

Adaptive Token Bank Fair Queuing Scheduling in the Downlink of 4G Wireless Networks

by

Feroz A. Bokhari

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
Faculty of Engineering
Department of Systems and Computer Engineering
Carleton University
December, 2007

© Copyright 2007, Feroz A. Bokhari

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

Adaptive Token Bank Fair Queuing Scheduling in the Downlink of 4G Wireless Networks

submitted by Feroz A. Bokhari
in partial fulfillment of the requirements for
the degree of Master of Applied Science in Electrical Engineering

**Thesis Co-Supervisor
Prof. Halim Yanikomeroglu**

**Thesis Co-Supervisor
Dr. William K. Wong**

**Chair, Department of Systems and Computer Engineering
Prof. Victor Aitken**

Carleton University
December, 2007

Abstract

Traditionally, the research on packet scheduling has focused mostly on QoS and fairness for different QoS classes or different applications, while opportunistic scheduling algorithms have focused on exploiting the time-varying nature of the wireless channels and to provide fairness to the different mobile users. This segregation between packet scheduling and radio resource allocation is not efficient since none of the two types of scheduling algorithms focus both on providing QoS for the applications and exploiting the time-varying characteristics of the wireless channel. For these reasons, it is necessary to merge the scheduling of packets and the allocation of radio resources to design cross-layer scheduling algorithms.

Adaptive Token Bank Fair Queuing (ATBFQ) algorithm is the proposed algorithm for cross-layer scheduling. The ATBFQ is the modified version of the Token Bank Fair Queuing (TBFQ) algorithm which was initially proposed for single carrier time division multiple access (TDMA) systems. It takes higher layer QoS attributes such as priorities, interflow fairness and delay constraints into account. By selecting the user terminals (UTs) in a certain prioritized manner derived from these QoS attributes, we can improve the performance of the UTs suffering from bad interference conditions and shadowing in particular. The ATBFQ algorithm is designed to accommodate the bursty nature of traffic. This is done by the graceful acceptance of traffic profile violation when bandwidth is available, provided the UT does not exceed its bandwidth allocation in the long term. This prevents sudden degradation of QoS experienced by the end user as a result of traffic profile violations or interference in the wireless environment.

For the radio resource allocation, channel feedback is required from every UT at the start of every scheduling instant. Based on the decisions made in the first level of scheduling with QoS provisioning, appropriate resources are assigned to the selected UTs taking into account the channel quality information (CQI). The maximum signal to interference noise ratio (SINR) method is used for the resource allocation where the best chunk is allocated to the selected UT.

The proposed scheduling algorithm is tested in a multicell environment in the presence of intercell interference. The performance is compared to that of the Score Based (SB) algorithm which is a variation of the Proportional Fair (PF) algorithm (the most widely adapted opportunistic fair scheduling technique) and the round robin (RR) method. The performance is studied in the context of the wide area scenario. QoS issues in terms of throughput, packet drop ratios, and queuing delays are addressed. Furthermore, a fairness analysis is shown highlighting the performance of ATBFQ, SB, and RR.

It is observed from simulation results that the proposed scheme provides better fairness in terms of queuing delays, and dropped packets for various loading factors, while the throughput remains comparable. A gain in the performance of cell edge users is also observed in the proposed scheme, this may result in substantial savings in the deployment cost since a fewer number of base stations (BS) will be needed to cover regions.

Acknowledgements

First and foremost, I would like to express my sincerest appreciation to my thesis co supervisors Prof. Halim Yanikomeroglu and Dr. William Wong for their guidance and instruction in achieving this work. Their profound knowledge and passion for research have greatly enhanced my enjoyment of this research.

In addition, I am grateful to Mr. Mahmudur Rahman for providing me constant motivation and feedback during the course of this project. I would also like to thank Mr. Jiangxin Hu and Mr. Ivan Lee for their help.

I am also grateful to the technical support staff in the Systems and Computer Engineering Department for installing the hardware and software needed as tools during the course of this project. I would also like to thank them for maintaining these tools; for being available to rectify software or hardware problems when they occurred.

Finally, I would like to recognize my family for their continual encouragement and understanding and without whose support this work would not have been possible.

Table of Contents

Abstract.....	iii
Acknowledgements	v
Table of Contents	vi
List of Figures.....	viii
List of Tables	x
List of Acronyms	xi
List of Symbols	xiv
Chapter 1 Introduction	1
1.1 Introduction.....	1
1.2 Scheduling in Wireless Networks	3
1.2.1 <i>Non-Queue-Aware Cross Layer Scheduling</i>	4
1.2.2 <i>Queue-Aware Cross Layer Scheduling</i>	5
1.3 Thesis Motivation and Objective	6
1.4 Scope of the Thesis	7
1.5 Organization of the thesis	8
Chapter 2 Token Bank Fair Queuing Algorithm	10
2.1 Introduction.....	10
2.2 Leaky Bucket	10
2.3 Token Bank Fair Queuing Algorithm	11
2.4 TBFQ Functionality	15
2.5 TBFQ Complexity	18
Chapter 3 Overview of the WINNER Architecture	19
3.1 Introduction.....	19
3.1.1 <i>Internet Protocol Convergence (IPC) layer</i>	20
3.1.2 <i>Radio link control (RLC) layer</i>	20
3.1.3 <i>Medium access control (MAC) layer</i>	20
3.1.4 <i>Physical (PHY) layer</i>	21
3.2 Physical Layer Modes	21
3.3 Chunk, Slot, Frame, and Super-frame Definitions.....	22
3.4 Resource Scheduling.....	25
3.4.1 <i>Score Based (SB) Algorithm</i>	26
3.4.2 <i>SB Parameter Selection</i>	28

Chapter 4 Adaptive Token Bank Fair Queuing Algorithm.....	31
4.1 Introduction to ATBFQ.....	31
4.2 ATBFQ Parameter Selection	35
Chapter 5 System Level Simulation Model	37
5.1 Introduction.....	37
5.2 Channel Model.....	38
5.2.1 Large Scale Path Loss Model	39
5.2.2 Shadowing.....	40
5.2.3 Fading	40
5.3 Background Noise Model	42
5.4 Adaptive Modulation and Coding.....	43
5.5 Interference Model.....	45
5.5.1 Wrap-around Model.....	45
5.5.2 Central Cell Model	45
5.6 Traffic Model	48
5.7 System View of the Simulation Model	51
5.8 Summary of the Simulation Assumptions	55
Chapter 6 Simulation Results.....	56
6.1 Assessment Criteria	56
6.2 Spectral Efficiency	57
6.3 Queuing Delay	59
6.4 Packets Dropped	63
6.5 SINR	66
6.6 Throughput.....	67
6.7 Performance vs. Distance.....	70
6.8 Fairness Analysis	75
Chapter 7 Conclusions and Proposals for Future Work.....	80
7.1 Conclusions.....	80
7.2 Thesis Contribution.....	81
7.3 Recommendations for Future Research Works.....	83
References.....	84
Appendix A	90
Appendix B	92

