section 6.3 we describe the system concepts and problem statement as well as introduce a simplified analysis for uplink interference. Then we present a prediction-based CQI reporting (P-CQI) scheme in Section 6.4. While in Section 6.5 we introduce the Node B packet scheduling scheme. In Section 6.6 the performance evaluation of the proposed scheme is discussed. Finally in Section 6.7 we briefly conclude this Chapter.

6.2 HSDPA System Architecture

The existing UMTS-WCDMA system already contains a downlink shared channel in terms of DSCH but HSDPA extends this concept to significantly provide higher throughput and hence more efficient use of the radio spectrum. This channel is called High Speed Downlink Shared Channel (HS-DSCH). Therefore, HSDPA is an enhanced version of WCDMA system in 3GPP standard employs new number of parallel shared channels and adaptive modulation and coding (AMC) scheme to be able to provide a high data rate transfer from base station (Node B) to user equipment (UE) [61]. Figure 6.1 illustrates the new channels and processing mechanisms in the HSDPA network architecture.

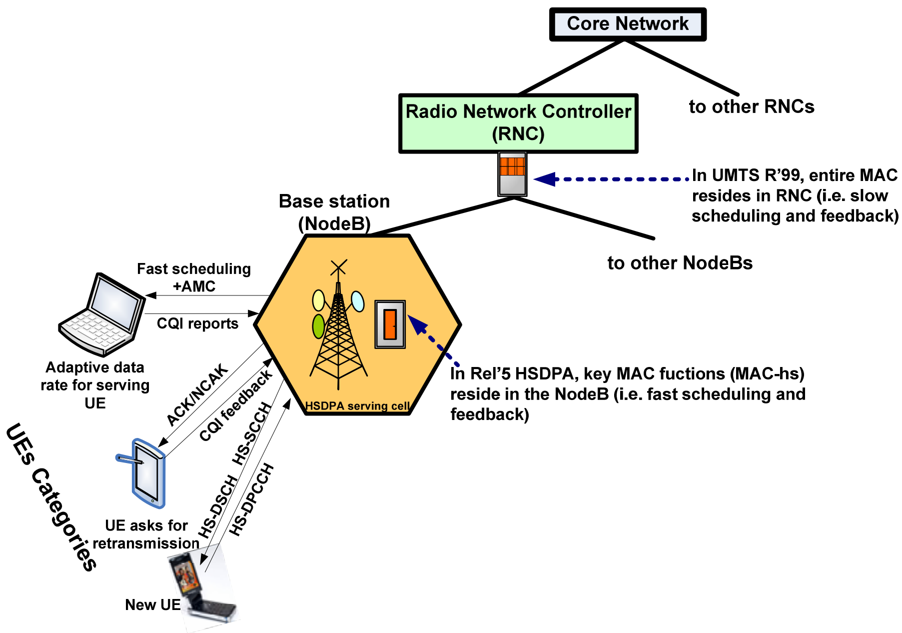

Figure 6-1: HSDPA network architecture.

The packet scheduler operation at the MAC-hs (high speed) sublayer in Node B is considered as the key component for QoS provisioning to UEs. The packet scheduler is shifted from the MAC layer at RNC in WCDMA system to MAC-hs sublayer at Node B in HSDPA. Therefore, one of the most important features of HSDPA is fast processing for packet scheduling. This is because the packet scheduler in HSDPA needs to track and monitor the variation of channel conditions for different UEs in the cell. In addition, the transmission time interval (TTI) is reduced to 2ms in order to allow the packet scheduler adaptively assigns the transmission rate by exploiting the variations of the wireless channel of UEs in the serving cell and then the scheduling decision becomes faster. Therefore, at every subframe, the scheduler can determine the number of codes and AMC scheme (i.e. peak data rate) assigned to each UE, based on the feedback CQI reports from the UEs. Furthermore, reduction of TTI allows for faster data processing, less buffer space at UE and Node B. This will lead to increase the end user throughput.

6.3 System Concepts and Problem Statement

6.3.1 System Concepts

The HSDPA operation utilizes a number of new channels [61]. The user data are transmitted through the High-Speed Downlink Shared Channel (HS-DSCH) while the associated signaling is transmitted through the high-speed shared control channel (HS-SCCH) at the forward link and at the high-speed dedicated physical control channel (HS-DPCCH) in the uplink as shown in Figure 6.1.

The feedback information from the UE to the Node B carried on the HS-DPCCH, is essential for the HSDPA operation as it makes possible the use of Adaptive Modulation and Coding. The HS-DPCCH frame consists of three slots and has duration of 2ms as shown at Figure 6.2. The first slot is used by the Hybrid Automatic Repeat request (HARQ) process while the other two slots are used for the transmission of CQI value. CQI is not a direct Signal to Interference and Noise (SINR) ratio measurement but instead it is an integer index to the Transport Format and Resource Combination (TFRC) which the UE requests from the packet scheduler located at Node-B [70]. The requested TFRC corresponds to the maximum transport block size which has a minimum of 90 per cent probability to be transmitted correctly. Two successive CQI values correspond approximately to a step of 1dB at the SINR of the HS-DSCH [61], [70]. At Release 5, 3GPP [96] proposes a periodic CQI feedback scheme which, as illustrated at Figure 6.3, has a report cycle of k. The possible values of k are [0, 1, 5, 10, 20, 40, 80] sub-frames or correspondingly [0, 2, 10, 20, 40, 80, 160] msec.

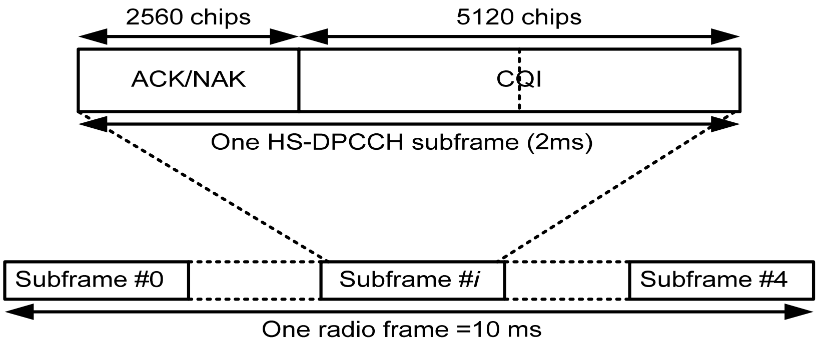

Figure 6-2: HS-DPCCH frame structure.

The Enhanced CQI reporting (E-CQI) scheme described at Release 6 [95] extends Release 5 specifications by introducing additional CQI reports during periods of downlink activity. As shown in Figure 6.4 the additional CQI reports are transmitted with every packet acknowledgment (ACK) (and/or Non-ACK). The aim of the "Enhanced CQI Reporting" is to use longer report cycles compared to the periodic CQI scheme and increases the number of the CQI reports when it is needed (i.e. when the downlink activity increases).

6.3.2 Problem Statement

It is obvious that the shorter the report cycle is the more advantageous for the AMC procedure of HSDPA is, as it provides better adaptation to the variations of the wireless channel. However, at the same time the frequent CQI signaling increases uplink interference and thus decreases the average UE throughput as well as the achievable energy-per-bit to interference (E_b/I_0) ratio at the reverse link. The "Enhanced CQI Reporting" scheme of [95] is a step towards the right direction, however it will still perform as a periodic scheme in the case of constant bit rate (CBR) services such as VoIP or even during the bursty periods of variable bit rate (VBR) services such as FTP.

In this Chapter we propose an improved CQI reporting scheme which aims to reduce the required CQI signaling even when the downlink data activity is high by employing a CQI prediction method. According to this scheme, Node B predicts all the intermediary CQI reports between two subsequent CQI reports by utilizing a FSMC model of the wireless channel. Thus, as shown in Figure 6.5, in our proposed mechanism a number of CQI reports can be predicted instead of transmitted.

The proposed prediction-based CQI reporting scheme (P-CQI) is based on the optimum periodic CQI feedback scheme [96] which has the minimum reporting cycle of 2ms. Therefore, every 2ms a CQI report is either transmitted by the UE or predicted at the Node B. By increasing the number of intermediary CQI predictions the actual reporting cycle is increased and consequently CQI transmissions are decreased. Thus, by employing P-CQI, a significant interference reduction in the uplink can be achieved.

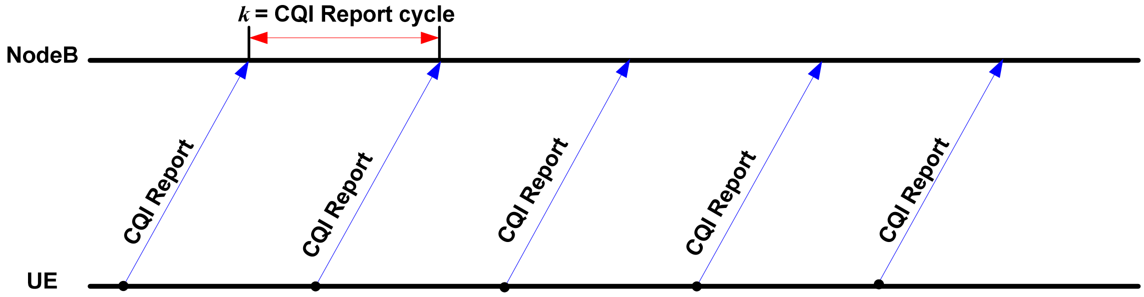

Figure 6-3: Periodic CQI reporting scheme.

