Fast Scheduling Strategies for WCDMA-High Speed Downlink Packet Access (HSDPA) System

By

Hua Sun, M.Eng

A thesis submitted to

The Faculty of Graduate Studies and Research

In partial fulfillment of

The requirements for the degree of

Master of Applied Science in Electrical Engineering

Ottawa-Carleton Institute for Electrical and Computer Engineering

Department of Systems and Computer Engineering

Carleton University

Ottawa, Ontario

Canada

April, 2003

© Copyright 2003, Hua Sun

National Library
of Canada

Bibliothèque nationale
du Canada

Acquisitions and
Bibliographic Services

Acquisisitons et
services bibliographiques

395 Wellington Street
Ottawa ON K1A 0N4
Canada

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence

ISBN: 0-612-83482-4

Our file Notre référence

ISBN: 0-612-83482-4

The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de cette thèse sous la forme de microfiche/film, de reproduction sur papier ou sur format électronique.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

L'auteur conserve la propriété du droit d'auteur qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

Canada

The undersigned hereby recommends to the Faculty of Graduate Studies and Research
acceptance of the thesis

Fast Scheduling Strategies for WCDMA-High Speed Downlink Packet Access (HSDPA) System

Submitted by Hua Sun

In partial fulfillment of the requirements for the Degree of
Master of Applied Science in Electrical Engineering

Thesis Supervisor

Prof. R. Hafez

Chair, Department of Systems and Computer Engineering

Prof. Rafik A. Goubran

Carleton University

April, 2003

Abstract

The purpose of scheduling in wireless downlink shared-channel is to allocate appropriate bandwidth and proper coding options to User Equipments at the appropriate times, with the aim of achieving optimal performance. The scheduling algorithm has a considerable influence on the system throughput and service fairness. Developing a simple and efficient scheduling algorithm for the High Speed Downlink Packet Access (HSDPA) channel within Wideband Code Division Multiple Access (WCDMA) is left as an open topic. This thesis studies several frame-to-frame fast scheduling strategies for WCDMA-HSDPA. We study the system performance by simulation, and compare the different schemes in terms of system throughput, average packet call completion time and user satisfaction rate.

The Round Robin (RR) and Maximum C/I (MaxC/I) are two well-known scheduling schemes. The RR algorithm allocates resources evenly among different users while the MaxC/I provides high channel throughput. We propose another scheme called Maximum Self Index (MSI). This proposed scheme provides a better trade-off between fairness and throughput. Furthermore, the MSI fast scheduling algorithm can well adapt to the rapidly changing wireless channel conditions by integrating adaptive modulation and coding techniques.

Acknowledgements

First, I would like to express my sincere and deep thanks to my thesis supervisor Dr. R.Hafez for his supervision and encouragement during the course of this work, as well as his kindness and support during the last two years. His passion for research and knowledgeable suggestions have greatly enhanced my enjoyment of this process. I also thankfully acknowledge the financial support provided by Nortel Networks and the Communications and Information Technology Ontario (CITO) and Carleton University which made it possible for me to finish my Master studies at Carleton University. Finally, My special thanks go to my parents, my dear husband, Chuan Hao, for their constant encouragement, support and love, and my most beloved new born son, Ray, for his enjoyable company during the process of my thesis accomplishment.

Table of contents

ABSTRACT.....	I
ACKNOWLEDGEMENTS.....	II
TABLE OF CONTENTS.....	III
LIST OF FIGURES	VI
LIST OF TABLES.....	X
LIST OF ACRONYMS	XI
CHAPTER 1. INTRODUCTION	1
1.1. Research Motivation	1
1.2. Thesis Summary.....	2
1.3. Thesis Organization	5
CHAPTER 2. BACKGROUND OF WCDMA-HSDPA.....	6
2.1. Background of WCDMA.....	6
2.2. WCDMA Packet Access.....	12
2.3. HSDPA --- High Speed Downlink Packet Access.....	13
CHAPTER 3. SURVEY IN PACKET SCHEDULING ALGORITHMS	18
CHAPTER 4. SYSTEM MODEL AND RELATED DEFINITIONS	22
4.1. System Assumption	22
4.2. Packet Data Traffic Model.....	23
4.3. C/I Distribution Model.....	26
4.4. Throughput Hull Model	26
4.5. Performance Measurement and Related Definitions	29
4.6. Simulation Environment and Software Validation	31

CHAPTER 5. PACKET SCHEDULER---ROUND ROBIN (RR) AND MAXC/I	32
5.1. Scheduling Procedures.....	32
5.1.1. UE Procedures.	33
5.1.2. Base Station Scheduling Procedures.....	33
5.1.2.1 Base Station Scheduling Procedures When Adopting RR Algorithm.....	33
5.1.2.2 Base Station Scheduling Procedures When Adopting MaxC/I Algorithm.	36
5.2. Simulation Result and Performance Analysis in the Whole Sector.....	39
5.3. Performance Analysis as a Function of Distance	43
5.3.1. Grouping Strategy	43
5.3.2. Simulation Results and Analysis	45
5.4. Problems With RR and MaxC/I Schedulers	51
CHAPTER 6. MSI – MAXIMUM SELF INDEX SCHEDULER.....	52
6.1. Initial Efforts for Better Scheduling Schemes	52
6.2. Maximum Self Index (MSI) – Idea Analysis and Algorithm Description..	54
6.2.1. Basic Concept of MSI.....	54
6.2.2. Base Station Scheduling Procedures Using MSI Strategy.....	56
6.3. Simulation Result and Performance Analysis in the Whole Sector.....	58
6.4. Simulation Result and Performance Analysis in Different Groups	62
CHAPTER 7. COMPARISONS AMONG RR, MAXC/I AND MSI BY THEIR GROUPED	
PERFORMANCES	65
7.1. Average UE Throughput: Near_Group, Middle_Group, Far_Group	65
7.2. Average Packet Call Completion Time: Near_Group, Middle_Group,	
Far_Group	68

7.3.	UE Satisfaction Rate: Near_Group, Middle_Group, Far_Group.....	71
CHAPTER 8.	FURTHER STUDY OF MSI	74
8.1.	Strategy for Sampling Study	74
8.2.	Results of Sampling Study and Analysis	75
8.2.1.	Status of Selected UEs	75
8.2.2.	Transmission Packet Length and Number of Decision Times in Each Frame	82
8.3.	Summary of Sampling Results.....	84
CONCLUSIONS AND FUTURE WORKS		87
REFERENCES		92

List of Figures

Figure 2-1.UMTS High-Level System Architecture.	7
Figure 2-2. Network Elements in UTRAN.....	8
Figure 2-3. WCDMA Frame Structure	11
Figure 2-4. Sharing By Means of Time Multiplex As Well As Code Multiplex in HSDPA	14
Figure 2-5. HSDPA Physical Layer Structure[2]	15
Figure 4-1. Typical Characters of A Packet Service Session	24
Figure 4-2. FER vs. CIR for multiple MCS levels	27
Figure 4-3. Throughput Hull, with 20 Codes, SF = 32, MCS Level = 7 , FER = 10^{-3} [11]	29
Figure 5-1. Time Reference for Packet Scheduling vs. Downlink Transmission.....	32
Figure 5-2. RR, Base Station Scheduling Procedures.....	34
Figure 5-3. MaxC/I, Base Station Scheduling Procedures	36
Figure 5_4. Example of a MaxC/I Scheduling Case, Number of Users = 3.....	38
Figure 5-5. RR and MaxC/I, System Throughput vs. System Load	39
Figure 5-6. RR and MaxC/I, Average Packet Call Completion Time vs. System Load ..	40
Figure 5-7. RR and MaxC/I, the UE Satisfaction Rate vs. System Load	40
Figure 5-8. User Grouping Based on Distance	43
Figure 5-9. RR, Average Throughput of Each UE in Different Groups vs. System Load	46
Figure 5-10. MaxC/I, Average Throughput of Each UE in Different Groups vs. System Load	46

Figure 5-11. RR, Average Packet Call Completion Time vs. System Load in Different Groups.....	47
Figure 5-12. Max C/I, Average Packet Call Completion Time vs. System Load in Different Groups	47
Figure 5-13. RR, User Satisfaction Rate vs. System Load in Different Groups.....	48
Figure 5-14. Max C/I, User Satisfaction Rate vs. System Load in Different Groups	48
Figure 6-1. Suggested Scheme Combining RR and MaxC/I.....	52
Figure 6- 2. MSI, Base Station Scheduling Procedures.....	57
Figure 6-3. Throughput Performance in the Whole Sector -- System Throughput vs. System Load	59
Figure 6-4. Delay Performance in the Whole Sector -- Average Packet Call Completion Time vs. System Load	59
Figure 6-5. Fairness Performance in the Whole Sector -- UE Satisfaction Rate vs. System Load	60
Figure 6-6. MSI, Average Throughput of Each UE vs. System Load in Different Groups	62
Figure 6-7. MSI, Average Packet Call Completion Time vs. System Load in Different Groups.....	63
Figure 6-8. MSI, UE Satisfaction Rate vs. System Load in Different Groups	63
Figure 7-1. Performance Comparison in Near Group -- Average Throughput of Each UE vs. System Load.....	65
Figure 7-2. Performance Comparison in Middle Group -- Average Throughput of Each UE vs. System Load	66

Figure 7-3. Performance Comparison in Far Group -- Average Throughput of Each UE vs. System Load.....	66
Figure 7-4. Performance Comparison in Near Group -- Average Packet Call Completion Time of Each UE vs. System Load.....	68
Figure 7-5. Performance Comparison in Middle Group -- Average Packet Call Completion Time of Each UE vs. System Load.....	69
Figure 7-6. Performance Comparison in Far Group -- Average Packet Call Completion Time of Each UE vs. System Load.....	69
Figure 7-7. Performance Comparison in Near Group -- Average UE Satisfaction Rate vs. System Load	71
Figure 7-8. Performance Comparison in Middle Group -- Average UE Satisfaction Rate vs. System Load.....	71
Figure 7-9. Performance Comparison in Far Group -- Average UE Satisfaction Rate vs. System Load	72
Figure 8-1. MSI, Probability Distribution of Locations of UEs While Being Selected, Mean = 1.0056 Km	76
Figure 8-2. MaxC/I, Probability Distribution of Locations of UEs While Being Selected, Mean =0.6689 Km	76
Figure 8-3. MSI, Probability Distribution of C/I Values of UEs While Being Selected, Mean = 17.5642 dB	77
Figure 8-4. MaxC/I, Probability Distribution of C/I Values of UEs While Being Selected, Mean =21.7636 dB	77

Figure 8-5. MSI, Probability Distribution of Buffer Queue Length of UEs While Being Selected, Mean = 18,596 bytes	78
Figure 8-6. MaxC/I, Probability Distribution of Buffer Queue Length of UEs While Being Selected, Mean =7,079.4 bytes	78
Figure 8-7. MSI, Probability Distribution of Valid Time Duration of UEs While Being Selected, Mean =31.8765 second	79
Figure 8-8. MaxC/I, Probability Distribution of Valid Time Duration of UEs While Being Selected, Mean =10.6784 second	79
Figure 8-9. MSI, Probability Distribution of Transmitted Packet Length in Each Frame, Mean = 12,072 byte	83
Figure 8-10. MaxC/I, Probability Distribution of Transmitted Packet Length in Each Frame, Mean = 10,894 byte.....	83

List of Tables

Table 2-1. Main Physical Parameters in WCDMA	9
Table 2-2. Features of Different Channels for Packet Transmission.....	13
Table 2-3. Maximum Information Rate Supportable By Each MCS Level, Chip Rate = 3.84 Mcps, SF = 32.....	17
Table 8-1. Summary of Sampling Results of MSI and MaxC/I	84