List of Figures

Figure 2-1 Leaky bucket mechanism	11
Figure 2-2 TBFQ downlink structure.....	13
Figure 3-1: The layer and services of WINNER (taken from [27])	19
Figure 3-2: a) Multi--carrier downlink physical channel structure. b) Layered time and frequency chunks for multiple input multiple output (MIMO) transmission (taken from [28])	23
Figure 3-3 Summary of chunk sizes in the two physical layer modes (taken from [28])	24
Figure 3-4 WINNER MAC Super Frame structure for the FDD case (taken from [28])	24
Figure 3-5 CDF of average user throughput for SB scheduler for different window sizes	29
Figure 3-6 CDF of average user queuing delay for SB scheduler for different window sizes	29
Figure 3-7 CDF of packets dropped per frame for SB scheduler for different window sizes.....	30
Figure 4-1 Overview of the ATBFQ scheduling operation	31
Figure 4-2 Pseudo-code for ATBFQ.....	34
Figure 5-1 Instantaneous power of the fading envelope (shown for 1 sec)	41
Figure 5-2 Time correlation of the fading samples (zoomed version of Figure 4-1).....	42
Figure 5-3 Throughput per chunk versus SINR for the baseline AMC shown for the hull curves.....	44
Figure 5-4 Network layout under study	47
Figure 5-5 State transition diagram of a 2IRP Process	50
Figure 5-6 2IRP video traffic model (Taken form [37])	51
Figure 5-7 Two levels of Scheduling.....	52
Figure 5-8 System level view of the simulation setup	54
Figure 6-1 Average spectral efficiency vs. number of users.....	58
Figure 6-2 Average spectral efficiency vs. activity factor for low and high loading.....	59
Figure 6-3 Average user queuing delay for medium and high interference scenarios.....	60
Figure 6-4 Average queuing delay vs. different activity factors for low and high loading	61
Figure 6-5 CDF of average queuing delay for low loading (8 users)	62
Figure 6-6 CDF of average user queuing delay for high loading (20 Users).....	62
Figure 6-7 Average packets dropped vs. number of users	64
Figure 6-8 Average packets dropped vs. different activity factor for low and high loading	64
Figure 6-9 CDF of packets dropped per frame for 8 users at AF=0.5, 0.7	65
Figure 6-10 CDF of packets dropped per frame for 20 users at AF=0.5, 0.7	65
Figure 6-11 CDF of SINR for 8 users on scheduled chunks.....	66
Figure 6-12 CDF of SINR for 20 users on scheduled chunks.....	67

Figure 6-13 CDF of throughput (bytes per frame per sector) for 8 users	68
Figure 6-14 CDF of throughput (bytes per frame per sector) for 20 users	69
Figure 6-15 Total sector throughput vs. number of users	70
Figure 6-16 Ratio of packets transmitted vs. distance from BS.....	71
Figure 6-17 Average queuing delay per user vs. distance from BS.....	71
Figure 6-18 CDF of average queuing delay for cell edge User 2 for low loading case	73
Figure 6-19 CDF of average queuing delay for User1 in a high loading scenario	74
Figure 6-20 CDF of average queuing delay for User 2 in a high loading scenario	74
Figure 6-21 Average short term fairness for low loading case (shown over 35 sec of simulation time)....	77
Figure 6-22 Average short term fairness for high loading	77
Figure 6-23 CDF for long term fairness for low loading case	78
Figure 6-24 CDF for long term fairness for high loading case	79
Figure 7-1 BLER vs SINR for Block length of 1728 bits.....	94

List of Tables

Table 4-1 Burst credit for ATBFQ for low loading (8 users)	36
Table 4-2 Burst credit for ATBFQ for high loading (20 users)	36
Table 5-1 Baseline modulation and coding scheme for adaptive modulation and coding.....	43
Table 5-2 AMC mode for information block-size of 1728 bits	44
Table 5-3Traffic Model Parameters of the Video Stream.....	51
Table 5-4 Summary of simulation assumptions.....	55
Table 6-1 Comparison of ATBFQ, SB and the RR algorithms for a low loading scenario.....	72
Table 6-2 Comparison of ATBFQ, SB and the RR algorithms for a high loading scenario.....	73
Table 7-1OFDM/GMC parameters.....	90
Table 7-2 Frame parameters in WINNER	91

List of Acronyms

1G	1 st Generation
2G	2 nd Generation
3G	3 rd Generation
4G	4 th Generation
AF	Activity factor
AMC	Adaptive coding and modulation
AWGN	Additive white Gaussian noise
BLDPCC	Block low density parity code check
BLER	Block error rate
BPSK	Binary phase shift keying
BS	Base station
CDF	Cumulative distribution function
CP-OFDM	Cyclic prefix- orthogonal frequency division multiplexing
CQI	Channel quality information
CSI	Channel state information
DL	Down-link
FDD	Frequency division duplex
FEC	Forward error correction
GMC	Generalized multicarrier
GPS	Generalized processor sharing
HARQ	Hybrid automatic repeat request
HOL	Head-of-the-line
IP	Internet protocol
IPC	Internet protocol convergence layer
IRP	Independent renewal process
ITU	International telecommunication union

LB	Leaky bucket
LS	Link to system
MAC	Medium access control
MCS	Modulation and coding size
NF	Noise figure
OFDM	Orthogonal frequency division multiplexing
OFDMA	Orthogonal frequency division multiple access
PDU	Protocol data unit
PF	Proportional fair
PFQ	Per-flow queuing
PHY	Physical layer
PL	Pathloss
PLM	Physical layer mode
QAM	Quadrature amplitude modulation
QoS	Quality-of-service
QPSK	Quadrature phase shift queuing
RAN	Random access node
RLC	Radio link control
RRM	Radio resource management
SB	Score based
SF	Super-frame
SINR	Signal-to-interference noise ratio
SISO	Single input single output
TBFQ	Token bank fair queuing
TDD	Time division duplex
TDMA	Time division multiple access
UT	User terminal
WA	Wide area
WFQ	Weighted fair queuing
WINNER	Wireless World INitiative New Radio