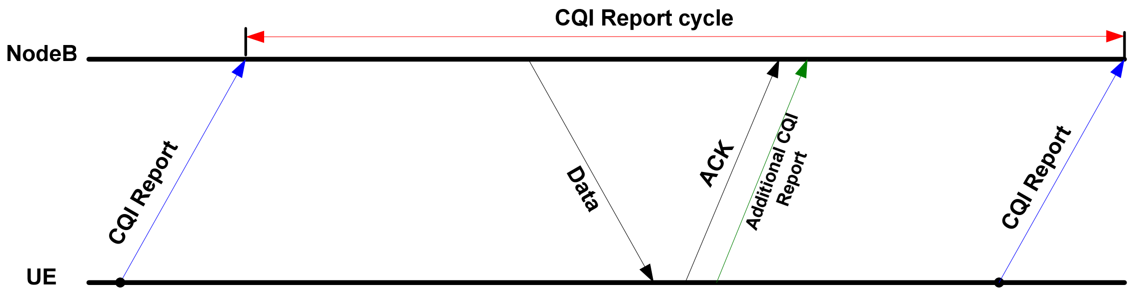

Figure 6-4: Enhanced CQI reporting scheme.

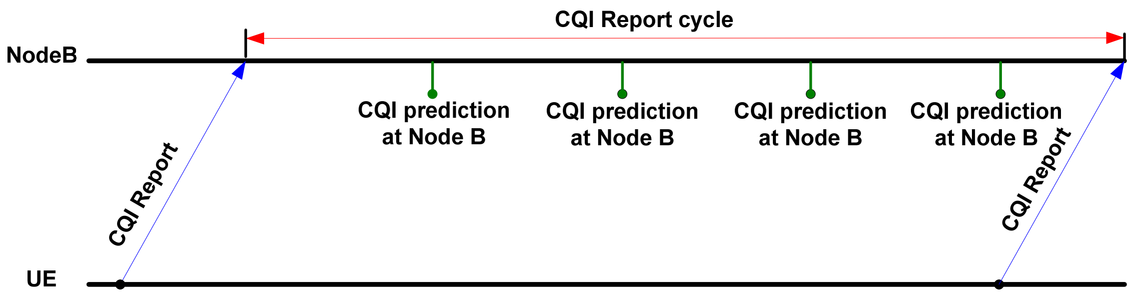

Figure 6-5: Prediction-based CQI reporting scheme.

6.3.3 Uplink interference analysis

By adopting the simplified interference analysis presented at [93] we can measure the benefit of using P-CQI. Under this model, the energy-per-bit to interference (E_b/I_0) requirement of uplink DPCCH and HS-DPCCH can be approximately calculated as follows:

$$E_b/I_0 = \frac{E_b \times R \times SF}{E_b \times R \times \left[M_u - 1 + M' \right]} \tag{6.1}$$

where SF=256 is spreading factor for both DPCCH and HS-DPCCH channels, R is the data transmission rate, M_u -1 denotes the interference of uplink DPCCH and M' denotes the interference of uplink HS-DPCCH due to the transmission update of CQI reports and equals to the average number of the CQI reports that Node B receives in one subframe.

Therefore, for the uplink interference evaluation, we firstly calculate the average number of CQI reports that Node B receives simultaneously in one subframe for both periodic CQI reporting scheme and the proposed scheme (P-CQI) as follows:

In case of periodic CQI reporting scheme, M' can be calculated as:

$$M' = \frac{M_u}{k} \tag{6.2}$$

In case of proposed P-CQI reporting scheme, M' is given by

$$M' = \frac{M_u}{k+1} \tag{6.3}$$

where k is the report cycle.

Considering a fixed number M_u=30 of users in the cell we calculate an approximate estimation of the signal reception quality gain at the Node B. Figure 6.6 shows how the gain of the E_b/I_0 ratio increases as the ratio of predicted CQI reports to the total number of CQI reports, increases. The Periodic-CQI scheme with a reporting cycle of 2ms is used as a reference base.

As we can see in Figure 6.6 the uplink reception quality improves as the number of CQI predictions increases. Although a high ratio of CQI predictions may be impractical in a real system, as prediction accuracy decreases when the number of predicted samples is increased, we can conclude from Figure 6.6 that even by predicting, and thus avoiding, the transmission of just one every two CQI reports we can achieve a significant gain of approximately 1.5 dB.

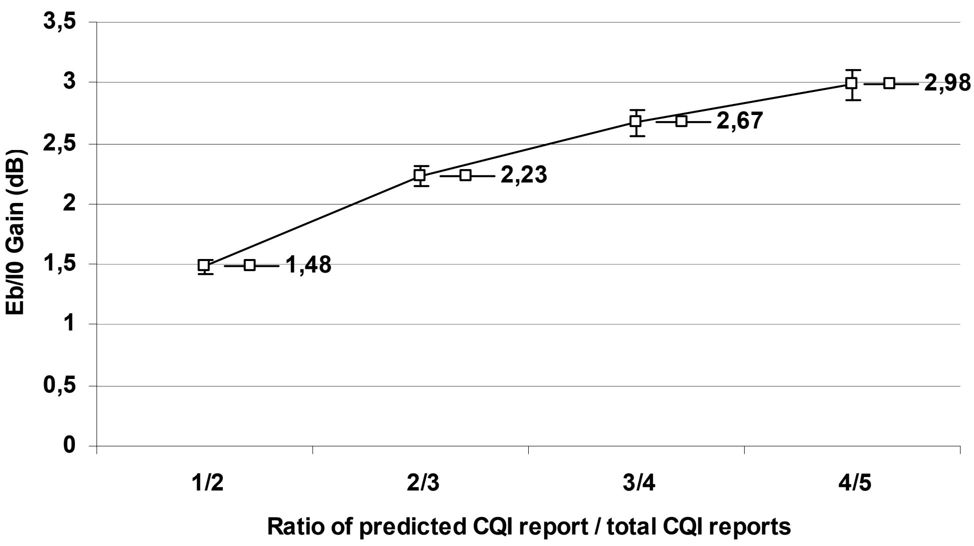

Figure 6-6: E_b/I_0 gain vs. increasing ratio of CQI predictions.

6.4 Prediction-based CQI Reporting Scheme

The CQI reports indicate the requested transport format by the UE and thus they reflect the current channel conditions. Therefore, they can be interpreted to SINR measurements. The obtained SINR values are then used by the P-CQI scheme to predict, through a FSMC model, the next state of the wireless channel and therefore the next CQI. Consequently, we can predict the next CQI at the Node B reducing thus the number of the needed CQI reports.

6.4.1 SINR respective to CQI value

As we mentioned before, the CQI report is not necessary equal to the SINR received at the Node B. However, by using the prediction method we utilize the HS-DSCH SINR value which reflects the varying of the wireless channel conditions and then can be translated at Node B to the respective CQI value (i.e. TFRC). The HS-DSCH SINR, experienced by UE i from a set of active UEs, M_u, in a serving HSDPA cell can be expressed as [61]:

$$SINR_i = \frac{SF_{16} \times P_{r,i}}{\alpha \cdot P_{own,i} + P_{other,i} + P_N} \qquad i \in M_u \tag{6.4}$$

where SF_{16} is the spreading factor (SF=16) of the assigned HS-PDSCH code, $P_{r,i}$ is the received power in HS-DSCH channel for UE i, $P_{own,i}$ is total interference from the serving HSDPA cell, $P_{other,i}$ is the total interference from the other cells, P_N is the thermal noise and α is the downlink code orthogonality factor.

6.4.2 Defining the FSMC model of the wireless channel conditions

Let γ_i be a random variable denoting the value of the received $SINR_i$ of user i in a Rayleigh multi-path fading environment, which is proportional to the square of the signal envelop. The probability density function of γ_i can be expressed as in [70].

$$p(\gamma_i) = \frac{1}{\gamma_0} e^{-\gamma_i / \gamma_0} \tag{6.5}$$

where γ_0=E{γ_i}is the mean value of SINR.

According to the FSMC model, K+1 increasing SINR values $\Gamma_0 = 0 < \Gamma_1 < \Gamma_2 < \cdots < \Gamma_K = \infty$ define K states that describe the channel condition as follows: the wireless channel of user connection i is considered to be in state S_k if the measured γ_i lies in the interval $\{\Gamma_k, \Gamma_{k+1}\}$. Assuming that the channel fades slowly with respect to the CQI feedback report cycle and the Doppler shift f_d in the carrier frequency f_c, is f_d=vf_c/c, then, we can determine the SINR threshold values by using the equal probability method as in Chapter 4, subsequently the Markov steady state and transition probabilities can be calculated. We also assume that the channel remains in one state at each HSDPA subframe (2ms) and then the transition probability between non-adjacent states is very small. Therefore we can assume that transitions happen at the end of the HSDPA subframe and occur only between adjacent states.