List of Acronyms

3G	The Third Generation
3GPP	the 3rd Generation Partnership Project
AMC	Adaptive Modulation and Coding
AMCS	Adaptive Modulation and Coding Schemes
AMPS	Advanced Mobile Phone Service
BPSK	Binary Phase Shift Keying
BoD	Bandwidth on Demand
C/I	Carrier over Interference
CDMA	Code Division Multiple Access
CN	Core Network
DL	DownLink
DS	Direct-Sequence
ETSI	European Telecommunications Standards Institute
FACH	Forward Access Channel
FCS	Fast Cell Selection
FDD	Frequency Division Duplex
FIFO	First In First Out

FTP	File Transfer Protocol
GPRS	General Packet Radio Service
H-ARQ	Hybrid-Automatic Repeat reQuest
HDR	High Data Rate
HS-DSCH	High Speed-Downlink Shared CHannel
HSDPA	High Speed Downlink Packet Access
MC	Multi_Carrier
MCS	Modulation and Coding Schemes
MINO	Multiple-Input Multiple-Output
MSI	Maximum Self Index
MaxC/I	Maximum Carrier over Interference
NRT	Non-Real-Time
OVSF	Orthogonal Variable Spreading Factor
PDF	probability density function
PDU	Protocol Data Units
PS	Packet Scheduler
QAM	Quadrature Amplitude Modulation
QPSK	Quadrature Phase Shift Keying
QoS	Quality of Service

RACH	Random Access CHannel
RNC	Radio Network Controller
RR	Round Robin
SF	Spreading Factor
SI	Self Index
SNR	Signal to Noise Ratio
TDD	Time Division Duplex
TDMA	Time Division Multiple Access
UE	User Equipment
UL	UpLink
UMTS	Universal Mobile Telecommunications System
UTRA	Universal Terrestrial Radio Access
UTRAN	UMTS Terrestrial Radio Access Network
VFEC	Variable Forward Error Correction
WCDMA	Wideband Code Division Multiple Access
WWW	World Wide Web

Chapter 1. Introduction

1.1. Research Motivation

The technology for the third generation (3G) CDMA wireless systems has been advanced at an ever-increasing speed in recent years. The focus of the technology development has shifted from improving systems that provides traditional voice and other circuit-switched services to enabling high-speed, high-efficiency wireless packet data systems. As the concept of wireless Internet gradually turns into reality, and with an increased demand for wireless data services, there is a need for an efficient high-speed data system. Specifically, the research effort has been mainly focused on improving the downlink capacity. The High Speed Downlink Packet Access (HSDPA) has been recently proposed by 3GPP [2] for future deployment in UMTS Release 5 (R5). Several enhancement techniques, such as Adaptive Modulation and Coding (AMC), Hybrid ARQ (H-ARQ), Fast Cell Selection (FCS), Multiple Input Multiple Output (MIMO) Antenna Processing, and Frame-by-Frame Scheduling, are proposed in order to provide very high speed downlink packet access by means of a high-speed downlink shared channel (HS-DSCH)[2]. Among these technologies, frame-by-frame scheduling is left as one of the open topics for further investigation. It is known that selecting a simple and efficient scheduling algorithm is very important in the sense that it has considerable impact on the system performance, such as channel throughput and service fairness. Based on this background, our research is focused on the studies of fast scheduling strategies for HSDPA system.

1.2. Thesis Summary

Round Robin (RR) and Maximum C/I (MaxC/I) are two well-known scheduling algorithms in packet data services. RR runs the scheduling in a fixed cyclic order. All active users are identified by their ID numbers, and each user is allocated an equal, fixed number of time slot(s) in a ring fashion. Transmission service will not be re-granted to the same user before every other user has been served. MaxC/I schedules packets in a way under which all active users are periodically ranked by their reported C/I values. At every scheduling event, the user with the highest C/I value is first selected, allocated a number of time slots (maximum 15 time slots, i.e., the whole frame), depending on the service requirement and the availability of slot resources. The same user can be re-scheduled in the next adjacent frame provided that it still possesses the highest C/I value at the next scheduling event. It's been widely accepted that RR has the advantage of being "fair" in the sense of giving equal slots to every individual user. It is also known that MaxC/I provides high channel throughput.

Under wireless channel conditions, in order to obtain the numerical performance of RR and MaxC/I in WCDMA-HSDPA system, simulation is carried out based on the system parameters of HSDPA, including the C/I distribution, and their corresponding throughput hulls. The performance is measured in terms of system throughput, packet delay and user satisfaction rate as functions of the system load. The impact of users' locations is also studied by grouping users according to their distance to the cell center.

Our simulation result shows that in RR, assigning equal chance and equal number of time slots to each user couldn't ensure equal amount of data transmission due to the impact of each user's location, and accordingly, the channel condition. RR is not able to provide

fairness as was expected before. This is a major difference between scheduling in the wireless and wired environments. RR is also very poor in system throughput performance, with only 30% utilization of the downlink shared channel capacity.

With MaxC/I, a considerably high channel throughput is achieved. However, this high throughput is obtained at the cost of fair allocation when the system load is heavy. In 3G wireless environments, it's very likely that users near the base station see statistically higher C/I values than those far from the base station, which can be translated that near users have higher possibilities to obtain transmission services. The inequality between near and far users in getting services is the main disadvantage of the MaxC/I.

In order to improve the fairness in service to all users without sacrificing much in the obtainable throughput, one should incorporate aspects of both scheduling methods. This is one of the main objectives and contribution of this thesis. The idea of splitting one frame into two parts, each of which uses either RR or MaxC/I was firstly simulated. The results obtained illustrate that the improvement in fairness wasn't worth the cost of the throughput that was sacrificed. We also designed a new scheduling scheme that provides a trade-off between throughput and fairness by introducing an indexed criteria. The scheme considers both of the user's C/I value and its buffered queue length at each scheduling event. However, the simulation results were not satisfactory either.

By incorporating an idea [9] proposed for cdma2000-HDR system and transplanting it to WCDMA-HSDPA, a fresh scheduling scheme is introduced and studied. In this scheme (we call it MSI), which is similar to the second suggested algorithm, users are selected for service based on two factors: the data rate that can be supported at the instantaneous scheduling event, and the data amount that has been transmitted for that user in the past.

It is the definition of the index that makes MSI different. Rather than applying a composite sum operation (as was tried before), the MSI applies the ratio of these two factors. By analyzing the performance of MSI and comparing it with RR and MaxC/I, two noticeable results were obtained:

- 1). MSI is superior to RR in terms of fairness, in that each user can get highly identical individual throughput in data transmission service, regardless of its location.
- 2). MSI is superior to MaxC/I in terms of system throughput. With MSI, the system can achieve an even higher throughput than with MaxC/I. This phenomenon seems to be an interesting paradox, since MaxC/I selects the user with the highest C/I value, thus the highest data rate, at each scheduling event, and is usually considered to provide the upper bound in system throughput over all other schedulers [2].

To explain the mechanisms behind these superior properties of MSI, further research is carried out by examining some detailed information at each scheduling event. The parameters of interest are: user locations and C/I values while being selected, their buffer queue length and valid time duration, the size of packets in each transmission frame, and the number of decision times in each frame. The explanation of the observed result is provided at the end of the thesis.

We concluded that MSI ensures not only a considerably high system throughput, but also a complete fair and satisfactory service to all users. In this sense, MSI is the best suitable option for packet scheduling in HSDPA system [4].

1.3. Thesis Organization

The thesis is organized as follows. Chapter 1 introduces the motivation and outline of the thesis. Chapter 2 explains the background and main features of the WCDMA and HSDPA system, on which our scheduling study is based. A brief survey of the-state-of-the-art scheduling algorithms is in chapter 3. Chapter 4 outlines the simulation layout, model and parameters of our simulation. By simulation, Round Robin and Max C/I schedulers are studied and the simulation results are presented in chapter 5. Chapter 6 investigates an improved algorithm--Max Self Index (MSI), which gives better performance in both system throughput and user satisfaction rate. Chapter 7 compares among RR, Max C/I and MSI by their grouped performances. The mechanisms behind MSI and the reason for its better performance are further studied and explained in chapter 8. The thesis conclusion and future work is presented at the end.

Chapter 2. Background of WCDMA-HSDPA

2.1. Background of WCDMA

Analog cellular systems, (e.g., AMPS), are commonly referred to as the first generation systems. The digital systems that are currently in use, such as GSM, CDMA (IS-95) and TDMA (IS-136), are referred to as second generation (2G) systems. These systems have enabled wireless telephony and other services such as Short Messaging Service (SMS) and dial-up access to data networks. These applications achieved high commercial success. Third generation (3G) systems, such as Cdma2000 and the UMTS (Universal Mobile Telecommunications System), are designed to support multimedia communications. They are expected to enhance person-to-person communication with high quality images and video, and access to information and services on public and private networks with higher data rates and new flexible communication capability [18].

In order to save resources and minimize cost, the existing cellular systems (GSM/GPRS) will be incorporated into the overall UMTS structure. Within the UMTS, transmission of voice will be supported using circuit switched techniques. Transmission of data however, will be accomplished using separate packet switched core networks, which are already a part of the existing GPRS systems, as shown in Figure 2-1. UMTS, therefore, uses the existing core network, which is currently part of GPRS, but an entirely new radio interface. The new network in UMTS is called UTRAN (UMTS Terrestrial Radio Access Network). UTRAN is organized in a similar fashion as the older GSM/GPRS cellular systems. Within the UTRAN all of the UE (User Equipment) in one cell communicate with the base station in that cell. A number of base stations are controlled by, and

communicate with one Node B. A number of Node Bs are in turn controlled by, and communicate with one RNC (Radio Network Controller), as shown in Figure 2-2. A system consisting of one RNC which controls multiple Node Bs, which in turn control multiple base stations is called the RNS (Radio Network Subsystem).

Figure 2-1.UMTS High-Level System Architecture.

Figure 2-2. Network Elements in UTRAN

Within the UTRAN, WCDMA (Wideband Code Division Multiple Access) is the most widely adopted air interface standard, and its specification has been created by 3GPP (the 3rd Generation Partnership Project), which is the joint standardization project, with contributions from standardization bodies in Europe, Japan, Korea, the USA and China.

The main physical parameters of WCDMA are listed in Table 2-1.

Table 2-1. Main Physical Parameters in WCDMA

Channel bandwidth	5 MHz
Duplex mode	FDD and TDD
Downlink RF channel structure	Direct spread
Chip rate	3.84 Mbps
Frame length	10 ms
Spreading modulation	Balanced QPSK (downlink), Dual-channel QPSK(uplink) Complex spreading circuit
Data modulation	QPSK (downlink), BPSK (uplink)
Channel coding	Convolutional and turbo codes
Coherent detection	User dedicated time multiplexed pilot (downlink and uplink), common pilot in the downlink
Channel multiplexing in downlink	Data and control channels time multiplexed
Channel multiplexing in uplink	Control and pilot channel time multiplexed I&Q multiplexing for data and control channel
Multirate	Variable spreading and multicode
Spreading factors	4–256 (uplink), 4–512 (uplink)
Power control	Open and fast closed loop (1.5 kHz)
Spreading (downlink)	OVSF sequences for channel separation Gold sequences $2^{18}-1$ for cell and user separation (truncated cycle 10 ms)
Spreading (uplink)	OVSF sequences, Gold sequence 2^{41} for user separation (different time shifts in I and Q channel, truncated cycle 10 ms)
Handover	Soft handover Inter-frequency handover

The main features of WCDMA are as follows[1]:

- WCDMA is a wideband Direct-Sequence Code Division Multiple Access (DS-SS-CDMA) system. User information bits are spread over a wide bandwidth by multiplying the user data with quasi-random bits (chips) derived from CDMA spreading codes. In order to support very high bit rates (up to 2Mbps), the use of a variable spreading factor and multi-code connection is supported.
- The system fixed chip rate of 3.84 Mcps leads to a carrier bandwidth of approximately 5MHz. The inherently wide carrier bandwidth of WCDMA supports high user data rates and also has certain performance benefits, such as increased multipath diversity.
- WCDMA supports variable user data rate and the concept of Bandwidth on Demand (BoD). Transmission takes place in frames of 10ms duration, during which the user data rate is kept constant. However, the data rate can change from frame to frame. This fast radio capacity allocation will typically be controlled by the network to achieve optimum throughput for packet data services.
- WCDMA supports two basic modes of operation: FDD and TDD. In the FDD mode, two separate 5 MHz bands are used for the uplink and downlink respectively, whereas in TDD only one 5 MHz is time-shared between up and down link.
- WCDMA supports the operation of asynchronous base stations, so unlike in the synchronous IS-95 system, there's no need for a global time reference, such as GPS. Deployment of indoor and micro base stations is easier when no GPS signal

(required line-of-sight) is needed, but the inter-frequency handover is different from IS-95 handover techniques.