WLAN

Wireless local area network

VoIP

Voice over IP

List of Symbols

ϕ_i	Priority assigned to user i
$\gamma(i)$	SINR in segment i of a packet
γ_{ij}	SINR of user i in chunk j
π	Constant, 3.141592654
ρ	Pareto parameter for location
β	Pareto parameter for shape
Λ	Traffic arrival rate
λ	Wavelength
β_1	Pareto parameter for ON time distribution
β_2	Pareto parameter for OFF time distribution
$\mu_{off \rightarrow on}$	State transition rate from OFF to ON
$\mu_{on \rightarrow off}$	State transition rate from ON to OFF
B	Channel bandwidth
b	Bucket size
C	Boltzman's constant
c_i	Burst credit for flow i
D_i	Data buffer size
d	Propagation distance
d_i	Debt limit for flow i
d_0	Reference distance
E_i	Token counter for flow i
f	Operating frequency
f_c	Carrier frequency
F_I	Fairness index
f_m	Maximum Doppler frequency

f_X	PDF of the Pareto distribution
h_{bs}	Height of BS
L	Length of packet
M	M-ary modulation
N	Number of samples to represent Doppler spectrum
N_{act}	Total number of active users
N_{chunks}	Total number of chunks
N_{UT}	Total number of user terminals
n	Propagation exponent
n_f	Number of frames
n_{sub}	Number of subcarriers for OFDM
n_{symp}	Number of symbols for OFDM
P_i	Priority index for user i
P_N	Average thermal noise power
PL	Path loss in dB
PL_{fs}	Free space path loss
P_{roff}	Probability being OFF state
P_{ron}	Probability being ON state
p_i	Token pool size
r	Rayleigh random variable
r_i	Token generation rate
r_c	Coding rate
r_s	Symbol rate
$S(f)$	Power spectral density
s_j	Score of user j
T_k	Ambient temperature in °Kelvin
T_c	Coherence time
T_{off}	Mean dwell time on OFF state
T_{on}	Mean dwell time on ON state
t_k	Slot at time k

W_i	Head of the line packet delay in user i 's buffer
X_σ	Gaussian random variable with a standard deviation of σ dB

Chapter 1

Introduction

1.1 Introduction

The last decade has witnessed a tremendous growth in the wireless market. 1G (analog voice) and 2G (digital voice/low-rate data) wireless networks have been ubiquitously deployed. To meet the growing demands in the number of subscribers, rates required for high speed data transfer and multimedia applications 3G (third generation) standards started evolving. Now, the approaching 4G (fourth generation) mobile communication systems (although currently in the research phase) are projected to solve still-remaining problems of 3G systems and to provide a wide variety of new services ranging from high-quality voice to high-definition video.

The International Telecommunication Union (ITU) is an international organization established to standardize and regulate international radio and telecommunications. The ITU Radio-communication division (ITU-R) is one of the three divisions of the ITU and is responsible for radio communication. Its role is to manage the international radio-frequency spectrum and satellite orbit resources and to develop standards for radio communications systems with the objective of ensuring the effective use of the spectrum. The ITU-R vision for systems beyond 3G comprises two major paths:

- existing and evolving access systems will be integrated on a packet-based platform to enable cooperation and interworking of these systems in the sense "optimally connected anywhere, anytime" and,
- the radio access system for new mobile access and new nomadic/local area wireless access will be developed to provide access with significantly improved performance compared to today's systems.

One such example of a 4G wireless network vision is being developed in the WINNER (Wireless World INitiative New Radio) project. The key objective of the WINNER project is to develop an innovative concept in radio access in order to address high flexibility and scalability

with respect to data rates and radio environments. The future converged wireless world requires in the long-term perspective ubiquitous radio system instead of disparate systems for different purposes (cellular, wireless local area networks, short-range access, etc.). The proposed vision of a ubiquitous radio system concept is to provide wireless access for a wide range of services and applications across all environments, from short-range to wide-area, with one single adaptive system concept. Compared to current and evolving mobile and wireless systems, the WINNER system concept aims to provide significant improvements in peak data rate, latency, mobile speed, spectrum efficiency, coverage, cost per bit and supported environments taking into account specified QoS requirements [1]. The focus of the WINNER project is the development of this radio access system by taking into account the interworking with other systems. The envisioned capabilities of the new components of future mobile and wireless communication systems have the following peak cell data rates:

- up to approximately 100 Mbps for the new mobile access and
- up to approximately 1 Gbps for new nomadic / local area wireless access.

The need for such a system arises due to the significant growth of the global demand for wireless bandwidth [2]. Compared with wireline networks, wireless resource is very scarce. While more wired network bandwidth is created when new physical resources (cable, fiber, router, etc.) are added to the network, wireless communication requires sharing a finite natural resource: the radio frequency spectrum. The data-rate capacity that a radio frequency channel can support is limited by Shannon's capacity laws [3]. Hence, the allocation and management of resources are crucial for wireless networks.

In the current dominant layered networking architecture, each layer is designed and operated independently to support transparency between layers. Among these layers, the physical layer is in charge of raw-bit transmission, and the medium access control (MAC) layer controls multiuser access to the shared resources. However, wireless channels suffer from time-varying multipath fading; moreover, the statistical channel characteristics of different users are different. The sub-optimality and inflexibility of this architecture result in inefficient resource utilization in

wireless networks. Therefore, cross-layer design across the physical and MAC layers are desired for wireless resource allocation and packet scheduling [4, 5].

The objective of this thesis is to establish an efficient cross-layer framework designed for resource allocation in downlink of the WINNER multicarrier network. This research focuses on studies on the mechanisms of spectral efficiency, fairness, as well as QoS provisioning and algorithm development for resource allocation in multiuser frequency-selective fading environments. This is a joint work done by Carleton University in collaboration with Communication Research Centre (CRC), Canada.

1.2 Scheduling in Wireless Networks

Research within the field of scheduling packets of wire-line networks has matured extensively during the last two decades. Much of this research has focused on scheduling algorithms similar to the Weighted Fair Queuing (WFQ) algorithm [6] which is a packet-based version of Generalized Processor Sharing (GPS) [7]. This is because GPS can guarantee to the different applications (sessions) that the network resources are allocated fairly and independently of the behavior of the other applications [8]. Most of the publications on packet scheduling assume that the throughput of the channel is constant.

For wireless networks, the research has mainly concentrated on how to schedule radio resources, e.g., time-slots, frequencies, powers and/or codes, to different mobile users. Most of these scheduling algorithms do not take the users' QoS requirements into account and mainly focus on how to exploit the time-varying nature of the wireless channels in order to increase the throughput. Such schedulers are also called as opportunistic scheduling schemes.