6.4.3 Prediction of the next CQI

The CQI prediction is based on the constructed FSMC model of the wireless channel. Given the current state of the channel which is computed by a CQI report, the next state *m* is the one in which the Markov chain will transit to with the highest state transition probability:

$$p_{k,m} = \max(p_{k,k}, p_{k,k+1}, p_{k,k-1}) \tag{6.6}$$

More future states may be predicted in the same manner using the same calculated transition probabilities. The SINR value is then translated at the Node B to the respective CQI value (i.e. TFRC). The state transition probabilities are updated periodically based on the real SINR levels received from the mobiles each time the real CQI measurements are collected. The steps involved in the prediction procedure are illustrated in Figure 6.7.

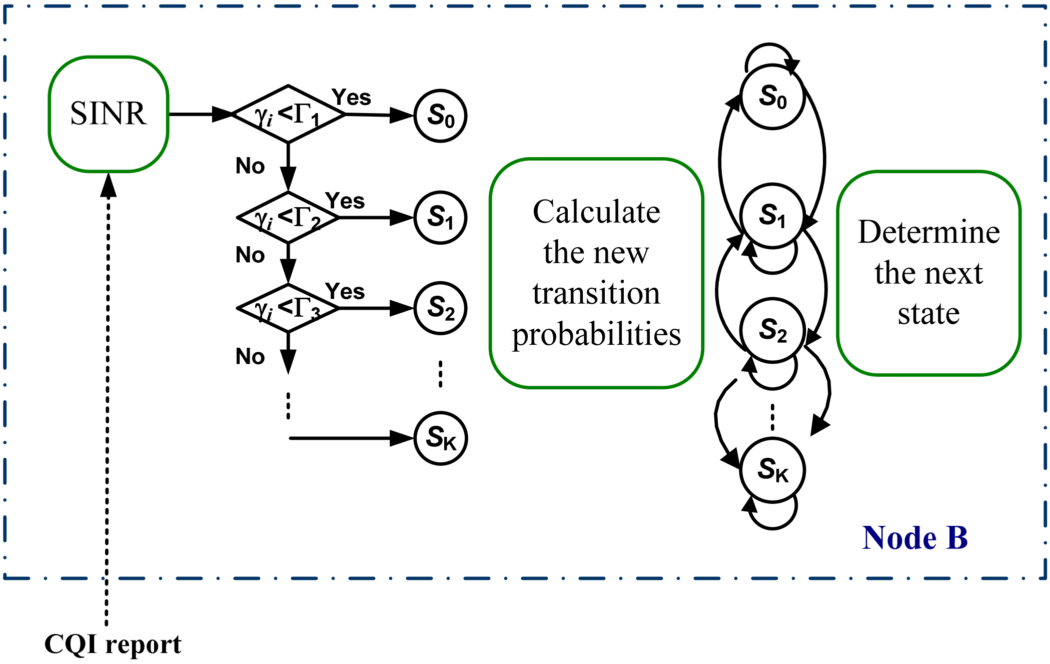

Figure 6-7: CQI prediction method.

6.5 Node B Packet Scheduling

For the packet scheduling we employ the scheduler presented at Chapter 5 and adapt it to HSDPA. Thus, the scheduling period T_s is decreased to 2ms while multi-code operation and AMC are employed.

6.5.1 Priority sorting

According to this scheme a priority P_i that can be calculated by the following equation is assigned to the traffic flow of user i:

$$P_i = \frac{T_{Q,i}(nT_s)}{T_{th,i}} \times \left(1 - P_{e,i}\right) \qquad i \in M_u, \qquad n = 0,1,2,... \tag{6.7}$$

where $T_{Q,i}$ is the delay of the Head Of Line (HOL) packet in the queue during the n-th scheduling period, $T_{th,i}$ is the delay threshold of packets for the specific service and $P_{e,i}$ is the bit error probability during the next frame determined by the state of the wireless channel which is predicted by the FSMC model of the channel. If k is the predicted next state of the wireless channel of user i, $P_{e,i}$ is given by Equation (6.8) [71]:

$$P_{e,i} = \frac{1}{\pi_k} \int_{\Gamma_k}^{\Gamma_{k+1}} p_{em}(\gamma_i) \cdot p(\gamma_i) d\gamma_i \tag{6.8}$$

where $p_{em}(\gamma_i)$ denotes the error probability for a specific modulation scheme.

In HSDPA the transmitted signal is modulated with either QPSK or 16QAM. The $p_{em}(\gamma_i)$ for the QPSK and 16QAM is given respectively by the following equations [97]:

$$p_{em}^{QPSK}(\gamma) = \frac{1}{2} erfc(\sqrt{\gamma}) \tag{6.9}$$

$$p_{em}^{16QAM}(\gamma) = \frac{3}{8} erfc(\sqrt{\tfrac{2}{5}\gamma}) - \frac{9}{64} erfc^2(\sqrt{\tfrac{2}{5}\gamma}) \tag{6.10}$$

6.5.2 Rate allocation

The flows are sorted in decreasing order of their priorities and the scheduler assigns to the highest priority connection the maximum TFRC for the current subframe according to the requested (or predicted) CQI and the available HSDPA capacity. The rate allocation procedure continues with the next connection in the sorted list until either (a) all the connections of the list are examined and all their respective queued packets are scheduled for transmission during the next subframe, or (b) all the available code capacity has been allocated.

6.6 Performance Evaluation

The performance of the proposed P-CQI reporting scheme is evaluated via event driven simulation. Then we evaluate the effect of the CQI reporting schemes in AMC efficiency.

6.6.1 Simulation Model

The simulation model is based on the 3GPP standards. Table 1 gives the major simulation parameters. We consider a cell with a radius of 1km. Node B is located at the centre of the cell. The session arrival process is modeled by a Poisson distribution, while the session duration is exponentially distributed with equal mean. The traffic load increases by increasing the number of users in the cell. At the downlink the user's data are transmitted through HS-DSCH while on the uplink the HS-DPCCH is employed for the transmission of the signaling data.

For each connection the traffic is assumed to arrive according to an "ON-OFF" model during the duration of the session. As long as the connection is in the "OFF" state, it has no arrivals. While in the "ON" state a batch of N packets arrives per timeslot. N is uniformly distributed between N_L and N_H, where $N_L, N_H \in R^+$. A packet is defined as the amount of bits that can be received during one timeslot at the lowest available rate R. The probability P_{ON} of being in the ON state, as well as the N_L and N_H are predefined for each connection.

The initial location of each UE is randomly distributed in the cell, the directions of movement of the users are uniformly distributed, while their velocity is also uniformly distributed in a [0, 3]km/h interval. The macro cell propagation model proposed in [64] is adopted for calculating the path loss at distance d_i (km) from the Node B. The modeling of the wireless channel is performed through a four state FSMC. The equal probability method (EPM) is used to determine the Markov steady state probabilities and by this means the transition probabilities. For all services the $SINR_T$ is equal to 7.9dB and $SINR_{\min}$ is equal to 3dB respectively.

Cell radius	1km
Max. number of UEs, M_u	30
TTI(subframe) T_s	2ms
Spreading factor (SF_{16})	16
Carrier frequency	2GHz
Slow fading model	Rayleigh-fading
Propagation model: $L(d_i)$	128.1+37.6 log (d_i)
Channel model	4 States Markov chain
$SINR_T$, $SINR_{\min}$	7.9dB, 3dB
BER_T	0.1
N_L and N_H	4 and 6
P_{ON}	1
Delay threshold (T_{th})	100ms

Table 6-1: Simulation Parameters.

6.6.2 Effect of the reporting scheme in AMC

If the CQI information is not accurate the AMC operation of HS-DSCH cannot be efficient. As a consequence, this affects various metrics, such as average Bit Error Rate, average packet delay and throughput. In the following, we evaluate the proposed P-CQI scheme by measuring the effect of prediction on the above metrics.

Considering the worst case where E-CQI performs as the Periodic-CQI due to high downlink activity, we assume services with high P_{ON} probability. Furthermore, given that the optimum performance of AMC is achieved when the CQI reporting is as frequent as possible we use the Periodic-CQI feedback scheme with a report cycle of 2ms as a reference. The evaluated P-CQI scheme has a report cycle of 2msec with a CQI prediction ratio of ½ (P-CQI_2ms_1/2). Hence, one every two CQI reports is predicted and the required CQI signaling is reduced by half corresponding to an actual report cycle of 4ms. The Periodic-CQI feedback scheme with a report cycle of 4ms, which causes the same uplink interference as P-CQI_2msec_1/2, is also evaluated for comparison.