- The WCDMA air interface has been crafted in such a way that advanced CDMA receiver concepts, such as multi-user detection and smart adaptive antennas, can be deployed by the network operator as a system option to increase capacity and/or coverage.
- WCDMA is designed to be deployed in conjunction with GSM. Therefore, handovers between GSM and WCDMA are supported in order to be able to leverage the GSM coverage for the introduction of WCDMA.
- In WCDMA, upon connection set up, terminal users shall tell the network a set of parameters indicating their radio access capabilities. Upon these parameters, the maximum user data rate is determined in particular radio configuration.

Figure 2-3. WCDMA Frame Structure

A general WCDMA MAC layer frame is shown in Figure 2-3. The frame is 10ms long, and consists of 15 time slots, each of which is 0.667ms. Each frame is part of a super frame, which is 720ms long, and consists of 72 frames.

2.2. WCDMA Packet Access

Packet Access is an important feature and focus of 3G cellular communications [1]. It provides data services up to 2Mbps. Packet allocations in WCDMA are controlled by the packet scheduler (PS), which is typically located in the Radio Network Controller (RNC) where the scheduling can be done efficiently for multiple cells. The functions of packet schedulers are:

- Properly allocate the available resources (time, code or power) among the packet data users
- Decide the allocated bit rates and the length of the allocation
- Determine which transport format to use on the current data
- Monitor packet allocations and the systems loads

WCDMA packet access allows NRT (non-real-time) UE to use common, dedicated or shared channels dynamically. The usage of different channels is controlled by the packet scheduler based on:

- Service type or UE parameters, e.g., delay requirement;
- Data amount;
- Load of the common channels and shared channels;
- Interference levels in the air interface;

- Radio performance of different transport channels.

The features of different channels are summarized in Table 2-2.

Table 2-2. Features of Different Channels for Packet Transmission.

	Channel characters	Suitable data type
Common channel (RACH/UL, FACH/DL)	Low set up time, but poor radio performance because there is no feedback channel, no fast closed loop power control.	Signaling data and small individual packets
Dedicated channel	Longer set up time, but improved link-level quality because of fast power control via feedback channels.	Suitable for medium and large data amount transmission, but not for bursty data.
Shared channel	Share one or more physical channels (i.e., orthogonal codes), between many users in a time division manner, code efficient. Fast power control, fast allocation and rate modification	Targeted to transfer bursty packet data, also suitable for medium or large amount data transmission.

2.3. HSDPA --- High Speed Downlink Packet Access

High Speed Downlink Packet Access (HSDPA) is a packet-based data service in WCDMA downlink with data transmission up to 10 Mbps (and 20 Mbps for MIMO systems) over a 5MHz bandwidth [20]. HSDPA incorporates several advanced features, such as: Adaptive Modulation and Coding (AMC), Multiple-Input Multiple-Output (MIMO), Hybrid Automatic Request (HARQ), fast cell search, advanced receiver design, and fast scheduling scheme. HSDPA reuses the WCDMA framework and functionality,

and coexists with current services on the same carrier. This thesis focuses only on the study of frame-by-frame scheduling, while integrating AMC.

On the physical layer, HSDPA transmission is carried out on HS-DSCH, which is a set of downlink physical channels (codes) shared by users at least in time domain and possibly also in the code domain, as illustrated in Figure 2-4. In the time domain, the chip rate of downlink packet channel is fixed as 3.84 Mcps. Each Radio Frame is 10ms, consisting of 15 slots. The HS-DSCH operates in TDM fashion. Furthermore, there is no power control in HSDPA system. In our study, the base station transmits at a fixed power level (17watt) to multiple downlink physical channels. All resources, power and channels (code sequences), are dedicated to one scheduled user during a very small time period. The user adapts to instant link condition by exploiting AMCS (Adaptive Modulation and Coding Scheme).

Figure 2-4. Sharing By Means of Time Multiplex As Well As Code Multiplex in HSDPA

The purpose of time domain fat-pipe scheduling is to provide very high data rate under instantaneous channel condition to one user only, this is in contrast to splitting resources among multiple users, where each individual user gets a low average data rate. This is different from the averaging effort that is well known in CDMA, and is adopted to well suit the properties of bursty packet data.

Figure 2-5 shows the HSDPA physical layer structure.

Figure 2-5. HSDPA Physical Layer Structure[2]

Since the scheduling algorithms in our research incorporate AMCS, we will describe the used AMCS schemes in more details.

The term AMC means that the modulation and coding can be changed from frame to frame based on instantaneous channel measurement. It is an alternative way of link adaptation functioning similarly as power control. The concept is that with good channel condition, higher order modulation, less coding or less spreading may be used. With bad channel condition, lower order modulation, more coding or more spreading should be adopted for better error protection. The main benefits of using AMC are:

- Higher data rates are available for users in favorable positions, which in turn increases the average throughput of the cell;
- Reduced interference variation due to link adaptation based on variations in the modulation/coding scheme instead of variations in transmit power;
- Effective when combined with fat-pipe scheduling techniques such as those enabled by the Downlink Shared Channel.

While AMC has the above mentioned advantages, its realization is complicated by requiring instantaneous channel measurement and feedback, detection of signal with higher order modulation, and fast switching between different transmission mode.

In 3GPP's study item, AMC scheme uses 7 Modulation and Coding Schemes (MCS), based on the combination of QPSK, 8-PSK, 16-QAM and 64-QAM as modulation, and $R = \frac{1}{2}$ or $\frac{3}{4}$ as Turbo Coding [2]. The maximum information rate that can be supported by each MCS level is listed in Table 2-3. These parameters are used to calculate the throughput hull, which is employed in this thesis to calculate the system throughput under different scheduling schemes.

Table 2-3. Maximum Information Rate Supportable By Each MCS Level, Chip Rate = 3.84 Mcps, SF = 32.

MCS	Chip Rate = 3.84 Mcps, SF = 32			
	Info rate per code Mbps	Max. info rate (20 codes) Mbps	Code rate	Modulation
7	0.54	10.8	$\frac{3}{4}$	64QAM
6	0.36	7.2	$\frac{3}{4}$	16QAM
5	0.24	4.8	$\frac{1}{2}$	16QAM
4	0.27	5.4	$\frac{3}{4}$	8PSK
3	0.18	3.6	$\frac{3}{4}$	QSK
2	0.12	2.4	$\frac{1}{2}$	QPSK
1	0.06	1.2	$\frac{1}{4}$	QPSK

Chapter 3. Survey in Packet Scheduling Algorithms

The packet scheduling function shares the available air interface capacity among packet users and aims at achieving the most efficient utilization of radio channels. The packet scheduling algorithm usually decides the following factors at each scheduling event:

- At what moment a user is to be served;
- Based on what parameters or indexes are users to be selected to be served;
- At what data rate will the data be transmitted;
- How many slots/codes should be allocated to a given data unit.

Recently, considerable attention was given to the research of efficient scheduling schemes in 3G WCDMA system.

Ref. [5] proposes a protocol based on the proposed 3G WCDMA standards that provides per-connection variable bandwidth guarantee to bursty data applications in the forward link of an OVSF-CDMA system. The heart of the protocol is a scheduler, which uses a credit-based reservation scheme to prioritize users and dynamically reassigns codes on a time slot basis to achieve statistical multiplexing of bursty traffic. The “credit” represents the difference between guaranteed transmissions and actual transmissions. A positive implies the connection is receiving less service than guaranteed. The main idea of the scheduling algorithms is to try to keep every connection’s credit upper-bounded by a small positive value, so that any connection is only lagging behind its guaranteed rate by a small amount.

Ref. [6] studies the MAC scheduling for uplink transmission in UMTS WCDMA system. It proposes a load-priority based algorithm and an algorithm with a fairness mechanism. The idea is to assign channels based on load-counter for pursuing a maximum throughput, or to reserve channels for fairness.

Ref. [7] compares the performance of four scheduling schemes in WCDMA downlink, such as round-robin and fastest-first schemes that are appropriate for uni-access mode, and an equal-rate and equal-weight schemes for multi-access mode transmission. Their result shows that the system employing the uni-access mode scheme performs better than the system with multi-access mode in terms of average delay, but performs worse in terms of fair allocation of data rates.

Ref. [8] examines the admission control and scheduling topics in WCDMA downlink, with the consideration of both real-time and non-real-time traffic (NRT) types. The scheduler always gives first priority to the real-time class. NRT can be scheduled based on largest queue priority.

Ref. [13] proposes an adaptive algorithm in UTRA/TDD mode that can deal with different satisfied user criteria by simply adjusting a single parameter. In this strategy, the available resources are divided among all users that have data to transmit. The data of every user are added up to a unique queue in the order they're generated. The retransmission data are also appended to the queue, which works somewhat like a FIFO (First-In-First-Out) buffer. Certain amount of Protocol Data Units (PDU) per frame is allocated to every user as long as resources are available. If the number of PDUs per user is small, the scheduler behaves like a RR strategy, and can be expected to be suited for a "fairness based" schedule. If the number of PDUs per user is chosen in a way that the

requested data amount for a user is larger than the available resource units, the scheduler behaves like a FIFO buffer, and is expected to be suited for “delay based” criteria. By simply changing the parameter, number of PDUs per user, a smooth transition can be achieved from RR to FIFO behavior of the scheduler. Therefore a simple adjustment for different user satisfaction criteria is possible. This strategy is called “merged RR-FIFO” strategy.

Ref. [14] presents an approach for unifying power control, variable forward error correction (VFEC), and scheduling for a downlink in a multi-user, multimedia indoor wireless communication system by allocating the system resource, namely, fixed bandwidth available for each user and the transmit power budget for each cell. Power control (downlink) is used to control the received SNR at the mobiles. FEC introduces redundancy to combat transmission errors. Because the bandwidth demands for multimedia applications vary greatly, scheduling (i.e., prioritizing) among data streams is needed when bandwidth demand exceed the channel capacity. The objective is to maximize the overall system utility. This objective is suggested to be achieved by applying a distributed algorithm which divides the overall optimization problem into a hierarchy of three levels – system, cell, and user ---, with each performing independent and parallel optimizations. This partitioning helps to localize the problems. The locally optimized results are then presented to the next leveling the hierarchy. This layering approach yields a distributed algorithm, which is essential given that any practical wireless system has to be scalable.

Ref. [12] studies the scheduling and other techniques in cdma2000-HDR system, a CDMA High Data Rate system proposed to address wireless access to data network.

Similar to HSDPA, the HDR system design improves the system throughput by utilizing fast channel estimation feedback, dual receiver antenna diversity, Hybrid-ARQ and a scheduling algorithm that takes the advantage of multi-user diversity. In this paper, the downlink packet may take from 1 to 16 time slots. The transmission of a multi-slot packet is interlaced such that there are three time slots between two adjacent time slots. This slot organization facilitates the functionality of H-ARQ, which allows early termination of multi-slot packet transmission.

Ref. [9] is another paper that studied the downlink data throughput of CDMA-HDR system. Data is transmitted to different UE in a TDM fashion. The rate of transmission to each UE is variable and depends on each UE's measured C/I, which is reported to the base station. In the scheduling algorithm, the scheduler at Node B determines the next UE to be served based on the supportable data rate and the amount of data that has been transmitted to each UE. Only the system throughput is provided with omni-antenna and beam-antenna in this paper. But the scheduling idea will be employed into our paper, and the performance will be compared with other algorithms, not only in terms of throughput, but also delay, user satisfaction rate and fairness analysis in the following chapters.