Traditionally, the research on packet scheduling has focused mostly on QoS and fairness for different QoS classes or different applications, while opportunistic scheduling algorithms have focused on exploiting the time-varying nature of the wireless channels and to provide

fairness to the different mobile users. This segregation between packet scheduling and radio resource scheduling is not efficient since none of the two types of scheduling algorithms focus both on

- (i) providing QoS for the applications and
- (ii) exploiting the time-varying characteristics of the wireless channel.

For these reasons, it is necessary to merge the scheduling of packets and the allocation of radio resources to design cross-layer scheduling algorithms [9].

To be able to improve the QoS experienced by the mobile users, cross-layer scheduling algorithms need to take both the time-varying characteristics of the wireless channels and the QoS demands of the applications into account. In addition, it is often necessary to consider the characteristics of the packet load of the buffers at the mobile users or the BS containing packets waiting to be transmitted over the uplink or downlink, respectively [10]. In this section, cross-layer scheduling algorithms that are designed to improve the QoS in the network will be described. Both *non-queue-aware* and *queue-aware* scheduling algorithms are considered. While the non-queue-aware algorithms do not consider how the queues of the buffers can affect the QoS, the queue-aware algorithms consider effects like queuing delay, buffer overflow, and probability of empty buffers.

1.2.1 Non-Queue-Aware Cross Layer Scheduling

Physical and MAC related design issues can be analyzed by assuming that all the users are *back-logged*, i.e., all the users in the system have nonempty buffers that always contain packets to send or receive. However, when analyzing the QoS performance of scheduling algorithms this assumption is not always correct since the number of packets in the buffers can vary significantly, and there is a relatively high probability that the buffers are empty [10,11]. However, since the scheduling algorithms in modern cellular networks operate on time-scales that are significantly shorter than the time-scale over which the population of back-logged users changes, it can nevertheless be assumed that the scheduling algorithms operate on a constant user population [11].

In [12], Andrews et al. assumed a constant user population and proposed scheduling algorithms that aim at offering throughput guarantees by giving different priorities to the users depending on how far they are from fulfilling their throughput guarantees. However, one of the problems with this algorithm is that it takes action only when a throughput guarantee already has been violated. As an alternative, Borst and Whiting proposed a scheduling algorithm that tries to fulfill the throughput guarantees *before* they are violated [13]. This algorithm is also based on assuming a constant user population and is based on a mathematical proof showing that the algorithm provides the highest theoretically attainable throughput guarantees to the mobile users in a cell.

1.2.2 Queue-Aware Cross Layer Scheduling

For time-slotted networks, the packets in the queues are aggregated into time-slots. Consequently, empty queues and partially filled time-slots will affect the system performance. In recent years, some research has considered how to integrate the packet scheduling and the radio resource scheduling into queue and channel-aware scheduling algorithms [9, 11, 14–17,18]. For example, one such publication handles how to implement WFQ when the largest share of the radio resources is given to the users with the instantaneously best channel conditions in a Code Division Multiplexing (CDM) based network [19]. However, the most well-known queue and channel aware scheduling algorithm is arguably the Modified Largest Weighted Delay First (M-LWDF) algorithm [18], where the scheduled user in time slot t_k is selected according to the following rule

$$i^*(t_k) = \arg \max_{1 \leq i \leq N} (\phi_i W_i(t_k) \frac{r_i(t_k)}{\bar{r}_i}), \quad 1-1$$

where $W_i(t_k)$ [seconds] is the head-of-the-line (HOL) packet delay in user i 's buffer, ϕ_i is a constant denoting the priority given to user i , $r_i(t_k)$ is the rate for user i in the time slot t_k , \bar{r}_i [bits/second] is the average rate for user i and N is the total number of users. This algorithm can

be used both for uplink and downlink scheduling since W_i can denote the delay of the HOL packets in either the users' output buffers on the uplink or the buffers at the base station containing packets for downlink transmission to each of the mobile users. The advantage of this algorithm is that it takes both the channel quality and the delay of the packets into account when performing scheduling. In addition, this algorithm is proven to be throughput optimal. This means that the algorithm manages to keep the queues stable if this is at all feasible to do with any other algorithm, where a stable queue is defined as having a finite expected queue length. The M-LWDF algorithm can also be reformulated to guarantee a certain throughput to the users if it is used in conjunction with a token bucket control [18, 34]. Another well-known queue and channel-aware scheduling algorithm is the *exponential rule* developed by Shakkottai and Stolyar [15]. This scheduling algorithm is also proved to be throughput optimal and can also be used to provide QoS guarantees in a cellular network.

In [9], a general queue and channel-aware scheduling algorithm providing QoS guarantees is developed. It is also thoroughly described how the adaptive coding and modulation and the scheduling algorithm is going to be implemented at the MAC layer of a IEEE 802.16-based network.

1.3 Thesis Motivation and Objective

Multiple Access (MA) techniques allow users to share the available bandwidth by allotting each user some fraction of the total system resources. Research has shown that dramatic performance differences are possible between various multiple access strategies. The diverse nature of the anticipated WINNER traffic – VoIP, data transfer and video streaming and the challenging aspects of the system deployment – mobility, neighboring cells, high required bandwidth efficiency, make the MA issues quite complicated. The proposed MA scheme for the WINNER network is Orthogonal Frequency Division Multiple Access/ Time Division Multiple Access (OFDMA/TDMA), whereby users share subcarriers and time slots. It is a hybrid of Frequency Division Multiple Access (FDMA) and TDMA. The advantages of OFDMA start with the advantages of OFDM in terms of robustness to inter-symbol interference (ISI) that arise

from frequency selective fading. In addition, OFDMA is a flexible MA technique due to its smaller granularity of resources which exploits the fact that each subchannel can be allocated, assigned power and adaptive rates independently. By using such measures, it can accommodate many users with widely ranging applications, data rates, and QoS requirements. This allows sophisticated time- and frequency domain scheduling algorithms to be integrated in order to best serve the user population.

Taking the above mentioned MA requirements into account, a scheduling scheme has to be designed that has low complexity, and is robust against severe wireless channel conditions. It has to be resilient to interference and at the same time sensitive to the various user QoS requirements. For this purpose, the scheduling in wireless systems can be split into two distinct operations, i.e., resource scheduling, and radio resource allocation. Resource scheduling deals with the task of selecting users based on their priorities, queue levels, and interflow fairness. The primary constraint at this level is to meet the QoS requirements. Radio resource allocation on the other hand, deals with the task of “channel aware scheduling”, where the allocation of resources is done adaptively and dynamically based on channel state information (CSI). The key idea of the channel-aware scheduling is to choose a user with good channel conditions to transmit packets [21]. Taking advantage of the independent channel variation across users, channel-aware scheduling can substantially improve the network throughput through multiuser diversity, whose gain increases with the number of users [20, 21]. The objective of this thesis is to develop a queue and channel aware scheduling scheme which takes QoS parameters into account for all users. For this purpose the TBFQ algorithm which was originally proposed for single carrier TDMA systems [22] is modified to suit OFDMA systems. This adaptive TBFQ (ATBFQ) scheme is investigated through extensive simulations considering WINNER system parameters.