As we can see at Figure 6.8 the average BER for the Periodic-CQI_4ms is significantly higher than that of P-CQI at medium to high traffic loads. On the other hand, the performance of P-CQI is close to the optimum performance of the Periodic-CQI_2ms at all traffic loads. The difference between P-CQI and Periodic-CQI_2ms is due to the fact that the CQI prediction procedure cannot always be accurate. Therefore, P-CQI can achieve performance comparable to the performance of the optimum Periodic-CQI_2ms while produces as low interference as the Periodic-CQI_4ms.

This conclusion can also be verified by Figure 6.9 and Figure 6.10 where the average packet delay and the average cell throughput are shown respectively. In both figures the performance of P-CQI is very close to the optimum performance while at the same time outperforms Periodic-CQI_4ms especially at medium to high traffic loads.

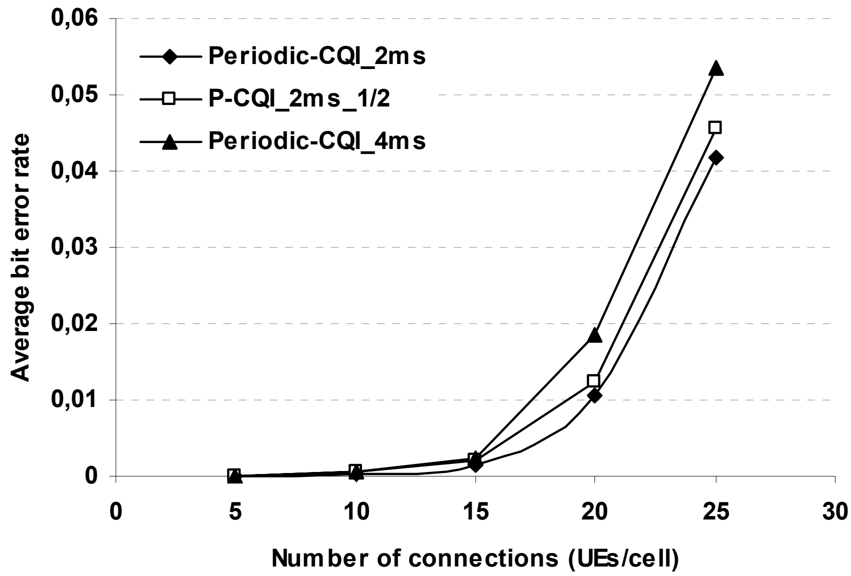

Figure 6-8: Average error rate under Periodic-CQI and P-CQI schemes.

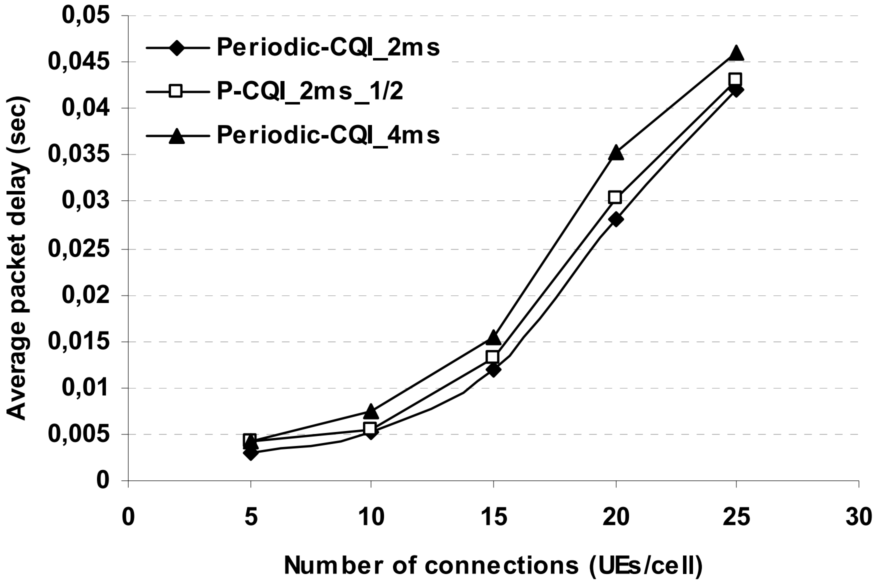

Figure 6-9: Average packet delay under Periodic-CQI and P-CQI schemes.

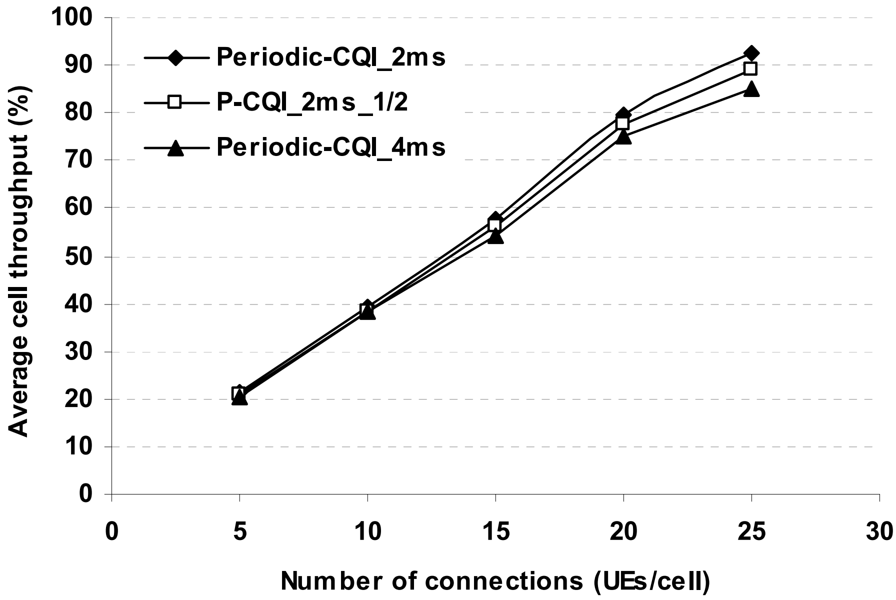

Figure 6-10: Average cell throughput under Periodic-CQI and P-CQI schemes.

6.7 Concluding Remarks

In this Chapter, a prediction-based CQI reporting scheme (P-CQI) in HSDPA system is proposed. P-CQI has better performance compared with a Periodic-CQI scheme which causes the same uplink interference. Due to the prediction of a number of CQI reports the P-CQI scheme reduces significantly the required CQI signaling while at the same time has a performance near to optimum. As a consequence, the reception quality at the uplink is increased.

Furthermore, the efficiency of Adaptive Modulation and Coding procedure in HSDPA depends on the frequency of the CQI reports transmitted by the UE to Node B. To measure the accuracy of the AMC based the proposed P-CQI scheme. We have measured the AMC efficiency by evaluation the Node B scheduling algorithm based on the P-CQI scheme. The efficiency of the proposed scheme in terms of Bite Error Rate, packet delay and throughput is evaluated via simulations. Our results have shown that, P-CQI scheme achieves performance comparable to the performance of the optimum Periodic-CQI_2ms while produces as low interference as the Periodic-CQI_4ms.

Chapter 7 - Conclusions and Future Research

This Chapter provides a summary of the works presented in this thesis and outlines a few directions for future research.

7.1 Conclusions

Due to the great success of wireless multimedia applications and wireless mobile communications, there has been a dramatic demand for wireless data access. To fulfill such extensive demands for wireless services, it is of great importance to develop new techniques for optimization and improving the spectrum efficiency of the existing mobile wireless networks. We note that within a layered architecture, it is possible to yield significant gains, if the system optimizes the performance by making use of the interaction across different protocol layers. Thus in this thesis we have focused on employing cross-layer techniques for resource allocation in mobile wireless networks; and a main objective is to improve the system performance by incorporating the information from the physical layer and MAC layer into the design of resource management.

We have studied data communications in the downlink of WCDMA systems. We first proposed a prediction criterion based on realistic SIR measurements. Then we made use of this prediction to propose a cross-layer framework for efficient packet scheduling which is able to accurately identify the connections which can be served with the required QoS. In particular with the introduction of the cross-layer framework smaller average packet delays, shorter queue lengths and increased bandwidth utilization are achieved as indicated in our results. Therefore, the proposed framework significantly improves the efficiency of the packet scheduling procedure in WCDMA based networks. Furthermore, we proposed a new scheduling algorithm, which serves the connections not only according to their delay sensitivity, but also according to the predicted state of their wireless channel. Also, we measured the major issues of wireless scheduling algorithms such as fairness and QoS differentiation. The simulation results confirmed that the features of the wireless scheduling algorithm are not effect by introducing the cross-layer technique into the scheduling disciplines.