Many other useful resources related to our study could be found in [16][17][18][19].

Chapter 4. System Model and Related Definitions

This chapter describes the system model, configuration and parameters for simulating different scheduling algorithms in HSDPA system, and some related definitions. The packet data traffic model follows ETSI's WWW browsing entity model [3], the C/I distribution of UEs and throughput hull model are derived in [11], and the results are used in our simulation.

4.1. System Assumption

In our study, we use parameters suggested by the 3GPP for the HSDPA channel. The data is spread over a 5MHz bandwidth, with a Spreading Factor $SF = 32$ and a fixed chip rate $= 3.84\text{Mcps}$. The information rate can be variable by applying different Modulation and Coding schemes for transmission. The UE terminal monitors the forward channel, measures the channel quality (C/I), and reports it to the base station on a frame-by-frame basis. The scheduler at the Radio Node Center (RNC) uses the C/I value to estimate its supportable rate by looking up the throughput hull curve, and determines the data rate by allocating different Modulation and Coding Schemes (MCS) and number of code sequences. A typical HSDPA frame length is 10 ms, which consists of 15 time slots, each of which is 0.667ms. A 19-cell structure, with a cell radius of 1.616 Km, is adopted for calculating the C/I samples seen by a UE at any point within the cell of interest. Path loss and lognormal shadowing is considered in radio transmission. Only out of cell interference is accounted for since the inter-cell interference is small between orthogonal codes in the downlink direction. In HSDPA, there's no power control. The base station transmits at a fixed power level. Link adaptation is implemented through variable

transmission data rate by applying AMCS. On a frame-by-frame basis, every UE obtains a random C/I ratio from an appropriate C/I distribution based on its' distance from the BS. Based on the obtained C/I value, the UE supportable data rate is calculated according to the throughput hull, which is the maximum achievable information rate for a given C/I value.

To study the effect of scheduling on performance, we only needed to perform a system level simulation. Our simulation is focused on one sector, wherein UEs are uniformly distributed and remain static. The system load increases by increasing the number of users, and all users, once fixed, are assumed to be in active sessions throughout the simulation.

4.2. Packet Data Traffic Model

Our simulation adopts the ETSI's typical Internet browsing entity as non-real time data traffic model. This model is widely adopted in wireless communication research. Figure 4-1 depicts the ETSI model for a WWW browsing session, which consists of a sequence of packet calls [3]. The user initiates a packet call when requesting an information entity. During a packet call several packets may be generated, which means that the packet call consists of a bursty sequence of packets. The burstyness during the packet call is an important feature that affects the packet transmission rate.

Figure 4-1. Typical Characters of A Packet Service Session

A packet service session contains one or several packet calls depending on the application. For example, in a WWW browsing session, a packet call corresponds to the downloading of a WWW document. After the document is entirely arrived to the terminal, the user is consuming certain amount of time for studying the information. This time interval is called reading time [10]. A session that contains only one packet call is representative to the case for a file transfer (FTP). A widely accepted distribution patterns and parameters of the WWW entity is as follows:

- Number of packet calls per session: geometrical distribution, mean = 5;
- Number of packets per packet call: geometrical distribution, mean = 25;

- Size of each packet: Pareto distribution with cut-off, $\alpha = 1.1$, $k = 81$, $m = 66666$ bytes, mean = 480 bytes;
- Reading time: exponential distribution, mean = 412 second; (we adjusted the reading time to shorten the simulation time.);
- Inter-arrival time between packets: exponential distribution, mean = 0.0277.

The probability density function (PDF) of each distribution is used to generate the random numbers according to reverse function techniques. Specifically, the packet size, noted as *PacketSize*, adopts a Pareto distribution ($\alpha = 1.1$, $k = 81.5$) with cut-off ($m = 66666$ bytes).

The normal Pareto distribution (without cut-off) is defined by:

$$f_x(x) = \frac{\alpha \cdot k^\alpha}{x^{\alpha+1}}, x \geq k \quad (\text{Eq 4.1})$$

$$F_x(x) = 1 - \left(\frac{k}{x}\right)^\alpha, x \geq k \quad (\text{Eq 4.2})$$

$$\mu = \frac{k\alpha}{\alpha-1}, \alpha > 1 \quad (\text{Eq 4.3})$$

$$\sigma^2 = \frac{k^2 \cdot \alpha}{(\alpha-2) \cdot (\alpha-1)^2}, \alpha > 2 \quad (\text{Eq 4.4})$$

PacketSize is defined with the following formula:

$$PacketSize = \min(p, m), \quad (\text{Eq 4.5})$$

where p is normal Pareto distributed random variable ($\alpha=1.1$, $k=81.5$ bytes) and m is maximum allowed packet size, $m=66666$ bytes.

The PDF of the *PacketSize* then becomes:

$$f_n(x) = \begin{cases} \frac{\alpha \cdot k^\alpha}{x^{\alpha+1}}, & k \leq x < m \\ \beta, & x \geq m \end{cases} \quad (\text{Eq 4.6})$$

where β is the probability that $x \geq m$, and

$$\beta = \int_m^\infty f_n(x) dx = \left(\frac{k}{m}\right)^\alpha, \alpha > 1 \quad (\text{Eq 4.7})$$

Then the mean packet size is calculated as:

$$\mu_n = \int_{-\infty}^{+\infty} x \cdot f_n(x) dx = \int_k^m x \cdot \frac{\alpha \cdot k^\alpha}{x^{\alpha+1}} dx + m \left(\frac{k}{m}\right)^\alpha = \dots = \frac{\alpha k - m \left(\frac{k}{m}\right)^\alpha}{\alpha - 1} \quad (\text{Eq 4.8})$$

With the parameters set above, *Mean PacketSize* = 480 bytes.

4.3. C/I Distribution Model

The C/I samples are collected every 100 meters along a radial line starting from the cell center to the cell edge. Refer to [11] for more information. A UE updates its C/I ratio by randomly picking up an appropriate sample according to its distance from the BS.

4.4. Throughput Hull Model

To translate a given C/I value seen by a UE to its supportable data rate, the throughput hull, which is the maximum achievable information rate for a given C/I value, is calculated. The method used in determining the maximum data rate is explained in [11] and it is not part of this thesis. Briefly, the scheduler calculates the channel throughput

for all possible MCS and “n” codes. There are seven MCS levels, and the number of code n ranges from 1 to 20. Obviously, the selected combination must be supported by the maximum available C/I.

Figure 4-2. FER vs. CIR for multiple MCS levels

Generally, the following steps apply in the derivation of the throughput hull.

1. A target FER is set to be 10^{-3} as the channel requirement;
2. Given a target FER, different MCS combination requires different target C/I value. Higher MCS level needs higher target C/I value than lower MCS level, as shown in Figure 4-2.
3. For a given C/I value measured by a UE, its received power is optimally allocated to a plurality of channels (maximum 20) in a way that different MCS may be applied to each individual channel, so that from a particular set of channel and MCS combination out of exclusive possible combinations, the maximum throughput is obtained;
4. By increasing the measured C/I with a preset increment step, repeat the optimal power allocation procedure in point 3, so as to derive the corresponding maximum throughput against the given C/I value. In this way, the throughput hull curve is obtained.
5. One point should be clarified is that during the optimal power allocation, the power allocated to each individual channel should be at least the target C/I of its corresponding MCS level, in order to meet the FER requirement.

Figure 4-3 illustrates the throughput hull as a function of the C/I value, using maximum 20 codes/channels and 7 possible MCS as listed in Table 2-3. The FER is set to be 10^{-3} .

Figure 4-3. Throughput Hull, with 20 Codes, SF = 32, MCS Level = 7 , FER = 10^{-3} [11]

4.5. Performance Measurement and Related Definitions

The performance of various scheduling algorithms is measured in terms of system and UE throughput, delay, and UE satisfaction rate. Definitions of these measurement and related terms are as follows.

Throughput

In this thesis, throughput is defined as the number of information bits transmitted in each second (bps). It's equal to the number of information bits that are received and accepted by the UE receiver within a time period T divided by that time period.

Delay

Delay is defined by the mean value of packet call completion time, measured from the instant when the first packet of a packet call arrives at the BS's buffer queue until the last packet of the packet call is received by the UE. The head and tail of a packet call is

therefore marked. The time of the first packet's arrival and the last packet's transmission is also recorded for the purpose of delay calculation.

UE satisfaction rate

UE satisfaction rate is the ratio of the number of satisfied users to the total number of users. A UE is defined as being "satisfied"[3] when the following criteria are fulfilled:

1. The user does not get blocked when arriving to the system;
2. Its throughput is greater than 10 % of the average bit rate of the source;
3. The user's packets do not get dropped.

Fairness

Fairness is a measure of the similarities of the QoS provided to each individual UE, where the QoS is defined in this thesis as the throughput an individual UE experiences, as opposed to providing equal number of time slots among UEs. Fair service indicates that a UE can obtain a consistent download speed regardless of its physical location. Being fair in service does not guarantee a UE to be satisfied, and vice versa. Having both fair and satisfactory service is most desirable to operators and customers.

Schedule event

A scheduling event is the process of selecting appropriate UEs and allocating appropriate time slots to them in the immediate frame that follows the scheduling. A scheduling event, executed on a frame-by-frame basis, follows the procedure stipulated by the scheduling algorithm.

Decision point

In a scheduling event, a decision point is the point of scheduling the next UE, if the previously selected UE needs only part of a frame, i.e., the remaining part can be further filled by other UEs. There is at least one, but no more than fifteen decision points in every scheduling event.

System load

System load is the amount of traffic arrived to the transmission queues measured in bps. In our thesis, where every user follows the same traffic pattern, system load varies by changing the number of users.

Maximum system capacity

The maximum system capacity is defined as the maximum obtainable throughput which is stipulated in HSDPA system as 10.8Mbps (highest modulation level and highest code rate over 20 parallel codes).

Active session

A user is defined to be in active session if it generates download requests according to the traffic model described earlier.

Valid time duration

Valid time duration is the summation of time slices when a user's buffer is not empty.

4.6. Simulation Environment and Software Validation

We use C++ as the simulation tool. Software is validated through validation examples, as well as by running some deterministic cases.

Chapter 5. Packet Scheduler---Round Robin (RR) and MaxC/I

5.1. Scheduling Procedures

Within the scheduling unit at the RNC, each UE that requests download packets is allocated a buffer associated with it. The requested download packets are temporarily stored in those buffers before being scheduled and transmitted to the end UE through the downlink shared-channels. At the end of the current frame, a UE reports its instantaneous C/I level, which is recorded and assumed fixed over one frame. By the end of the current frame, the schedule for the next frame is ready, as shown in Figure 5-1. (The signalling between base station and end UEs is beyond discussion in this thesis.) The performance of different scheduling algorithms are examined under various level of offered traffic load.

Figure 5-1. Time Reference for Packet Scheduling vs. Downlink Transmission

5.1.1. UE Procedures.

From UE's perspective, the requested downloading packets of each UE arrive at base station and are appended to the end of the queue in its associated buffer. If the buffered queue size exceeds the maximum buffer size, the following incoming packets will be discarded, which makes the UE of interest "not satisfied". (Refer section 4.5 for the definition of "user satisfaction".) Once the concerned UE is selected for service, data will be drawn from its buffer and transmitted to the UE's end.

5.1.2. Base Station Scheduling Procedures.

5.1.2.1 Base Station Scheduling Procedures When Adopting RR Algorithm

In RR, each UE is allocated an ID, which is used to poll the UE sequentially. Starting from the first time slot of a frame, the scheduler checks if a UE has data in its buffer. If no, the scheduler skips the current UE and checks the next one. Otherwise, it assigns one time slot to the current UE. This procedure is executed cyclically until the last time slot in the frame has been assigned, or until all UEs' buffers are empty before all the fifteen time slots are used up. The last scheduled UE's ID is recorded at the end of the current scheduling event. The next scheduling cycle (next frame) starts with the user whose ID is the next one to the currently recorded ID. And the round robin continues. All non-empty queues must be serviced before re-servicing a user.