1.4 Scope of the Thesis

The performance of the ATBFQ scheme is studied in the context of the 4G WINNER system and is compared to the Score Based (SB) and Round Robin (RR) schedulers. A

simulation model for the downlink is built adherent to specifications of this system. The simulation model built caters to the following needs:

- A traffic model which realistically models the burstiness of the video streaming service class,
- An inter-cell interference model which takes the interference from the first tier of BS into account,
- A channel model which accurately depicts the large scale path loss, shadowing and fading for a macro-cell urban environment,
- The ATBFQ algorithm for the multi-carrier WINNER system,
- A modified version of the SB algorithm for multicarrier networks (original SB was proposed by for single carrier systems)
- Adaptive coding and modulation (AMC),
- For radio resource allocation, the maximum SINR algorithm.

1.5 Organization of the thesis

The remainder of this thesis is organized as follows. In Chapter 2, the features and characteristics of the original TBFQ algorithm are described. Chapter 3 provides an overview of the WINNER system architecture. It states the various protocol layers especially the MAC layer in detail. The timescale of the scheduling operation along with the resource partitioning is also discussed. Details of the reference SB algorithm are also provided. Orthogonal frequency division multiple access (OFDMA) and its parameters are also described.

The ATBFQ algorithm along with its parameter selection is outlined in chapter 5. Chapter 6 describes the simulation models, parameters, and assumptions. The inter-cell interference model, video streaming traffic model, adaptive coding and the AMC techniques are all explained in detail. The resource scheduler is also described with the ATBFQ algorithm. In Chapter 7, the simulation results are presented. The ATBFQ algorithm is compared with the SB and the performance is shown with respect to various loading levels and different interference

conditions. Finally, conclusions along with outline proposal for future research are provided in Chapter 6.

Chapter 2

Token Bank Fair Queuing Algorithm

2.1 Introduction

The TBFQ algorithm was initially developed for wireless packet scheduling in the downlink channel by William Wong [22] and was later modified for wireless multimedia services using uplink as well. Its concept was based on the leaky bucket (LB) mechanism which polices flows and conforms them to a certain traffic profile. In this chapter, a description of the LB mechanism is provided followed by a detailed description of the TBFQ algorithm along with its characteristics and properties.

2.2 Leaky Bucket

This section explains the basic leaky bucket model, its parameters and usage [23-25]. Although there are different versions of the leaky bucket scheme, they all share the common basic idea of regulating the rate and burstiness of information entry into the network. A leaky bucket controller is comprised of a controller buffer and a token bucket as shown in Figure 2-1. Packets arrive at the buffer and are queued. For a packet in the buffer to leave the controller and be admitted into the network, it must obtain a token from the token bucket. Tokens are generated in the bucket periodically with a specified rate r . The token bucket has a fixed size b . If the token bucket is full at the time of token generation, the newly generated token is discarded. This scheme is specified by two parameters: the token generation rate r and the bucket size b . The token generation rate quantifies the allowed rate of admissions, and the bucket size quantifies the allowed burstiness of the traffic admitted. The leaky bucket scheme has been studied by numerous researchers under many different formulations.

The LB implementation does not efficiently use available network resources because its leak rate is a fixed parameter; there will be many instances when the traffic volume is very low and large portions of network resources (bandwidth in particular) are not being used. Therefore no mechanism exists in the leaky-bucket implementation to allow individual flows to burst up to port speed, effectively consuming network resources at times when there would not be resource contention in the network.

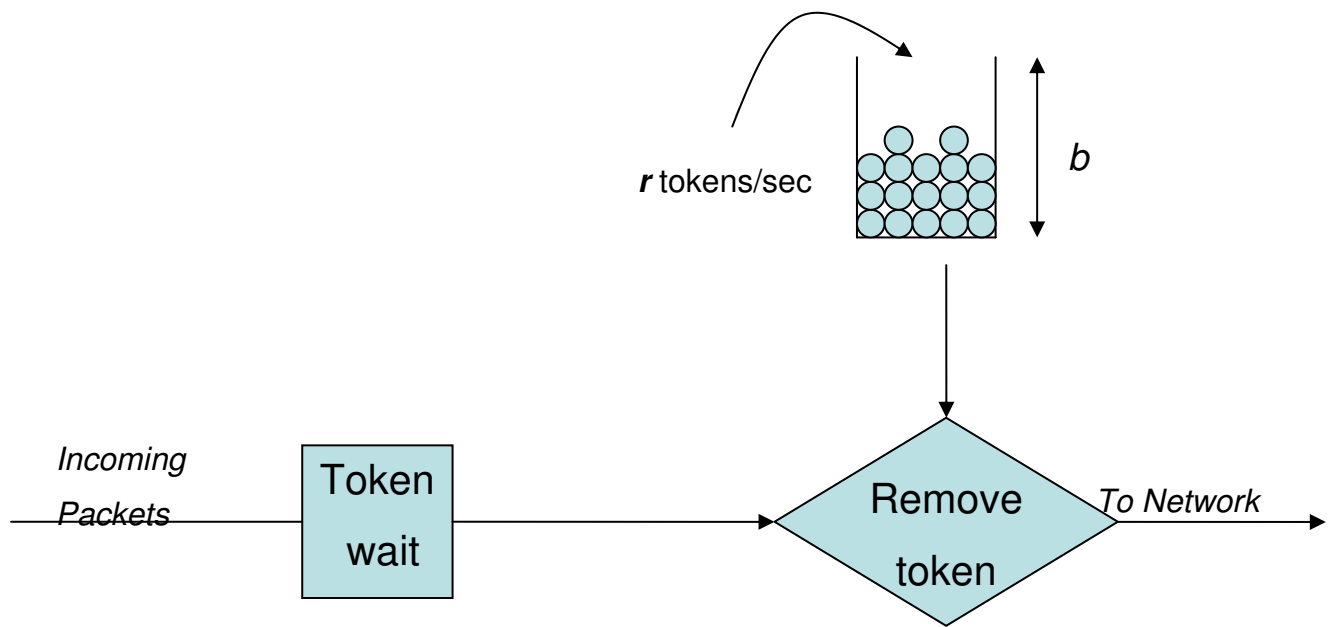


Figure 2-1 Leaky bucket mechanism

2.3 Token Bank Fair Queuing Algorithm

In wireless mobile networks, strict scheduling schemes may not seem appropriate as flows cannot *a priori* determine their exact behavior as most schedulers would require. The environment of wireless scheduling requires soft handling of packets [45] - this is the philosophy behind TBFQ. We define soft QoS provision of a session to be the graceful acceptance of traffic profile violation when excess bandwidth is available, provided the UT does not exceed its bandwidth allocation in the long term. This prevents sudden degradation of quality of service

experienced by the end user as a result of traffic profile violations. TBFQ penalizes violating traffic less severely as it is able to service a packet, which might otherwise be discarded by per flow policing mechanism, by distributing unused bandwidth from other connections.