Next we proposed radio resource management scheme for handover provisioning QoS in 3G WCDMA networks. To realize QoS provisioning in these networks, there exist many challenges: 1) Limited radio resource; 2) Time-varying wireless channel; 3) Limited battery power supply of the users; 4) Interference-limited capacity; 5) Grade of Service (GoS) such as new blocking and handover dropping probabilities. In our work, we have taken into account most these parameters. We introduced a new guard code scheme in WCDMA systems which employ OVSF codes as a channelization codes. The scheme belongs to the well known family of guard channel schemes and reserves some code capacity to prioritize the HO calls over new calls. Although this scheme belongs to the well known family of guard channel schemes, it is adapted to the OVSF code tree structure. The numerical analysis of this scheme performance is analyzed by using Markov chain with multiple transitions. The performance evaluation of the proposed scheme is evaluated by using different code thresholds reservation, different traffic patterns and different HO traffic loads. The performance metrics such as new call blocking probability, HO failure rate, and the code utilization are discussed.

Then a new design for effective combination between a CAC mechanism and a traffic scheduling discipline at MAC layer of the WCDMA-3G networks is proposed. This design is based on the proposed cross layer technique. We have proposed a CAC scheme at the call level which is based on the guard code reservation in the OVSF channelization codes. The calculation of the code reservation threshold and adaptation to the OVSF codes are given. Then we proposed a delay-aware scheduler discipline namely, delay driven traffic scheduling (DDS) which exploits the information from the physical layer in order to avoid erroneous packet transmission and increase system performance. The scheduling procedure is not only based on the delay sensitivity but also based on the predicted error probability of each user in the cell. Compared to the existing literature of radio resource management schemes for handover provisioning which focuses only on provisioning the QoS requirements of handover by using efficient CAC mechanism at call level, our scheme has a capability of efficiently utilizing the available radio resources and also supporting the QoS provisioning at both call and packet levels. The proposed scheme introduced an efficient integration between the CAC scheme and the traffic scheduling algorithm at the MAC layer for provisioning QoS. We have also investigated the performance of the CAC scheme at different code reservation thresholds. However, when the total available capacity is shared among the users, the performance evaluation of the scheduler is evaluated in terms of the average packet delay, average queue size and the code utilization.

Finally, we turned our attention to HSDPA-based networks. Specifically, we studied the performance enhancement of HSDPA system. We proposed an improved CQI reporting scheme based on our prediction method. The proposed scheme (P-CQI) can achieve higher throughput while at the same reduces the overhead signaling between the UEs and Node B. Firstly, we evaluated the gain in the uplink reception quality by employing a simplified model for HS-DPCCH uplink interference evaluation when the P-CQI reporting scheme introduced to HSDPA system. Furthermore, AMC of HSDPA operation depends on CQI reporting scheme. Thus we introduced a packet scheduling discipline based on P-CQI reporting scheme in order to measure the system performance. The simulation results have shown the efficiency of the proposed scheme. As our results indicated, the main benefit of the proposed scheme is the improvement of the uplink reception quality and at same time achieves a comparable performance compared to the optimum periodic CQI reporting scheme.

7.2 Future work

We outline a few directions for further research in the area of cross-layer techniques, in the next-generation wireless systems as follows:

7.2.1 More and More Cross-Layer Techniques

7.2.1.1 4G-based cross-layer techniques

In our work we have considered the cross-layer technique among the physical layer and the MAC layer. However, new network architectures and core technologies will be required to support “4G” high data-rate connections from 2 to over 100Mb/s with various required QoS. Provisioning of various real-time multimedia services to mobile users is the main objective of the next generation wireless networks that will be expected to seamlessly integrate with the Internet backbone. Therefore, the overall performance improvement in these networks can be achieved by cross-layer analysis, optimization and design taking physical layer, MAC layer, network layer and transport layer into account. Furthermore, these new technologies will possibly require cross-layer techniques, distributed CAC and distributed scheduling schemes for radio resource management in such networks. HSPA is the first step towards the 4G mobile networks; however, the existing HSPA needs efficient enhancements based on cross-layer design to solve some technical problems such as adaptive scheduling algorithm for MAC layer, handover strategy for network admission control, and algorithms for the avoidance of erroneous TCP at the transport layer.

7.2.1.2 WiMAX-based cross-layer techniques

The development of next-generation wireless systems (e.g., the fourth-generation (4G) mobile cellular systems, IEEE 802.11n, etc.) aims to provide high data rates in excess of 1Gbps. However, the prediction of the future 4G solution is quite hard. The reason for this is simply the limited knowledge of the future environment in which 4G will be presumably implemented. Moreover, assumptions are often coupled to existing or preceding technologies. One distinctive fact is that 4G will be implemented among various competing access technologies.

Nowadays, one promising access technology named Worldwide Interoperability for Microwave Access (WiMAX) is introduced in the wireless markets. The most important feature of mobile WiMAX system is not only providing high coverage area but also supporting high capacity. A mobile WiMAX system which is based on IEEE 802.16 standard consists of at least one base station (BS) and one or more subscriber stations (SSs). The PHY layer of the standard were designed in such a way that different burst profiles and adaptive modulation and coding (AMC) are supported based on the channel conditions. Although the IEEE 802.16 standard provides the outlines for implementing the link adaptation in which different AMC schemes can be used based on the channel conditions, it does not give the procedures for enhancement. Since there is also a possibility that the wireless channel deteriorates or becomes unavailable at a certain period of time, the scheduling algorithm at the MAC layer, which has been left undefined in the standard, must have a knowledge of the channel condition so that only a connection with a good channel condition is scheduled for transmission to achieve multi-user diversity gain. At the same time however, the QoS requirements of each connection should be fulfilled.

It is therefore clear that cross-layer interactions between the MAC and PHY layer are needed so that the limited wireless resource could be optimally used while satisfying the QoS requirements. By deploying interactions between both MAC and PHY layers, the optimum parameters can be obtained and eventually performance gain should be achieved. Therefore, our future work may focus on applying our proposed cross-layer technique in mobile WiMAX standard to improve the system performance.

7.2.1.3 Integration WiMAX and Mobile Networks

In the near future the next-generation wireless networks will be an effective combination of the mobile WiMAX system together with the 4G mobile networks. Adaptive resource reservation and bandwidth adaptation in radio resources management algorithms (CAC and scheduler) are the main concern in such networks. Many challenges will arise in such integration due to the different carrier frequency, inter-mobility, capacity and coverage area. We believe that cross-layer techniques will have a great influence in these networks. The deployment of cross-layer techniques will make the integration between WiMAX and cellular mobile networks such as 3G and 3.5G possible. One simple solution is deploying the mobile WiMAX network as an overlay system to cellular mobile networks which have the same physical characteristics. In addition, deploying hybrid scheduling algorithms instead of fixed schedulers may be a sudation; the scheduling algorithm should be adaptively allocated the radio resource in both networks. Thus, integration of WiMAX based-cross layer techniques in next wireless generation networks (NWGNs) is becoming an emerging research area and should be carefully examined and studied.

Appendix I - Generation Function of the Number of States

In this Appendix we define the generation function to determine the total number of states in the OVSF code tree. Depending on the possible transmission rates in the system, we can obtain all the feasible occupancy states in the OVSF code tree by using a generation function as follows [81]:

$$f(s) = \frac{1}{1-s} \times \frac{1}{1-s^2} \times \cdots \times \frac{1}{1-s^g} \tag{I.1}$$

If the number of allowable transmission rates in the OVSF code tree are $1R$, $2R$, $4R$, and $8R$ then g takes the values 1, 2, 4, and 8 respectively. The number of states which have capacity of c can be obtained by summing the coefficients of s^c at $f(s)$. We applied Taylor theorem to calculate these coefficients as follows. According to the feasible transmission rates $1R$, $2R$, $4R$, and $8R$, $f(s)$ can be expressed as follows:

$$f(s) = \frac{1}{(1-s)(1-s^2)(1-s^4)(1-s^8)} \tag{I.2}$$

By calculating the coefficients of $f(s) = \frac{f^n(s_0)}{n!}$ where, $f^n(s_0)$ is the n^{th} derivative function of $f(s)$ at $s = s_0 = 0$, then $f(s)$ can be written as follows:

$$f(s) = 1 + s + 2s^2 + 2s^3 + 4s^4 + 4s^5 + 6s^6 + 6s^7 + 10s^8 + 10s^9 + \cdots \tag{I.3}$$

The summation of all the coefficients of s^c, gives the total number of states. For example, when C_T =8, then the summation coefficients of s^8 is 36. If C_T =16, then the summation of the coefficients of s^{16} will be 201 which equals to total number of states in the OVSF code tree.

Le us study this example when the C_T =16 and C_H =6, in order to obtain the flow balance equations of transition probabilities at Equation (5.2). If the system is at state (0112) with green color as shown in Figure I.1, then we can determine the feasible transitions states into and out this state based on the C_T and C_H values. As shown in the Figure I.1, the red states refer to the possible states which transit into state (0112) and the blue states refer to the possible states when transiting from the state (0112). Then we can obtain the flow balance equations at state (0112) by using the Markov chain transition diagram as shown at Figure 5.3. Manual analysis to obtain the steady state probabilities at each state was impossible. Thus we employed a C++ program to produce the steady state equations before we have fed these equations to Mathematica to solve them.