Figure 5-2. RR, Base Station Scheduling Procedures

Specifically, the scheduling results can be categorised into three cases.

Case 1. The number of users (whose buffers are not empty) is greater than the number of time slots in one frame (i.e., 15), and all the fifteen time slots has been filled at the end of the current scheduling event.

In this case, the top fifteen waiting will be scheduled with one time slot each. The next scheduling procedure starts with the sixteenth user. Till a particular scheduling frame wherein the last user has been served, or equivalently, every users has been served at least once, the service continues by turning around to the first user for the second round.

Case 2. The number of active users is less than the number of time slots in one frame (i.e., 15), and all the fifteen time slots has been filled at the end of the current scheduling event.

In this case, since the number of users is less than the number of time slots in one frame, some users can be assigned multiple time slots, as opposed to case 1, where only one time slot is permitted for each user. In other words, all time slots in one frame are divided up among small number of users, for the purpose of maximizing the utilization of resources. Practical protocols for transmitting individual user's data could be designed for either continuous or intermittent time slots, depending on the complexity of signalling and hardware realization.

Case 3. Redundancy in time slot resource

Under light traffic load, after all the users' service requests have been fulfilled, there may still be some residual time slots remain unused at the end of a scheduling procedure. In this case, the channel capacity is more than the service requests.

5.1.2.2 Base Station Scheduling Procedures When Adopting MaxC/I Algorithm

Figure 5-3. MaxC/I, Base Station Scheduling Procedures

In MaxC/I, at every scheduling event, all UEs with non-empty source queues are ranked by their C/I values, which is updated every frame. The scheduler first selects the UE with the highest C/I among those non-empty-buffered UEs, and calculates the needed number of time slots based on the C/I value and the buffered packet size. If the needed number of slots exceeds the available slots, then all the available slots will be allocated to the selected UE. Otherwise, the remaining slots can be assigned to the next-highest-C/I-valued and non-empty-buffered UE, and so forth till all the slots in a frame are assigned. Figure 5.4 illustrates an example of three consecutive frames and three UEs. We make the following points:

- Adaptive and multiple number of time slots, ranging from one to maximum fifteen, can be assigned to a particular selected UE, depending on its request and the availability of slot resource.
- The scheduler will skip an empty-buffered UE, even if it has the highest C/I level.
- If a UE is selected in the current scheduled frame, it can be re-selected in the following scheduling event provided that all the requirements of the scheduling algorithm are met, (i.e., it has non-empty buffer and highest C/I at a particular decision point in one scheduling frame,), even if other UEs haven't been served. Refer to section 4.5 for the definition of decision point.
- As can be seen from the example in Figure 5-9, the whole frame can be assigned to only one UE (Frame 3 in the example), or it can be shared among multiple UEs (Frame 4 in the example). It's also possible that some portion of frame is left unused in some scheduling case when the system load is small (Frame 2 in the example).

	Reported C/I (dB)			Requested number of slots to transmit buffered data (calculating from UE's C/I and throughput hull)		
	U1	U2	U3	U1	U2	U3
Frame 1	10	9	2	0	9	2
Frame 2	8	13	0	8	17	9
Frame 3	12	11	-1	10	4	25

Figure 5_4. Example of a MaxC/I Scheduling Case, Number of Users = 3.

5.2. Simulation Result and Performance Analysis in the Whole Sector

The simulation results of both RR and Max C/I schedulers are shown in this section. The statistics are collected from all users throughout the whole sector.

Figure 5-5 shows the system throughput vs. the system load and Figure 5-6 shows the delay vs. the system load. Here we use the average packet call completion time as delay measurement. Figure 5-7 shows the UE satisfaction rate vs. the system load.

Figure 5-5. RR and MaxC/I, System Throughput vs. System Load

Figure 5-6. RR and MaxC/I, Average Packet Call Completion Time vs. System Load

Figure 5-7. RR and MaxC/I, the UE Satisfaction Rate vs. System Load

From the above curves, it can be seen that:

1. When the system load is small, there is no big difference between the two schedulers regarding the system throughput, data delay and UE satisfaction rate. All the data of every UE can be transmitted out of the buffers. With the system load increasing, different schedulers start to show different properties.
2. Considering the system throughput vs. system load, when the system load approaches about 30% of the maximum system capacity, RR starts to get saturated by showing a flat throughput. The buffer of each UE at BS will grow.
3. Under the MaxC/I strategy, the system throughput goes up with the increase of system load, until about 80% of the maximum system capacity (10.8Mbps), at the point of which MaxC/I also shows a saturation. Comparing the two scheduling schemes in the sense of system throughput, MaxC/I is much more efficient than RR.
4. Regarding the packet delay, measured by the average packet call completion time, the curves show similar behavior to the system throughput curves. The delay starts to shoot up at the point where the corresponding throughput gets saturated. This is explainable since a packet call completion time is measured from the arrival of the first packet at the base station until the last packet is received by the UE station. In the light load zone, there's no data accumulated at the buffer. Accordingly, when the packet call head arrives at the buffer, it almost always gets transmitted immediately, generating a very short waiting delay in the queue. When the system saturates under heavy load, packet data wait longer in the queue, resulting a steep increase in the delay.

5. UE satisfaction rate. In the light load zone, all users can remain satisfied under both RR and Max C/I. The satisfaction rate for RR begins to drop earlier than Max C/I. In heavy load zone, RR drops much faster than Max C/I. (The explanation is given in point 6 and 7.)
6. In RR, each UE is allocated the same amount of time for transmission, although the transmission speed varies due to UE's channel condition. Resources are divided up evenly among all UEs. With the number of UEs increasing, it takes longer time for a UE to be polled for transmission, resulting in a smaller throughput of a particular UE. When the throughput of an individual UE drops below a certain threshold, that UE falls into the category of being unsatisfied. As an extreme case when the sector is very heavily loaded, all UEs' active throughput may drop below the threshold, leading to an almost zero UE satisfaction rate.
7. In Max C/I, resources are not evenly distributed among all UEs. UEs closer to the base station are more likely to have good channel conditions (higher C/I), thus have more chances to occupy transmission frames. When the sector load increases, it is those UEs at the cell edge that are first sacrificed, since they are in very disadvantageous positions to compete with those "strong" UEs. Their throughput can be expected to drop quickly below the satisfaction threshold. Even in very heavy load, there will always be a group of UEs in advantageous positions to monopolize the radio resources and remain satisfied. This is why the drop rate of UE satisfaction rate in Max C/I is slower than that of RR.

On one hand, MaxC/I provides a much higher system throughput, since it allocates the frame resource to users with good channel condition that can support higher transmission rate. On the other hand, this high throughput is achieved at the expense of fairness, since the resource may be occupied for a long time by only a small portion of users close to the base station, while leaving other users little chance to be served. In order to further study the relationship between the schedulers' performances and UE locations, it's natural to group UEs according to their distance to the base station, and collect their statistics independently, as we will do in the following section.

5.3. Performance Analysis as a Function of Distance

5.3.1. Grouping Strategy

Figure 5-8. User Grouping Based on Distance

In our simulation, all users are uniformly distributed in a sector and remain static. Let ‘R’ denotes the cell radius, ‘d’ denotes the random variable of each user’s position, and ‘r’ denotes a certain distance, we have:

$$\Pr ob(d < r) = \frac{r^2}{R^2} \quad , \text{ where } r < R; \quad (\text{Eq 5.1})$$

According to the cdf inverse technique, we can derive the formula to generate the random variable d, whose cdf conforms with Eq 5.1.

Let

$$\frac{r^2}{R^2} = u \quad , \quad (\text{Eq 5.2})$$

where u is a random variable uniformly distributed from 0 to 1. By solving Eq 5.2, we get:

$$r = \sqrt{u} \times R \quad , \text{ where } u \sim U[0,1] \quad (\text{Eq 5.3})$$

Eq 5.3 is the equation used to generate the position variable of each user so as to make them evenly distributed in a sector.

We now divide all users into three groups according to their distance so that the number of users in every group will be the same. As illustrated in Figure 5.8, let ‘a’ denotes the radius of the first group, ‘b’ denotes the radius of the second group, and ‘R’ denotes the radius of the third group, which is also the cell radius. To find out the values of ‘a’ and ‘b’, we have:

$$\frac{a^2}{R^2} = \frac{1}{3} \quad (\text{Eq 5.4})$$

$$\frac{b^2}{R^2} = \frac{2}{3} \quad (\text{Eq 5.5})$$

By solving the above equations, we have

$$\begin{cases} a = 0.93Km \\ b = 1.32Km \end{cases}, \text{ where } R = 1.61Km .$$

UEs can be grouped into three groups, i.e., near group, middle group and far group, according to their locations to the cell center.

5.3.2. Simulation Results and Analysis

Figure 5-9 and Figure 5-10 shows the average throughput of each UE in different groups vs. system load for RR and Max C/I schedulers. Figure 5-11 and Figure 5-12 shows the average packet call completion time vs. system load in different groups for RR and Max C/I. Figure 5-13 and Figure 5-14 shows the user satisfaction rate vs. system load in different groups for RR and Max C/I.

Figure 5-9. RR, Average Throughput of Each UE in Different Groups vs. System Load

Figure 5-10. MaxC/I, Average Throughput of Each UE in Different Groups vs. System Load

Figure 5-11. RR, Average Packet Call Completion Time vs. System Load in Different Groups

Figure 5-12. Max C/I, Average Packet Call Completion Time vs. System Load in Different Groups

Figure 5-13. RR, User Satisfaction Rate vs. System Load in Different Groups

Figure 5-14. Max C/I, User Satisfaction Rate vs. System Load in Different Groups

From the simulation results, it can be seen that:

1. When the system load is very light, say less than 30% of the maximum channel capacity, the average throughput of each UE is almost the same, regardless of where UEs are located and which scheduler is used. Within the load range of 30% ~80% of system capacity (which is 10.8Mbps), it starts to show the difference between RR and Max C/I in the UE average throughput.
2. In RR, the UE average throughput in the near group maintains its high throughput till 50% of system capacity before it drops at the rate about 3% with each 10% increase in system load. In the middle group and far group, the UE average throughput drops at a similar rate with the increase of load, except that the drops begin at a lighter load. Although the UE average throughput in each group declines at almost the same rate, the absolute values of throughput in each group are different. Users near base station have higher average throughput than those far from base station, given that each UE is allocated equal amount of transmission time. In this sense, it can be seen that the obtainable UE throughput is affected by its location.
3. In Max C/I, the service quality remains the same in each group till the system load reaches 80% of maximum channel capacity. After that, there shows a great difference in different groups. In the near group, UEs can still maintain an average throughput as high as before, and data packets don't get accumulated in the buffer. This high throughput can be maintained even when the system load is much greater than the channel capacity, say 150% of it. UEs in the near group will always get the best quality of service. In the middle group, the average

throughput declines very slowly from 80~100% of system load. The majority of users can still get high QoS. But in the far group, the average throughput drops dramatically to 50% when the system load increases from 80% to 100%. This fact tells that within MaxC/I, though a considerably high channel throughput is achieved, this high throughput is obtained at the cost of fair allocation when system load increases to more than 80% of the system capacity. In 3G wireless environment, it's very likely that users near base station see statistically higher C/I values than those far from the base station, therefore, near users have far more possibilities than far users to be selected by the MaxC/I scheduling algorithm. The inhomogeneous allocation, when the system works with 80% load or over, is the downside of MaxC/I.

4. The average packet call completion time in each group is closely related to the throughput performance. In RR, the delay increases at a similar rate when the system load increases, through the absolute values vary from every group due to the UE positions. In MaxC/I, delay in the near group is almost negligible as a result of very short waiting time in the buffer. In the far group however, the average delay shoots up after 80% of system load. This phenomenon can be explained by the lower average throughput in the far group. Especially, when the average throughput is lower than the packet arriving rate, packets get accumulated in buffers, causing a much longer waiting time before transmission is permitted.
5. UE satisfaction rate. In RR, all users become unsatisfied at 50% or higher system load in the middle and far group. Even in the near group, more users, if not all, will also get unsatisfied at heavier load. In Max C/I, however, all users in the near

group can remain satisfied at any system load, contrasting with the case in the far group, where the satisfaction rate falls steeply from 1 to 0 when the load changes from 80% to 100%.