The generic TBFQ algorithm as proposed in [22] is a frame based algorithm similar to round-robin type algorithms. Each frame can be thought of as a round. It is a work conserving algorithm. A work conserving scheduler is never idle while there are packets waiting to be transmitted in the service queues. On the other hand, a non-work conserving scheduler may be idle even if their packets waiting to be served. In a wireless network it may be better to postpone a mobile terminal from transmitting if the channel condition is poor, this gives the opportunity for other terminals to utilize the bandwidth while their channel condition may be better. The postponing of scheduling service due to impaired channel condition can turn a work conserving algorithm to a non-work conserving one.

The parameters of the algorithm are defined with respect to a packet based system. Each frame in a TBFQ can be considered as a round in a round robin based system except that in each frame the order of service will change according to the priorities assigned to the users. A round is generally defined as having varied time intervals and the length of which depend on the completion of servicing all the flows in turn in the system. Frames are intervals with fixed time period. The structure for TBFQ in the downlink is shown in Figure 2-2 .

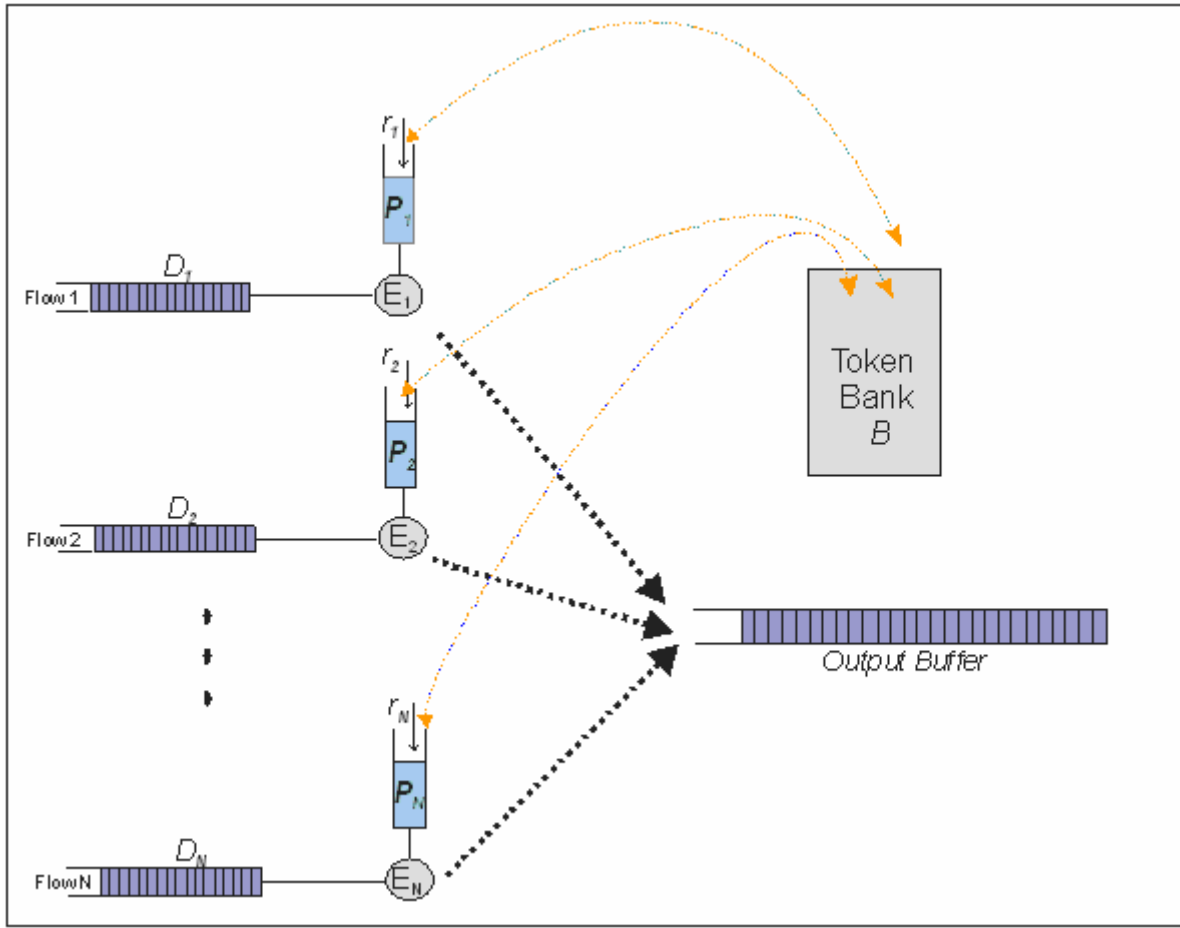


Figure 2-2 TBFQ downlink structure

A flow is defined as stream of packets belonging to one user. Each flow has its own data buffer.

Flow i is characterized by the following LB parameters:

r_i : Token generation rate

D_i : Data buffer size

P_i : Token pool size

Furthermore, each flow is also defined by:

λ_i : Packet arrival rate of flow i

E_i : Counter that keeps track of the number of tokens borrowed from or given to the token bank by flow i

Each L -byte packet consumes L tokens. For each flow i , E_i is a counter that keeps track of the number of tokens borrowed from or given to the token bank. As tokens are generated at rate r_i , the tokens overflowing from the token pool are added to the token bank, and E_i is incremented by the same amount. When the token pool is depleted and there are still packets to be served, tokens are withdrawn from the bank by flow i , and E_i is decreased by the same amount. Thus, during periods when the instantaneous incoming traffic rate of flow i is less than its token generation rate, the token pool always has enough tokens to service arriving packets, and E_i becomes positive and increasing. On the other hand, during periods when the instantaneous incoming traffic rate of flow i is greater than its average token generation rate, the token pool is emptied at a faster rate than it can be refilled with tokens. In this case, the connection may borrow tokens from the bank. The priority of a connection in borrowing tokens from the bank is determined by the priority index given by

$$\frac{E_i}{r_i}. \quad 2-1$$

The connection that has the highest index value has the highest priority in borrowing tokens from the bank; hence it will be serviced first. The amount of tokens a flow can borrow from the bank is vital because it

- I. defines the amount of bursty traffic which can be accommodated,
- II. maintains fairness among all flows such that no one is affected.