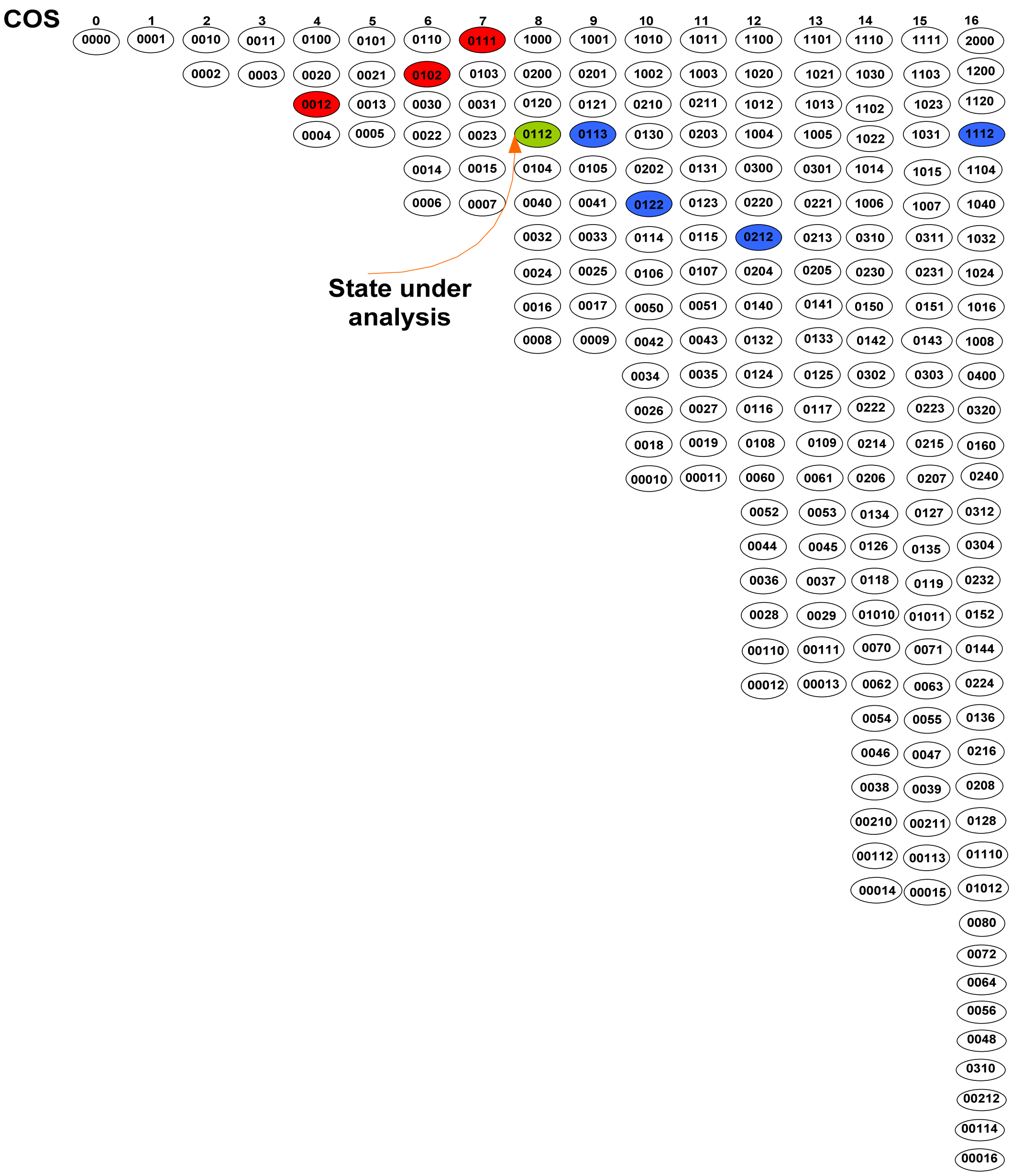

Figure I-1: Total number of states at C_T =16 and C_H =6.

Bibliography

[1] A. J. Goldsmith and S. B. Wicker, "Design challenges for energy-constrained ad hoc wireless networks," IEEE Wireless Communications, Vol. 9, 2002, pp. 8–27.

[2] T. S. Rappaport, P. C. Karlsson and S. Shakkottai, "Cross-layer Design for Wireless Networks," IEEE Communications Magazine, Vol. 41, October, 2003, pp. 74-80.

[3] V. Kawadia and P. R. Kumar, "A Cautionary Perspective on Cross Layer Design," IEEE Wireless Communications, Vol. 12, Feb. 2005, pp. 3–11.

[4] N. Nikaein and C. Bonnet, "A Glance at Quality of Service models in Mobile Ad Hoc Networks," in Proceedings. DNAC 2002, France, Paris, 2002.

[5] I. Koutsopoulos Leandros Tassiulas and Ulas C. Kozat, "A framework for cross-layer design of energy-efficient communication with QoS provisioning in multi-hop wireless networks," in Proceedings IEEE INFOCOM 2004, Vol. 23, March 2004, pp. 1447-1457.

[6] V. Srivastava and M. Motani, "Cross-Layer Design: A Survey and the Road Ahead," IEEE Communications Magazine, December 2005, pp. 112–119.

[7] Simon Dobson, "Putting meaning into the network: some semantic issues for the design of autonomic communications systems," in Proceedings of the 1st IFIP Workshop on Autonomic Communications, Vol. 3457, of LNCS, Springer Verlag, 2005, pp. 207-216.

[8] G. L. Stuber, *Principles of Mobile Communications*, Kluwer Academic Publishers, 1996.

[9] E. N. Gilbert, "Capacity of a burst-noise channel," Bell Syst. Tech. J, Vol. 39, Sept. 1960, pp. 1253 –1265.

[10] E. O. Elliott, "Estimates of error rates for codes on burst-noise channels," Bell Syst. Tech. J., Vol. 42, Sept. 1963, pp. 1977–1997.

[11] W. Turin, *Digital Transmission Systems: Performance, Analysis and Modeling* 2nd edition, New York: McGraw-Hill, 1999.

[12] H. Fattah and C. Leung, "An Overview of Scheduling Algorithms in wireless multimedia networks," IEEE Wireless Communications, Vol. 9, Oct. 2002, pp. 76-83.

[13] Dapeng W. and Negi R, "Effective capacity: a wireless link model for support of quality of service," IEEE Transactions on Wireless Communications, Vol. 2, July 2003, pp. 630 – 643.

[14] T. S. Rappaport, *Wireless Communications: Principles & Practice*, Englewood Cliffs, NJ: Prentice-Hall, 1996.

[15] W. C. Jakes, *Microwave Mobile Communications*, IEEE Press, Piscataway, NJ, 1993.

[16] ETSI TR, Universal Mobile Telecommunications System (Umts), Selection procedures for the choice of radio transmission technologies of the UMTS, ETSI TR 101 112, UMTS 30.03 version 3.2.0.

[17] Antti Toskala and Harri Holma, *WCDMA for UMTS radio access for third generation mobile communications*, John. Wiley, 2001.

[18] W. Turin and R. Van Nobelen, "Hidden Markov Modeling of Flat Fading Channels," IEEE Journal on Selected Areas in Communications, Vol. 16, December 1998, pp. 1809–1817.

[19] H. S. Wang and N. Moayeri, "Finite-State Markov Channels—A Useful Model for Radio Communications Channels," IEEE Transactions on Vehicular Technology, Vol. 44, 1995, pp. 163–171.

[20] S. Sivaprakasam and K. S. Shanmugan, "An Equivalent Model for Burst Errors in Digital Channels," IEEE Transactions on Communications, Vol. 43, February/March/April 1995, pp. 1347–1355.

[21] H. S. Wang and P.-C. Chang, "On Verifying the First-Order Markovian Assumption for a Rayleigh Fading Channel Model," IEEE Transactions on Vehicular Technology, Vol. 45, May 1996, pp. 353–357.

[22] O. E. Kelly, F. Babich, and G. Lombardi, "Generalized Markov Modeling for Flat Fading," IEEE Transactions on Communications, Vol. 48, April 2000, pp. 347–351.

[23] R. R. Rao, M. Zorzi, and L. B. Milstein, "A Markov Model for Block Errors on Fading Channels," in Proceedings PIMRC '96 Conference Vol. 3, October 1996, pp. 1074 - 1078

[24] M.I Mushkin and I.Bar-David, "Capacity and Coding for the Gilbert-Elliott Channels," IEEE Trans. Inform. Theory, Vol. 6, November 1989, pp.1277-1290.

[25] Shengli Z., Georgios B. and Giannakis Qingwen L., "Cross-Layer Scheduling with Prescribed QoS Guarantees in Adaptive Wireless Networks," IEEE Journal on Selected Areas in Communications, Vol. 23, May 2005, pp. 1056-1066.