5.4. Problems With RR and MaxC/I Schedulers

In this chapter, we studied RR and Max C/I schedulers.

RR can provide fair service in packet scheduling provided that each user can support the same transmission rate over the air. However, for wireless users, the instantaneous supportable data rate is variable due to their locations and the instantaneous channel conditions. As result, the service quality varies with users' locations. Fairness is not achievable as was expected. RR is also very poor in system throughput, delay and UE satisfaction rate performance.

Max C/I is featured by offering high system throughput. In our simulation, the maximum system throughput can be about 80% of the maximum channel capacity. It works well under 80% system load. But when the load exceeds 80%, the average throughput of each UE in the far group drops quickly, while that in the near group still maintains a high level. The inhomogeneity in scheduling is the shortcoming in MaxC/I, especially when the system load is heavy.

None of the two methods are perfectly desirable in a practical system. Trying to find out other scheduling schemes that trade off between fairness and throughput, or if possibly, combine the advantage of the two, is another task of our study.

Chapter 6. MSI – Maximum Self Index Scheduler

6.1. Initial Efforts for Better Scheduling Schemes

In order to find a compromising scheduling scheme, it's natural to combine RR and Max C/I methods into one scheme. A frame is divided two parts, as shown in Figure 6-1. The first part uses RR, the remaining part of the frame is allocated to the users with good channel condition.

Figure 6-1. Suggested Scheme Combining RR and MaxC/I

Two different approaches are proposed for implementing the frame partitioning:

1. Fixed partitioning, where a fixed number of slots are reserved for RR, and the remaining slots use Max C/I scheduling.
2. Adaptive partitioning, where the partitioning is adaptive according to the system load or the length of buffer queue.

Either way, it can be expected that the throughput and fairness performance fall between Max C/I and RR. This expectation is verified by our simulation results.

Another possible scheme is based on the idea of tradeoff between throughput and delay. Rather than partitioning the whole frame, an index is invented based on which the scheduler determines the transmission order among users. This index is defined as:

$$Index = normalized(C/I) + weight_factor \times normalized(buffer_length) \quad (Eq\ 6.1)$$

where *weight_factor* is adjustable and affects the tradeoff between C/I and buffer length.

In this index, three factors are to be considered:

1. UE's instantaneous C/I value is included for the purpose of achieving high throughput;
2. The instantaneous length of buffer queue of a user is included in order to shorten the waiting delay in the buffer;
3. A weighting factor is introduced to balance the above two elements.

The scheduler chooses the UE with the biggest index value for packet transmission. Intuitively, this index favors the users with both good and bad channel conditions. Good channel condition offers high C/I value, which leads to a big index. Also, bad channel condition permits lower UE throughput, which leads to longer buffer queue, which in turn leads to a bigger index. This index is an integration of a pair of contradictions. The normalized C/I tends to contribute to a high system throughput, while the normalized buffer queue length tends to improve the fairness. The weighting factor is adjusted to keep balance between the two contradictions. As an extreme case when the weighting

factor equals zero, the index scheduler turns to the Max C/I scheduler. A bigger weighting factor grants more resource to users in bad channel conditions. This, on one side, equalizes the data transmission amount among all users and provides more fairness, on the other side however, decreases the whole system throughput. By changing the weighting factor from 0 to 1, the performance of index scheduling is expected to fall between MaxC/I and RR. Our simulation results are consistent with this view.

All the above-mentioned proposals cannot exceed the boundaries of throughput and fairness performance, as limited by MaxC/I and RR respectively. Attention must be re-directed to some fresh ideas in seeking for breakthrough.

6.2. Maximum Self Index (MSI) – Idea Analysis and Algorithm

Description

6.2.1. Basic Concept of MSI

Along the way of searching for new scheduling schemes, we noticed an idea based on the proportional fairness criteria described in [9]. Similar to our second suggestion in 6.1, where we used an index as the selection criteria at an scheduling decision point, they proposed an index, which we called the “self index”. In that paper, the scheduler determines the next UE to be served based on two factors: the current C/I, translating into the current supportable data rate of a UE, noted as $R_i(t)$, and the amount of data that has already be transmitted to that UE till the referred time, or equivalently, the average data rate that has been experienced by that UE, noted as $AVG_R_i(t)$. The UE with the highest ratio of $R_i(t)/AVG_R_i(t)$, which is defined the Max Self Index (MSI), is selected by the scheduler at decision point for data transmission.

$$SI(i, t) = \frac{R_i(t)}{avg_R_i(t)} \quad (\text{Eq 6.2})$$

where $SI(i, t)$ is the self index of user i at time t .

In fact, this self index contains the same considerations as in our index in Eq 6.1, but in a form of ratio rather than composite sum. To demonstrate this, we have:

$$\begin{aligned} SI(i, t) &= \frac{R_i(t)}{avg_R_i(t)} \sim \frac{(C/I)_i(t)}{Total_Tx_i(0, t)/t} \\ &\sim \frac{(C/I)_i(t)}{(Total_Arr_i(0, t) - Buffer_Len_i(0, t))/t} \end{aligned} \quad (\text{Eq 6.3})$$

where

$\sim$ stands for “can be translated into”;

$(C/I)_i(t)$ is the instant C/I value of user i at time t ;

$Total_Tx_i(0, t)$ is the total amount of data that had been transmitted by time t for user i ;

$Total_Arr_i(0, t)$ is the total amount of data that had arrived at the buffer of user i by time t ;

$Buffer_Len_i(0, t)$ is the buffer queue length of user i at time t ;

In (Eq 6.3), the total arrived data is statically identical for all users, since every user follows the same traffic pattern. The only two variables are the instantaneous C/I value and the length of buffer queue, same as in the definition of “Index”, as previously described in (Eq 6.1).

In reference [9], the self-index algorithm was applied in cdma2000-HDR system, focusing purely on the effect of sector deployment upon system throughput. Based on the

idea of taking a division operation instead of addition on the related factors, we modified the algorithm and tailored it as MSI to the HSDPA system. In addition, MSI is compared with RR and MaxC/I in the whole sector as well as in different groups.

The definition of Self Index indicates that the MSI scheduler favors those UEs with high C/I values and with long buffer lengths at the same time. It first serves those UEs with relatively high rates and long chunk of data, from the practice of which some improvement in scheduling fairness can be reasonably expected. The system throughput under MSI scheduler however, is hard to foretell and needs to be assessed by simulations.

6.2.2. Base Station Scheduling Procedures Using MSI Strategy

MSI scheduling follows the same procedures as MaxC/I except that MSI ranks UEs by their self indexes, as opposed to C/I values in MaxC/I. Specifically,

- In MSI, channel measurement (C/I) and average data rate (by time t) of a UE is updated for every frame. The UE's average data rate equals the total transmitted data for that UE by time t , divided by the valid time duration of that UE, where the valid time duration is defined as the summation of time slices when the UE's buffer is not empty.
- At every scheduling event, the scheduler calculates all UE's SIs (self index) based on their current supportable data rates derived from their reported C/I values, and average data rates by the scheduling time, and assumes the SIs unchanged through the next coming frame period.

Figure 6- 2. MSI, Base Station Scheduling Procedures

- All UEs with non-empty source queues are ranked by their SI values.
- The scheduler first selects the UE with the highest SI among those non-empty-buffered UEs, and calculates the needed number of time slots based on its currently supportable data rate and its buffered packet size.
- If the needed number of slots exceeds the available slots, then all the available slots will be allocated to the selected UE. Otherwise, the remaining slots can be assigned to the next-highest-C/I-valued and non-empty-buffered UE, and so forth till all the slots in a frame has been exploited.

In MSI, adaptive and multiple number of time slots, ranging from one to maximum fifteen, can be assigned to a particular selected UE, depending on its request and the availability of slot resource. The whole frame can be assigned to only one UE, or be shared among multiple UEs. It's also possible that some portion of frame is left unused in some scheduling case when the system load is extremely small. Similar to MaxC/I, the MSI algorithm permits a previously selected UE to be re-scheduled in the following adjacent frame, so long as that UE possesses the highest SI at some decision point in that frame.

6.3. Simulation Result and Performance Analysis in the Whole Sector

In order to find out the performance of MSI in comparison with RR and Max C/I, simulation for MSI is carried out and the comparison results are shown below.

Figure 6-3. Throughput Performance in the Whole Sector -- System Throughput vs.
System Load

Figure 6-4. Delay Performance in the Whole Sector -- Average Packet Call Completion
Time vs. System Load

Figure 6-5. Fairness Performance in the Whole Sector -- UE Satisfaction Rate vs. System Load

Figure 6-3 shows the throughput vs. the sector load, Figure 6-4 shows the average packet call completion time vs. the sector load, and Figure 6-5 shows the UE satisfaction rate vs. the sector load. All the above results are collected from the whole sector.

Result Analysis:

1. *System throughput:* The system throughput under MSI outperforms not only RR, but also Max C/I! This conclusion is somewhat out of our expectation, in that MaxC/I is commonly believed to bind the high end in system throughput since at each decision moment, the scheduler serves the UE who can support the highest data rate. However, our simulation result shows that the MSI can utilize almost 98% of the channel capacity at 98% system load, as compared to a maximum of

80% channel capacity utilization at 80% system load in Max C/I. The seemingly contradiction will be further investigated and explained in later chapters.

2. *Delay*: The delay shoots up under RR at the load of less than 30% of the system capacity. For MaxC/I and MSI, the delay remains very small up to 75% of system load. From 75% to 95% load, the delay remains small for MSI but it starts to increase gradually for MaxC/I. From 95% to 110% load zone, the delay of Max C/I stops going up while that of MSI starts to increase rapidly. After 110% system load, MSI shoots up, as opposed to the almost constant value in Max C/I. Usually, the system won't work more than 100% load of system capacity, so we're more interested in the 80~100% load zone, where the delay of MSI is shorter than Max C/I.
3. *UE satisfaction rate*: In contrast to RR and Max C/I schedulers where the average UE satisfaction rate of the whole sector goes down slowly as system load increases, the curve in MSI is almost a square wave that turns from a 100% satisfaction rate steeply to zero at 100~110% system load. From this feature, it can be assumed that unlike the other two schedulers, in MSI, there's no big difference in packet transmission service that every user can get. In other words, every user could get the same amount of data transmission regardless of its location, the way in which service fairness among users is achieved to the most extent. This assumption is verified in the next section when MSI is studied in near, middle and far groups.

6.4. Simulation Result and Performance Analysis in Different Groups

Following the same grouping strategy in chapter 5, we further studied MSI in different groups in order to find out the relationship between scheduling performance and users' locations. Simulation results are listed in Figure 6-6, 7, 8.

Figure 6-6. MSI, Average Throughput of Each UE vs. System Load in Different Groups

Figure 6-7. MSI, Average Packet Call Completion Time vs. System Load in Different Groups

Figure 6-8. MSI, UE Satisfaction Rate vs. System Load in Different Groups

Result analysis:

From the above results, it's surprising that the statistics in all three groups are highly identical regarding the average throughput of each UE, the delay and the group- averaged UE satisfaction rate. The average throughput in each group starts to decline at the same rate at the same system load, and users in each group simultaneously become unsatisfied when their throughput drop below the threshold. Any user, located at either the cell center or the edge, will get the same quality of service from base station.

In short, a complete fairness and uniformity is achieved among users within the whole sector under MSI, where the QoS of a near user won't get any better, nor will a far user get any worse. This is extremely different from all other schedulers, and is regarded as one of the most important discoveries in our study. Another important contribution is that we find that MSI can achieve an even higher system throughput than Max C/I. The reasons of these superior properties of MSI will be further investigated and analyzed in chapter 8.