For this purpose the following concepts are defined:

Debt limit (d_i): A limit is placed on the amount of tokens a flow can borrow in order to avoid starvation of other flows. If E_i reaches d_i , the connection can no longer borrow from the bank. The debt limit is initially defined as a negative value.

Burst credit (BC_i): The maximum number of tokens connection i can borrow from the bank each time is defined by BC_i . For a constant bit rate source, r_i equals the source peak rate λ_i and there's no need to borrow tokens from or deposit tokens into the bank. E_i ideally would stay zero all the time and c_i would have no relevance whatsoever. However for bursty sources, c_i should be set large enough such that the bursty nature of the traffic is taken into account.

Creditable threshold: A connection may borrow token from the bank until its debt limit is reached, then it must wait until it has deposited enough tokens to the bank.

2.4 TBFQ Functionality

A pseudo-code implementation of the generic TBFQ scheduling is shown in . The operation of the TBFQ algorithm itself can be summarized by the following functions:

- I. *Initialize*
- II. *Enqueue*
- III. *Token Generation*
- IV. *TBFQ*

Before the scheduling operation begins, the *Initialize* function is invoked. This is for the memory allocation of the TBFQ parameters.

The *Enqueue* function is invoked when a packet arrives. This inserts the packet into the proper queue. If the packet belongs to a previously inactive flow, then all the TBFQ parameters have to be initialized using the *Initialize* function.

The *Token Generation* function is invoked whenever a token is generated by flow i . If the token pool for flow i is empty, the new token will fill the pool; otherwise, the overflowing token is added to the bank and the counter E_i is incremented.

The *TBFQ* function is invoked at the beginning of every frame. Depending on the number of available scheduling slots, this function can be split into two rounds. In the first round, if token pool is full for a backlogged flow i , then a packet is admitted to the output buffer. This condition is checked for all backlogged flows. In the second round, priority is calculated for all backlogged flows which do not satisfy the first condition based on the priority index

$$P_i = \frac{E_i}{r_i}. \quad 2-2$$

Using the standard *Quicksort* function, the priority of each flow is sorted in an ascending manner and this result is stored in the *Priority* array in form of flow id's. The first term in the *Priority* array contains the id of the flow with the highest priority index. Resources are then allocated to this flow. The backlogged flow with the highest priority can continue to borrow tokens from the bank until it no longer has the highest priority or the bank gets depleted.

In each frame the total amount of resources that are available to serve all the flows is set to B . Thus the transmission rate during the frame is given by

$$R = \frac{B}{T}, \quad 2-3$$

where T is the period of a frame. After the first round of scheduling where each flow is served based on its token pool size, the remaining resources that can be used in the second round are given by

$$B' = B - \sum_{i \in n} \left\lfloor \frac{p_i}{L} \right\rfloor, \quad 2-4$$

where p_i is the token pool for flow i , L is the size of the packet and $0 \leq p_i \leq L_i$.

Initialize()	<i>//invoked before the scheduler can be used</i> <i>Constants : NUMFLOW, MAXQSIZE, MINTOKENRATE, PACKETSIZE, BANKSIZE</i> <i>ALLOCATE (InputQueue[]); //contains incoming packets</i> <i>ALLOCATE (Rate[]); //token generation rate for each flow</i> <i>ALLOCATE (TokenPool[]); //token pool</i> <i>ALLOCATE (E[]); //counter that keeps track of tokens</i> <i>ALLOCATE (PriorityIndex[]); //keeps track of the priority of each flow</i> <i>ALLOCATE (Priority[]); //prioritized flow order</i> <i>ALLOCATE (Backlogged[]); //FALSE if queue is empty</i> <i>ALLOCATE (DebtLimit[]); //keeps the debt limit of all flows</i> <i>ALLOCATE (BurstCredit[]); //keeps the burst credit of all flows</i> <i>ALLOCATE (CreditableThreshold[]); //keeps the creditable threshold of all flows</i> <i>For (i=0;i<NUMFLOW; i++)</i> <i>Qsize[i] = MAXQSIZE;</i> <i>Rate[i] = MINTOKENRATE;</i> <i>TokenPool[i] = PACKETSIZE;</i> <i>E[i] = 0;</i> <i>Bank = BANKSIZE;</i>
Enqueue(i, j, packet, r, d, c, h)	<i>//packet: the actual packet data</i> <i>//i : flow id number</i> <i>//j : Service class number</i> <i>//r : token generation rate for flow i</i> <i>//d : debt limit for flow i</i> <i>//c : flow credit for flow i</i> <i>//h : creditable threshold for flow i</i> <i>If(InputQueue(j,i) isEmpty) //this is to check if it is a new flow or not</i> <i>Rate[j,i] =r;</i> <i>DebtLimit[j,i] =d;</i> <i>BurstCredit[j,i] =c;</i> <i>CreditableThreshold[j,i] =h;</i> <i>E[i] =0;</i> <i>TokenPool[j,i] =0;</i> <i>packet → InputQueue[j,i];</i> <i>Else packet → InputQueue[j,i];</i>
TokenGenerated(j,i)	<i>//The following function is invoked when a token is generated for Queue (j,i)</i> <i>If(token pool is empty)</i> <i>Then fill the pool with the new token</i> <i>Else E[j,i] +=1;</i>
TBFQ() (1st stage)	<i>For(i=0....n-1)</i> <i>If(Backlogged[j,i] && Tokenpool[j,i] ==PACKETSIZE)then</i> <i>Admit one packet form InputQueue[j,i]to output buffer</i> <i>Tokenpool[j,i] = 0;</i>
(2nd stage)	<i>Quicksort(E[] ,r[], Priority[]);</i> <i>While (Bank>0)</i> <i>For (i=0....n-1)</i> <i>q = Priority[i];</i> <i>If Backlogged[j,q] then Break;</i> <i>Admit one packet from InputQueue[j,i] to output biffer</i> <i>E[q] = E[q] - PACKETSIZE;</i> <i>Bank = Bank – PACKETSIZE;</i> <i>Quicksort(E[] ,r[], Priority[]);</i> <i>End while</i>

Figure 2-3 Pseudo-code of generic TBFQ

2.5 TBFQ Complexity

The complexity of the TBFQ scheduler is defined by three operations: enqueueing, dequeing and the priority calculation function. The enqueueing function is executed whenever a new packet arrives at a flow. Whether the packet belongs to an existing queue or to a new one, it will be identified and appended to the end of the appropriate queue. This operation has a complexity of $O(1)$. The dequeing function is independent from the enqueueing function. Although there is a priority operation to be determined by sorting which is computed once during each scheduling interval, the sorting is not computed on a per packet basis but rather on a per flow basis. Therefore when a packet is dequeued, it is of a complexity of $O(1)$ also. The quicksort function is used to calculate the priorities and the complexity of that on the average is $O(n \log n)$ where n is the number of users [26].