[26] X. Wang, G. Giannakis and Q. Liu, "A Cross-Layer Scheduling Algorithm With QoS Support in Wireless Networks," IEEE Transaction on Vehicular Technology, Vol. 55, MAY 2006, pp. 839 - 847.

[27] Yaxin Cao and Victor O.K. Li, "Scheduling Algorithms in Broad-Band Wireless Networks," IEEE Proceedings Vol. 89, January 2001, pp. 76-87.

[28] S. Keshav and S. Shenker "A. Demers, Analysis and Simulation of a fair queueing algorithm," in Proceedings ACM SIGCOMM'89, 1989, pp. 3-12.

[29] Norival R., Figueira and Joseph Pasquale, "An upper bound on delay for the Virtual Clock service discipline," IEEE/ACM Transactions on Networking, Vol. 3, August 1995, pp. 399 – 408.

[30] K. Shin, D. Ferrari and D. Kandlur, "Real-time Communication in multi-hop networks," in Proceeding IEEE 11th Distributed Computer System, May 1991, pp. 300-307.

[31] Abhay K. Parekh and Robert G. Gallager, "A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks: The Single-Node Case," IEEE/ACM Transactions on Networking, Vol. 1, June 1993, pp. 344-357.

[32] V. Bhargavan, R. Srikant and S. Lu, "Fair Scheduling in Wireless Packet Networks," IEEE/ACM Trans on Networking, Vol. 7, August 1999, pp. 473-489.

[33] D. A. Levine, I. Joe and I. F. Akyilidiz, "A Slotted CDMA Protocol with BER Scheduling for Wireless Multimedia Networks," IEEE/ACM Trans. on Networking, Vol. 7, April, 1999, pp. 146-158.

[34] J.-K. Lain, J.-H.Wen, and Y.-W. Lai, "Performance evaluation of a joint CDMA/NC-PRMA protocol for wireless multimedia communications," IEEE Select. Areas Commun, Vol. 19, Jan. 2001., pp. 95–106.

[35] X. Wang, "Wide-band TD-CDMA MAC with minimum-power allocation and rate- and BER-scheduling for wireless multimedia networks," IEEE/ACM Trans. Network, Vol. 12, Feb. 2004, pp. 103–116.

[36] R. Padovani A. Jalali, and R. Panaaj, "Data Throughput of CDMA HDR a High Efficiency-High Data Rate Personal Comm. Wireless System," in Proceedings IEEE VTC'2000-spring, Vol. 3, 2000, pp. 1854-1858.

[37] A. Annamalai, T.S. Rappaport, R.M.Buehrer and W.H. Tranter, "Wireless Communications: Past Events and a Future Perspective," IEEE Communications Magazine, Vol. 40 May 2002, pp. 148-161.

[38] K. Kumaran, M. Andrews, K. Ramanan, A.L. Stolyar, P. Whiting and R. Vijayakumar, "Providing Quality of Service Over a Shared Wireless Link," IEEE Communications Magazine, Vol. 39 February 2001 pp. 150-153.

[39] K. Kumaran, M. Andrews, K. Ramanan, A.L. Stolyar, R. Vijayakumar and P. Whiting, "CDMA Data QoS Scheduling on the Forward Link with Variable Channel Conditions," Bell Labs Technical Report, April 2000.

[40] S. Shakkottai and A.L. Stolyar, "A Study of Scheduling Algorithms for a Mixture of Real and Non-Real Time Data in HDR," in Proceedings. 17th International Teletraffic Congress, September 2001.

[41] D. Niyato and E. Hossain, "Service differentiation in broadband wireless access networks with scheduling and connection admission control: a unified analysis," IEEE Trans on Wireless Comm., Vol. 6, 2007, pp. 293-301.

[42] H. Zhang, "Service Disciplines for Guaranteed Performance Service in Packet-switching Networks," Proceeding IEEE, Vol. 83, Oct. 1995, pp. 1374–1396.

[43] R. Guerin and V. Peris, "Quality-of-service in Packet Networks: Basic Mechanisms and Directions," Computer Networks, Vol. 31, 1999, pp. 169–189.

[44] P. Stuckmann, *The GSM Evolution: Mobile Packet Data Services*, John Wiley & Sons, Chichester, UK, 2003.

[45] E. Dahlman and et al., "WCDMA-The Radio Interface for Future Mobile Multimedia Communications," IEEE Trans. on Vehicular Technology, Vol. 47, Nov. 1998, pp. 1105 -1118.

[46] 3GPP2 Web Site, http://www.3gpp2.org.

[47] W. Mohr, R. Parsad, and W. Konhauser, *Third Generation Mobile Communication Systems*, Artech House Publisher, Boston - London, 2000.

[48] L. Hanzo and R. Steele, *Mobile Radio Communications Second and Third Generation Cellular and WATM System*, John Wiley and Sons, New York, 1999.

[49] A. Annamalai, M. Zeng, and V. Bhargava, "Advances in Cellular Wireless Communications," IEEE Communications Magazine, Sept. 1999.

[50] 3GPP Technical Specification, Spreading and Modulation (FDD), 3G TS 25.213,Ver. 3.2.0, Mar. 2000.

[51] 3GPP Technical Specification, UE Radio transmission and Reception (FDD), 3G TS 25.101, Ver. 3.2.2, Apr. 2000.

[52] 3GPP Technical Specification, Physical Channels and Mapping of Transport Channels onto Physical Channels (FDD), 3G TS 25.211, Ver. 3.2.0, Mar. 2000.

[53] 3GPP Technical Specification, Multiplexing and Channel Coding (FDD), 3G TS 25.212, Ver. 3.2.0, Mar. 2000.

[54] N. Parameshwar and R. Rajagopalan, "A comparative study of cdma2000 and W-CDMA," in Proceedings Wireless Communications and Systems, 1999 Emerging Technologies Symposium, 1999 pp. 15.11 - 15.19.

[55] 3GPP Technical Specification, Radio Resource Management Strategies, 3GPP TR 25.922 v.5.0.0 Release 5 ETSI TR 125 922.

[56] B. Sarikaya, "Packet Mode in Wireless Networks: Overview of Transition to Third generation," IEEE Communications Magazine, Vol. 38, Sept. 2000, pp. 164 -172.

[57] Kun I. Park, *QoS in Packet Networks*, Springer, Boston, 2005.

[58] H. Jonathan Chao and Xiao Lei Guo, *Quality of Service Control in High-Speed Networks*, John Wiley & Sons, Inc., 2002.

[59] 3GPP Technical Specification, QoS concept and architecture, 3GPP TS 23.107 V5.4, 2002–2003.

[60] W. Zhuang H. Jiang, and X. Shenrman, "Cross-Layer Design for Resource Allocation in 3G Wireless Networks and Beyond," IEEE Communications Magazine, Dec. 2005, pp. 120-126.

[60] W. Zhuang H. Jiang, and X. Shenrman, Cross-Layer Design for Resource Allocation in 3G Wireless Networks and Beyond, IEEE Communications Magazine, Dec. 2005, pp. 120-126.

[61] Harri Holma and Antti Toskala, *HSDPA/HSUPA for UMTS High Speed Radio Access for Mobile Communications*, John Wiley and Sons, 2006.

[62] J. Ruela, G. Carneiro, and M. Ricordo, "Cross-layer Design in 4G Wireless Terminals," IEEE Wireless Communications, Vol. 11, Apr. 2004, pp. 7–13.

[63] Mihaela Van Der Schaar and Sai Shankar, "Cross-Layer Wireless Multimedia Transmission: Challenges, Principles, and New Paradigms," IEEE Wireless Communications, Vol. 12, August 2005, pp. 50-58.

[64] 3GPP Technical Specification, Radio frequency (RF) system scenarios, 3GPP TR 25.942 V6.3.0, 2004–2006.

[65] Thit Minn, Anthony C. Kam, and Kai-Yeung Siu, "Supporting rate guarantee and fair access for bursty data traffic in W-CDMA," IEEE Journal on Selected Areas in Communications, Vol. 19, 2001, pp. 2121–2130.

[66] J. Pérez-Romero, O. Sallent, R. Agustí., and F. Casadevall, "Provisioning multimedia wireless networks for better QoS: RRM strategies for 3G W-CDMA," IEEE Communications Magazine,, Vol. 41, 2003, pp. 100–106.

[67] Yu-Chee Tseng, Chih-Min Chao, and Li-Chunwang, "Dynamic bandwidth allocation for multimedia traffic with rate guarantee and fair access in WCDMA systems," in Proceedings ACM of the 6th MSWiM, San Diego, CA, USA, 2003, pp. 136–139.

[68] Xuemin Shen, Liang Xu, and Jon W. Mark, "Dynamic fair scheduling with QoS constraints in multimedia wideband CDMA cellular networks," IEEE Transactions on Wireless Communications, Vol. 3, 2004, pp. 63–73.