Chapter 7. Comparisons among RR, MaxC/I and MSI by Their Grouped Performances

In Chapter 5 and Chapter 6, we studied the RR, Max C/I and MSI scheduling schemes individually, with statistics from both the whole sector and three groups. In this chapter, instead of analyzing each scheduling method in different groups, we will be focusing on one certain group in which the three scheduling methods are compared with each other.

7.1. Average UE Throughput: Near_Group, Middle_Group, Far_Group

Figure 7-1. Performance Comparison in Near Group -- Average Throughput of Each UE vs. System Load

Figure 7-2. Performance Comparison in Middle Group -- Average Throughput of Each UE vs. System Load

Figure 7-3. Performance Comparison in Far Group -- Average Throughput of Each UE vs. System Load

Figure 7-1 shows the average throughput of each UE in the near group, within which the three scheduling schemes are compared. Figure 7-2 and Figure 7-3 shows middle and far group respectively.

Result analysis:

1. *In the near group*, Max C/I is superior to MSI in that its average throughput of each UE always remains highest (, same as packet arrival rate,) under any system load. The average throughput for MSI starts to decline gradually after 95% system load, and that of RR starts to drop at 40% system load. This is true because for MaxC/I, higher service priority is inclined to be given to users with high C/I levels in the near group, who can support fast data rate and guarantee high average throughput of each UE.
2. *In the middle group*, Max C/I and MSI has almost the same behavior regarding the average UE throughput, remaining the highest before 95% of system load and starting to decline afterwards. The throughput of RR starts to drop even before 30% of system load.
3. *In the far group*, things get reversed between Max C/I and MSI as MSI exhibits a better performance than Max C/I in this group. The average throughput of Max C/I declines significantly after 80% system load, while that of MSI keeps constant as in the near and the middle groups. The reason for this performance transition is that MaxC/I tends to select users in the far group less frequently than in the near group, whereas MSI can schedule packets transmission consistently homogeneously and efficiently in any group. The average throughput of RR in this group starts to drop at an even lighter load than in the middle group.

7.2. Average Packet Call Completion Time: Near_Group, Middle_Group, Far_Group

Figure 7-4. Performance Comparison in Near Group -- Average Packet Call Completion Time of Each UE vs. System Load

Figure 7-5. Performance Comparison in Middle Group -- Average Packet Call

Completion Time of Each UE vs. System Load

Figure 7-6. Performance Comparison in Far Group -- Average Packet Call Completion

Time of Each UE vs. System Load

Figure 7-4 shows the average packet call completion time of each UE in the near group, within which the three scheduling schemes are compared. Figure 7-5 and Figure 7-6 show middle group and far group respectively.

Result analysis:

1. *In the near group*, the delay for Max C/I is almost zero. The delay of MSI remains the same as in Max C/I before 95% of system load, and starts to increase after that. The delay of RR begins to rise at about 40% load.
2. *In the middle group*, the delay of RR increases much faster than in the first group and the other two schedulers. Max C/I and MSI have similar delay performance in this group.
3. *In the far group*, the delay of RR shoots up very steeply even when the system load is very light (lighter than 20% of system load). The delay of Max C/I also increases very rapidly when the system load reaches 70% of its maximum capacity. MSI shows a better delay performance than the others do in this group. Its delay character is the same in this group as in other two groups.

7.3. UE Satisfaction Rate: Near_Group, Middle_Group, Far_Group

Figure 7-7. Performance Comparison in Near Group -- Average UE Satisfaction Rate vs. System Load

Figure 7-8. Performance Comparison in Middle Group -- Average UE Satisfaction Rate vs. System Load

Figure 7-9. Performance Comparison in Far Group -- Average UE Satisfaction Rate vs. System Load

Figure 7-7 shows the average UE satisfaction rate in the near group, within which the three scheduling schemes are compared. Figure 7-8 and Figure 7-9 shows middle group and far group respectively.

Result analysis:

1. *In the near group.* For Max C/I, all UEs' average throughput is above the threshold, therefore, all users remain satisfied even at load as heavy as 150% of maximum system capacity. The UE satisfaction rate in MSI is also 100% before a heavy load of about 110% of maximum system capacity, from where all UEs' throughput drop right below the threshold simultaneously, causing a sudden drop of the satisfaction rate to be zero. The satisfaction rate in RR drops from 100% to

0% when the system load increases steadily from 40% to 110% of its maximum capacity.

2. *In the middle group.* For RR, the satisfaction rate drops more quickly since users in this group get lower throughput than in the near group. The satisfaction performance for RR is much worse than that of the other two. For Max C/I, some users turn to be unsatisfied after 90% load, but there is still a portion of users that can get a satisfactory service when the load increases, although this portion will get smaller and smaller. This is because in this group, there're still some users who can get good channel conditions in some cases, and can keep a throughput higher above the threshold. The satisfaction rate in MSI is similar to the near group.
3. *In the far group.* For MSI, all UEs can still remain satisfied at a heavy load of 100%, while the satisfaction rate for Max C/I drop dramatically from 100% to 0% when the system load increases from 80% to 100%. MSI shows superiority to Max C/I in this group in every sense.

Till now we've compared the performances of the three schedulers in near, middle and far group respectively. We can tell from the results that although Max C/I outperforms MSI in the near group, it loses in the far group where MSI can schedule packet transmission consistently homogeneously and efficiently. MSI works the best among the three schedulers in respect of providing high system throughput and homogeneous fair service under 95% system load, therefore, is the best scheduling scheme as compared to the other two.

Chapter 8. Further Study of MSI

In the previous study of Maximum Self Index scheduler, we've got two important results. One is its uniformity in performance regardless of user's location, the other is the high obtainable system throughput in MSI, which is even higher than that in Max C/I. In this chapter, we'll further study the details of MSI scheme, and provide more explanations for these results.

8.1. Strategy for Sampling Study

In this chapter, different from collecting the statistics of system throughput, delay and user satisfaction, we'll keep track of the samples of some detailed information at each scheduling decision point, such as the locations of UEs that are selected for packet transmission, their C/I values while being selected, the length of buffer queue and the duration of valid time. All these information are categorized as the UE status at the time of being selected for service. Besides, we will also collect the samples of the number of packets in each transmission frame, translated to the system throughput, and the number of decision times in each frame, translated to the number of scheduled users in each frame.

For the purpose of sampling study, the system load is set fixed at 92% of maximum system capacity, yielding to 650 users generating 10M bps load. The figure is chosen based on the fact that at 92% load, both MSI and Max C/I are in crucial transitional stage, yet MSI has a higher throughput than Max C/I, leaving the room for discovering the

reasons behind. The user distribution is also set the same in these two methods. Related samples are collected at each decision point in each frame.

Through this sampling study, we hope to be able to explain the following questions:

1. Why users at the cell edge get as high throughput as those in cell center under MSI, i.e., the homogeneous fair service;
2. Why MSI provides an even higher system throughput than Max C/I.

8.2. Results of Sampling Study and Analysis

8.2.1. Status of Selected UEs

Figure 8-1 and 8-2 shows the probability distribution of locations of UEs while being selected in MSI and Max C/I; Figure 8-3 and 8-4 shows the probability distribution of their C/I values; Figure 8-5 and 8-6 shows the distribution of buffer queue length, and Figure 8-7 and 8-8 shows the distribution of valid time duration.

Figure 8-1. MSI, Probability Distribution of Locations of UEs While Being Selected,

Mean = 1.0056 Km

Figure 8-2. MaxC/I, Probability Distribution of Locations of UEs While Being Selected,

Mean = 0.6689 Km

Figure 8-3. MSI, Probability Distribution of C/I Values of UEs While Being Selected,

Mean = 17.5642 dB

Figure 8-4. MaxC/I, Probability Distribution of C/I Values of UEs While Being Selected,

Mean = 21.7636 dB

Figure 8-5. MSI, Probability Distribution of Buffer Queue Length of UEs While Being Selected, Mean = 18,596 bytes

Figure 8-6. MaxC/I, Probability Distribution of Buffer Queue Length of UEs While Being Selected, Mean = 7,079.4 bytes

Figure 8-7. MSI, Probability Distribution of Valid Time Duration of UEs While Being Selected, Mean = 31.8765 second

Figure 8-8. MaxC/I, Probability Distribution of Valid Time Duration of UEs While Being Selected, Mean = 10.6784 second

Analysis:

1. From Figure 8.1 and 8.2, it can be seen that while Max C/I allocates more resources to the users at nearer location (mean distance=0.6689km), MSI allocates more slots to users farther away (mean distance=1.0056km), as a compensation for bad channel conditions. In MSI, while packets to near user can be transmitted at higher speed, transmissions to far users are scheduled more frequently. This mechanism ensures a user, wherever located, experience coherent transmission rate. In addition, by the definition of SI, longer time of being un-served leads to longer buffered queue and bigger SI, which further leads to more likelihood of being scheduled in the next scheduling event. Reversely, executed transmission service provided to a user increases its average data rate in the past, therefore reduces its SI, which further reduces the possibility of being scheduled in the following frame. With this balancing, MSI scheduler breaks the resource monopoly by those users at advantageous locations, providing homogeneity and complete fairness among all users in the whole embedded sector.
2. Regarding the C/I samples, MSI selects a UE with an average C/I value of 17.5642 dB, in comparison with a 21.7636 dB in Max C/I. It is no surprise that Max C/I has a higher mean C/I value than MSI, but there're three phenomena worth more discussion.
 - Although MSI allocates more time slots to users at farther locations, its average C/I value is not much sacrificed than in Max C/I. This tells us that MSI is able to allocate time slots intelligently to the far users at such moment that they see higher C/I values comparing the historical experiences of their own.

That is, the scheduler holds to serve each UE until it reports an instantaneous supportable data rate comparatively higher than its own average throughput. Having more users in the sector allows the scheduler more choices, which increases the likelihood that when an UE is served, it is always served near the peak data rate during a particular period of time, i.e., at a local maximum[12].

- The MSI scheduler could have selected a UE with C/I value as low as -15dB , as long as the length of its buffer queue is long enough to lead a highest self index at the decision point of scheduling. This could also explain why a UE in disadvantageous location can still be served. In contrast, in Max C/I, the UE, while being selected, must have a C/I higher than 3 dB .
- By looking up the throughput hull in chapter 4, we know that when the C/I value is greater than 16dB , the supportable data rate reaches the maximum 10.8Mbps , and won't go up with any higher C/I. Since the average selected C/I values in Max C/I (21.8dB) and MSI (17.7dB) methods both exceed that threshold, the supported throughput in MSI won't suffer much than Max C/I from a statistical point of view. Put it in other way, under our simulation circumstances, the Max C/I scheduler may have selected a user with unnecessary high C/I value (but probably with short packets arriving at the buffer), given a BER requirement is met. This could partially explain why Max C/I scheme doesn't have an advantage over MSI in throughput performance. The block size of transmission in each frame will further explain why MSI has an even higher throughput, as analyzed in point 3.

3. Now look at the samples of the buffer queue of UEs while being selected. In MSI, the scheduler chooses a UE with an average queue of 18,596 bytes for service. In Max C/I, this queue length is 7,079.4 bytes, less than half of that in MSI. Accordingly, the average valid time duration (the time when buffer is not empty) of a UE while being selected in MSI and Max C/I are 31.8765 seconds and 10.6784 seconds respectively. These values imply that in MSI, the data tend to be held on purpose in the buffer, even that user may have the highest C/I value at that moment, until more packets arrive at the buffer so that the scheduler can schedule a bigger chunk of packets for transmission at the most appropriate moment. In Max C/I, however, when a user has the highest C/I value among all users at the decision moment, it's selected for service no matter how many packets are in the buffer at that moment. This practice could result in squeezing more users with short queues in one frame, leaving more unused gaps in slots, which in effect reduce the amount of data transmitted in one frame, in other words, reduce the UE throughput and system throughput. The transmission length in one frame and the number of decision times in each frame is listed in next section.

8.2.2. Transmission Packet Length and Number of Decision Times in Each Frame

Figure 8-9 and 8-10 shows the probability distribution of transmission packet length in each frame. We also collected the number of decision times in each frame, with a mean of 5 times in Max C/I and 2 times in MSI.

Figure 8-9. MSI, Probability Distribution of Transmitted Packet Length in Each Frame,

Mean = 12,072 byte

Figure 8-10. MaxC/I, Probability Distribution of Transmitted Packet Length in Each

Frame, Mean = 10,894 byte

Analysis:

In MSI, the transmission length in each frame is greater than 12,000 bytes for most cases. In Max C/I, the mean transmission length is 10,894 bytes per frame, translated to a less throughput than MSI. Also considering the number of decision times, or equivalently, the number of users being served in each frame, (as is 5 times in Max C/I and 2 times in MSI,) it can be seen that the MSI schedules bigger chunk of data of each user to fill up the time slots as much as possible, therefore, leaving less slot gaps in between. This gives the reasons of why MSI can schedule more data in a frame, and can gain a higher throughput than Max C/I.

8.3. Summary of Sampling Results

Table 8-1 summarizes the sampling results.

Table 8-1. Summary of Sampling Results of MSI and MaxC/I

Mean Value	Max C/I	MSI
Location	0.6689 Km	1.0056 Km
C/I value	21.7636 dB	17.5642 dB
Buffer length	7,079.4 byte	18,596 byte
Valid time duration	10.6784 second	31.8765 second
Tx length in each frame	10,894 byte	12,072 byte
Number of decisions in each frame	About 5 times	About 2 times

In this chapter, we further studied MSI scheduler in comparison with Max C/I, by sampling the status of UEs while being selected for data transmission, the transmission length and number of decision times in each frame, and answered two questions.

The complete uniformity in MSI is explained by two contributing factors. First, MSI allocates more resources to users at physically farther locations as a compensation for bad channel conditions. In MSI, while packets to near user can be transmitted at higher speed, transmissions to far users are scheduled more frequently. This mechanism ensures a user, wherever located, experience coherent transmission rate. In addition, by the definition of SI, longer time of being un-served leads to longer buffered queue and bigger SI, which further leads to more likelihood of being scheduled in the next scheduling event. Reversely, executed transmission service provided to a user increases its average data rate in the past, therefore reduces its SI, which further reduces the possibility of being scheduled in the following frame. With this balancing, MSI scheduler breaks the resource monopoly by those users at advantageous locations, providing homogeneity and complete fairness among all users in the whole embedded sector.

The higher throughput in MSI is explained by a). transmitting bigger chunk of data, leaving less gap in slots, and b). scheduling a user to be served when its current supportable rate is closer to the peak compared to its historical average data rates. This is done, in effect, by holding some data in the queue before transmitting them out at the most appropriate moment.

In MSI, the scheduler not only compares a particular user's index with other users' indexes, but also compares this user's current supportable data rate with the historical average data rate of its own, a reason why we called this scheme MSI -- maximum self

index. It is right the two-dimensional comparisons that entitles the superiorities in MSI in both system throughput and complete fair service.

Conclusions and Future Works

This thesis studies the frame-by-frame scheduling schemes for WCDMA-HSDPA system. Performances of each scheme are analyzed in terms of system throughput, average packet call completion time and user satisfaction rate. Simulations are carried out in one sector of a 19-cell cellular layout, and statistics are collected both from the whole sector and from different groups, namely, near, middle and far group, in order to find out the relationship between the scheme performances and the users' locations.

Round Robin and Max C/I scheduling schemes are studied in the first half of our thesis.

In RR, assigning equal chance and equal number of time slots to each user couldn't ensure equal amount of data transmission due to the impact of each user's location, and accordingly, the channel condition. RR is not able to provide fairness as was expected before. RR is also very poor in system throughput performance, with only 30% utilization of the downlink shared channel capacity.

Max C/I is characterized by providing high system throughput. In our simulation, the maximum system throughput can reach 80% of the channel capacity. It's high channel efficient and fair under 80% system load. Beyond that, however, the average UE throughput in the far group drops quickly while UEs in the near group can still obtain high average throughput. The inhomogeneity in MaxC/I scheduler heavily affects the

QoS depending on users' physical locations. The unfairness is a shortcoming in Max C/I scheduler, especially when the system load is heavy.

None of the two methods are perfectly desirable in practical system. Trying to find out other scheduling schemes that have a better tradeoff between fairness and throughput is the task in the second half of our study.

The idea of MSI scheme is that the scheduler determines the next UE to be served based on two factors: the current C/I, equivalently, the current supportable data rate of a UE, and the amount of data that has already be transmitted to that UE till the referred time, or equivalently, the average data rate that has been experienced by that UE. The UE with the highest ratio of the two, called maximum self index, is selected for service.

Scheduling protocol is designed based on the above-mentioned idea, and two important results are obtained in MSI scheme. One of them is the uniform performance with any system load, regardless of user's location. A complete fairness among all users is achieved in that any user can get the same service quality no matter where it locates. The other important result is that MSI can achieve an even higher system throughput than MaxC/I, which contradicts the traditional belief that Max C/I binds the upper boundary of the system throughput. In our simulation, MSI can utilize almost 95% of the channel capacity at 95% system load, as compared to a maximum of 80% channel utilization at 80% system load in Max C/I. MSI outperforms both RR and Max C/I in both system

throughput and fairness, therefore, is the best scheme for packet scheduling in the said system.

In order to explain the superiority in MSI, further investigation is carried out by collecting the samples and probability distributions of the status of UEs while being selected for service, the transmission length and number of decision times in each frame.

The complete uniformity in MSI is explained by two contributing factors. First, MSI allocates more resources to users at physically farther locations as a compensation for bad channel conditions. In MSI, while packets to near user can be transmitted at higher speed, transmissions to far users are scheduled more frequently. This mechanism ensures a user, wherever located, experience coherent transmission rate. In addition, by the definition of SI, longer time of being un-served leads to longer buffered queue and bigger SI, which further leads to more likelihood of being scheduled in the next scheduling event. Reversely, executed transmission service provided to a user increases its average data rate in the past, therefore reduces its SI, which further reduces the possibility of being scheduled in the following frame. With this balancing, MSI scheduler breaks the resource monopoly by those users at advantageous locations, providing homogeneity and complete fairness among all users in the whole embedded sector.

The higher throughput in MSI is explained by a). transmitting bigger chunk of data, leaving less gap in slots, and b). scheduling a user to be served when its current

supportable rate is closer to the peak compared to its historical average data rates. This is done, in effect, by holding some data in the queue before transmitting them out at the most appropriate moment.

In MSI, the scheduler not only compares a user's index with other users' indexes, but also compares this user's current supportable data rate with the historical average data rate of its own, this is where the name of maximum self index (MSI) arises. It is right the two-dimensional comparisons that entitles the superiorities in MSI in both system throughput and fair service.

Future works can be done from several aspects based on our current system model.

1. The traffic model can be improved as a combination of different traffic types rather than just web browsing, thus a more realistic multimedia data packets can be depicted and simulated.
2. In our simulation, all users are stationary. Future simulation can be improved by introducing user mobility.
3. In our research, the system load is increased by increasing the number of users, where every user follows the same traffic model. As an alternative, we can fix the number of users and increase their traffic load. This could reduce the probability of having users in favorite positions.

4. In our simulation, each user is associated with a unique buffer. If one user is able to open up multiple connections of different service types, the buffer design should be modified to accommodate the multi-session, different service scenario.

References

1. Harri Holma, Antti Toskala, **“WCDMA for UMTS: Radio Access for Third Generation Mobile Communication”**, Published by John Wiley & Sons Inc., June, 2000.
2. 3GPP; Technical Specification Group Radio Access Network; Physical Layer Aspects of UTRA High Speed Downlink Packet Access (Release 2000), Serial Code: 3G TR25.848 V0.6.0 (2000-05).
3. ETSI, TR UMTS 30.03, **“Selection procedures for the choice of radio transmission technologies of the UMTS”**, version 3.2.0, April, 1998.
4. R.Hafez, Hua Sun, **“Fast Scheduling Strategies for WCDMA-High Speed Downlink Packet Access (HSDPA) System”**, Submitted for publication, Page1-6, 2002.
5. Kam, A.C.; Minn, T.; Kai-Yeung Siu, **“Supporting rate guarantee and fair access for bursty data traffic in wcdma”**, Selected Areas in Communications, IEEE Journal on , Volume: 19 Issue: 11 , Nov 2001, Page(s): 2121 -2130
6. Dimou, K.; Godlewski, P. **“MAC scheduling for uplink transmission in UMTS WCDMA”**, Vehicular Technology Conference, 2001. VTC 2001 Spring. IEEE VTS 53rd , Volume: 4 , 2001, Page(s): 2625 -2629
7. Vannithamby, R.; Sousa, E.S., **“Resource allocation and scheduling schemes for WCDMA downlinks”**, International Conference on Communications, 2001.

- ICC 2001. IEEE International Conference on , Volume: 5 , 2001, Page(s): 1406 - 1410
8. Kazmi, M.; Godlewski, P.; Cordier, C., **“Admission control strategy and scheduling algorithms for downlink packet transmission in WCDMA”**, Vehicular Technology Conference, 2000. IEEE VTS Fall VTC 2000. 52nd , Volume: 2 , 2000,Page(s):674-680,vol.2.
9. A.Jalali, R Padovani, R.Pankaj, **“Data Throughput of CDMA-HDR, a High Efficiency High Data Rate Personal Communication Wireless System”**, IEEE 51st Vehicular Technology Conference, Tokyo, Japan - May - 2000, page 1854-1858.
10. Paolo Goria, Davide Sorbara, **“System Performance Evaluation of Packet Data Transfer in UTRA-FDD Standard”**, 2000 IEEE Radio and Wireless Conference, Denver, Colorado, USA, September, 2000, page 95-98.
11. Hong Lu, **“Optimal Allocation of Power to AMCS for Maximum Throughput in WCDMA”**, Graduate Thesis in System and Computer Engineering Department, Carleton University, Sept. 2002.
12. Y.Jou, **“Development in Third Generation CDMA Technology”**, Spread Spectrum Techniques and Applications, 2000 IEEE Sixth International Symposium on Volume: 2, 2000, Page(s): 460 -464 vol.2.
13. Gerald Ostermayer, Peter Slanina, Christian Hölzl, Christoph F. Mecklenbräuer, Fariba Raji and Thomas Stadler: **“Scheduling Algorithm for UTRA TDD**

- Mode**", in Proc. AFCEA/IEEE EuroComm 2000, pp. 212-216, Munich, Germany, May 17-19, 2000.
14. Yuming Lu, Robert W. Brodersen, "**Integrating Power Control, Error Correction Coding and Scheduling for a CDMA Downlink System**", IEEE Journal on Selected Areas in Communications, Vol. 17, No.5, Page 978-989, June, 1999.
 15. <http://mathworld.wolfram.com/ParetoDistribution.html>
 16. Ok-Sun Lee; Younghan Han; Jiwoong Lee, "**Transmission scheduling scheme for multicode WCDMA/TDD networks**", Vehicular Technology Conference, 2001. VTC 2001 Fall. IEEE VTS 54th , Volume: 2 , 2001, Page(s): 812 -816
 17. Mihailescu, C.; Lagrange, X.; Godlewski, P., "**Radio resource management for packet transmission in UMTS WCDMA system**", Vehicular Technology Conference, 1999. VTC 1999 - Fall. IEEE VTS 50th , Volume: 1 , 1999, Page(s): 573 -577 vol.1.
 18. Téllez Labao C., Stopped González E., Ruiz Garcí'a M., and Diaz Star To, "**Capacity Analysis of UTRA/TDD Systems** ", Proceedings of IEEE Vehicular Technology Conference (VTC'2000), Boston, September 2000
 19. Jeongrok Yang, Insoo Koo, Yeongyoon Choi and Kiseon Kim, "**A Dynamic Resource Allocation Scheme to Maximize Throughput in a Multimedia CDMA System**", VTC'99-Fall, Vol. 1, pp. 348-351, Sep., 1999.
 20. <http://www.umtsworld.com/technology/hsdpa.htm>