[69] N. Madsen, J. Wigard, P. Gutierrez, I. Sepulveda, and P. Mogensen, "Packet scheduling with QoS differentiation," An International Journal Springer, Wireless Personal Communications, Vol. 23, Kluwer Academic Publishers, 2002, pp. 147–160.

[70] J. G. Proakis, *Digital Communications* 3rd Edition, New York, McGraw-Hill 1995.

[71] Q. Zhang and Kassam S., "Finite-state Markov model for Rayleigh fading channels," IEEE Transactions and Communications, Vol. 47, 1999, pp. 1688–1692.

[72] H.S. Wang and N. Moayeri, "Modeling, capacity, and joint source/channel coding on Rayleigh fading channels," in Proceedings 43rd IEEE VTC, NJ, USA, 1993, pp. 473–479.

[73] Aytac Azgin and Marwan Krunz, "Impact of Channel Modeling on the Performance of Wireless Scheduling Schemes," in Proceeding 58th IEEE VTC'03-Fall, Vol. 5, Orlando, USA, 2003, pp. 3532-3536.

[74] Dimitrios N. Skoutas and Angelos N. Rouskas, "A dynamic traffic scheduling algorithm for the downlink shared channel in 3G WCDMA," in Proceeding IEEE ICC, Vol. 27, Paris, France, June, 2004, pp. 2975–2979.

[75] 3GPP Technical Specification, Physical Layer General Description, 3G TS 25.201, Ver. 3.2.0, Mar. 2000.

[76] D. Niyato and E. Hossain, "Call-Level and Packet-Level Quality of Service and User Utility in Rate-Adaptive Cellular CDMA Networks: A Queuing Analysis," IEEE Transactions on Mobile Computing, Vol. 5, Dec. 2006, pp. 1749-1763,.

[77] C. C. Lee and R. Steele, "Effects of soft and softer handoffs on CDMA system capacity," IEEE Transactions on Vehicular Technology, Vol. 47, Aug. 1998 pp. 830-841.

[78] 3GPP Technical Specification, Handover procedures, 3GPP TS 23.009 V3.10.0 2002-06.

[79] M. Sawahashi, F. Adashi, and K.Okawa, "Tree structured generation of orthogonal spreading codes with different lengths for forward link of DS-CDMA mobile radio," Electronic Letters, Vol. 33, Jan 1997, pp. 27–28

[80] K.Y. Siu and T. Minn, "Dynamic Assignment of Orthogonal Variable Spreading Factor Codes in W-CDMA," IEEE Journal on Selected Areas in Communications, Vol. 18, August 2000, pp. 1429-1440.

[81] Y. C. Tseng and C. M. Chao, "Code Placement and Replacement Strategies for Wideband CDMA OVSF Code Tree Management," IEEE Trans. on Mobile Computing, Vol. 1, Oct.-Dec. 2002, pp. 293-302.

[82] A. Rouskas and D. Skoutas, "OVSF code assignment and reassignment at the forward link of W-CDMA 3G systems," in Proceeding IEEE PIMRC, Vol. 5, Lisbon, Portugal, Sept. 2002, pp. 2404-2408.

[83] B. Li, Y. Fang and X. Cheng, "A Dynamic Multiple-Threshold Bandwidth Reservation (DMTBR) Scheme for QoS Provisioning in Multimedia Wireless Networks," IEEE Trans. Wireless Comm, Vol. 4, March 2005, pp. pp.583-592.

[84] Y. Zhang, E. Regentova and J. Zheng, "Virtual Guard Channel for Handoff Calls in Integrated Voice/Data Wireless Networks," IEEE Communications letter Vol. 10, April 2006., pp. 263-266.

[85] Lee S. and Park S., "Handoff with Dynamic Guard Channels in ATM-Based Mobile Networks," in Proceeding IEEE ICOIN, Tokyo, Japan, January 1998, pp. 21 – 23.

[86] Ideguchi T., Tian X. and Okuda T., "Improvement of the dropping probabilities in mobile communication networks by using guard channels protocol with delay of

new user," in Proceeding IEEE 60th VTC-Fall, Vol. 6, Los Angeles, CA, USA, Sept. 2004, pp. 4558- 4562.

[87] J. Mark, K. Chua and T. Wong, "Multi-level Dynamic Guard Channels for Priority Access in Cellular Systems," in Proceeding NETWORKING 2005, Springer LNCS 3462, 2005, pp. 598 – 609.

[88] S. Kotsopoulos, I. Panoutsopoulos, and V. Tountopoulos, "Handover and New Call Admission Policy Optimization for G3G Systems," Springer Wireless Networks, Vol. 8, July 2002, pp. 381-389.

[89] C.L.P. Chen, B. Wang and Y. Xiao, "Bandwidth Degradation QoS Provisioning for Adaptive Multimedia in Wireless/Mobile Networks," Computer Communication, Vol. 25, 2002, pp. 1153-1161.

[90] V. Kawadia and P. R. Kumar, "QoS-aware Radio Resource Management Scheme for CDMA Cellular Networks Based on Dynamic Interference Guard Margin (IGM)," Computer Networks, Vol. 46, 2004, pp. 867-879.

[91] V. Wong, V. Leung and Y. Fei, "Efficient QoS Provisioning for Adaptive Multimedia in Mobile Communication Networks by Reinforcement Learning," Springer Mobile Networks and Applications, Vol. 11, 2006, pp. 101-111.

[92] K. Meier-Hellstern and E. Alonso, "The use of SS7 and GSM to support high density personal communications," in Proceedings ICC, 1992, pp. 1698-1702.

[93] Noriyuki Fukui, "Study of channel quality feedback in UMTS HSDPA," in Proceedings IEEE 14th PIMRC, 2003, pp. 336-340.

[94] Soo-Yong Jeon and Dong-Ho Cho, "Channel Adaptive CQI Reporting Schemes for HSDPA Systems," IEEE Commun. Letters Vol. 9, May 2005, pp. 432-434

[95] 3GPP Technical Specification, High Speed Downlink Packet Access (HSDPA) enhancements, 3GPP TS25.899 v6.1.0.

[96] 3GPP Technical Specification, High Speed Downlink Packet Access: Physical Layer Aspects, 3GPP TR 25.858 V5.0.0

[97] Hiroshi Harada and Ramjee Prasad, *Simulation and Software Radio for mobile communications*, Artech House, 2002.

Vita

Saied M. Abd El-atty Soliman was born in Menouf, Egypt, on December 15, 1972. After finishing Menouf High School he attended the Faculty of Electronic Engineering of Menoufia University, where he received the five-year Diploma (B.Sc.1995) in Electronic Engineering and Master Degree (M.S. 2001) with emphasis in Communication Engineering. In 1995 and 2001 he worked at the Department of Electronics and Communications Engineering in the Faculty of Electronic Engineering as demonstrator and assistant lecturer respectively. On August 2003, he awarded the National Foundation Scholarship of Greece (IKY) for PhD studies. On February, 2004, he joined the Department of Information and Communication Systems Engineering at University of the Aegean to pursue a PhD. His current research interests include cross-layer design, wireless channel modeling, packet scheduling algorithms and QoS control in WCDMA systems as well as radio resource management in 3G and beyond mobile communication systems. He is the author of several papers in scientific journals and conferences in that field.

List of Publications

Journals:

- Saied M. Abd El-atty, Dimitrios N. Skoutas, Angelos N. Rouskas and George T. Karetsos, "A Cross Layer Scheduling Framework for Supporting Bursty Data Applications in WCDMA Networks", *International Journal of Wireless Personal Communications, WPC*, Springer, 2007.
- Saied M. Abd El-atty, Dimitrios N. Skoutas and Angelos N. Rouskas, "A Guard Code Scheme for Handover Traffic Management in WCDMA Systems" accepted for publication in *International Journal of Wireless Information Networks, IJWIN*, Springer, 2008.
- Saied M. Abd El-atty, Dimitrios N. Skoutas and Angelos N. Rouskas, "Reducing CQI Signaling Overhead in HSPA," accepted for publication in *Journal of Research Letter in Communications*, Hindawi Publishing Corporation, 2008.

Conferences:

- Saied M. Abd El-atty, Dimitrios N. Skoutas and Angelos N. Rouskas, "A Cross-Layer Framework for Efficient Packet Scheduling in Wireless CDMA Networks ," 7th, Seventh IFIP International Conference on Mobile and Wireless Communication Networks IFIP (*MWCN*'05), Marrakech, Morocco, 19-21 September 2005.
- Saied M. Abd El-atty, Dimitrios N. Skoutas and Angelos N. Rouskas, "Handover Provisioning in WCDMA Systems," the Third International Mobile Multimedia Communications Conference (*MobiMedia*'07) Nafpaktos, Greece, 27-29 August 2007.
- Saied M. Abd El-atty, Dimitrios N. Skoutas, Angelos N. Rouskas and George T. Karetsos, "Radio Resource Management for Handoff Provisioning in WCDMA Systems," 18th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (*PIMRC*'07) Athens, Greece, 3 – 7, September 2007.