Chapter 3

Overview of the WINNER Architecture

3.1 Introduction

This chapter describes the functionalities within a WINNER Random Access Node (RAN). The architecture is shown in context of the different system layers. A brief description is given of each layer along with an explanation of its main function (those that are relevant to this study).

There are four system layers in the WINNER system concept [27]. These layers are further divided into *user plane* and *control plane*. The services that need to operate on individual Internet Protocol (IP) packets or lower layer Protocol Data Units (PDUs) have been placed in the user plane. The control plane services operate on longer time scales and control the operation of the user plane services by way of control signaling.

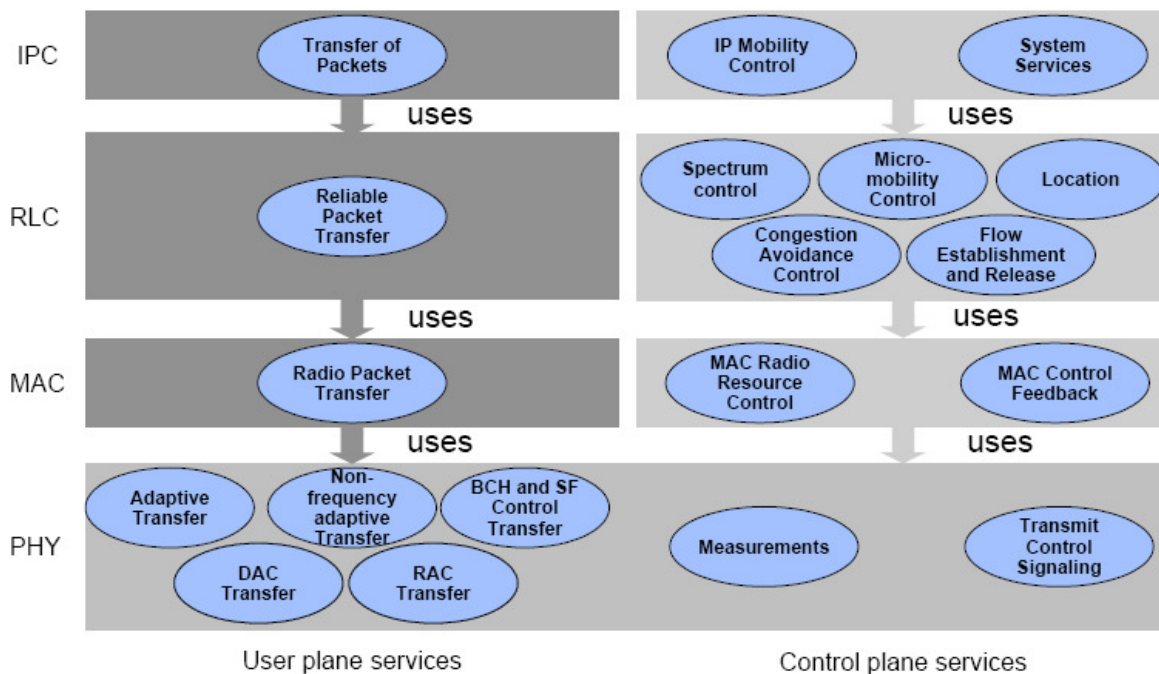


Figure 3-1: The layer and services of WINNER (taken from [27])

The functional role of each system layer is as follows:

3.1.1 Internet Protocol Convergence (IPC) layer

The user plane of the IPC layer receives IP packets from the user of the WINNER RAN, maps them into flows and performs header compression and decompression. The control plane is responsible for RAN association functions as well as macro-mobility (IP level mobility).

3.1.2 Radio link control (RLC) layer

The user plane of the RLC layer provides reliable packet transfer over the radio interface. It also performs confidentiality protection and packet prioritization in order to meet the QoS goals. The control plane takes care of flow establishment and release, location services, load, spectrum, and micro-mobility control.

3.1.3 Medium access control (MAC) layer

The MAC user plane provides the service “radio packet transfer” i.e., transmission and reception of packets over the radio interface. An important part of this service is the scheduling of packets. The control plane provides the “MAC radio resource control” service i.e., acceptance and execution of control messages from higher layers that specify required transmission parameters and boundary conditions.

Furthermore, it implements “MAC control feedback”, i.e., messaging that supports the flow control, the QoS control and the spectrum assignment and other functions at the RLC system layer. There is a tight inter-layer interaction between MAC and physical layers and this is crucial for the performance of the WINNER system. Some functions, such as encoding and decoding, that are traditionally placed in the physical layer are in the WINNER system concept placed in the MAC system layer [28].

3.1.4 Physical (PHY) layer

The PHY system layer handles the physical transmission of flows and of measurements and control signaling directly related to the radio interface. The PHY system layer is not separated into user plane and control plane since it is assumed that all control functionality for the PHY layer resided within the control plane of the MAC system layer.

3.2 Physical Layer Modes

The WINNER architecture should be able to handle deployments from wide area coverage to high capacity hot spots. A basic goal is that the WINNER radio interface should present a unified set of services to higher layers, yet include some specific parts that provide the required flexibility. To provide flexibility and convergence in a structured way, the definition of *modes* is helpful. A *physical layer mode (PLM)* has been defined where there is a significant impact of PHY functionality on the radio interface concept. Two PLMs have been defined:

- I. **Frequency division duplex (FDD):** transmissions are performed over paired bands and supporting half-duplex FDD terminals.
- II. **Time division duplex (TDD):** transmissions are carried out over unpaired band.

Although any PLM can be configured for any kind of deployment, the FDD mode is evaluated primarily in *wide-area cellular deployment scenarios*, using frequency bands of different width. The TDD PLM has so far primarily been evaluated in short-range cellular deployment. A *system mode* represents a specific combination of physical layer modes and MAC modes [27]. All higher layer functions are designed to be mode-independent (generic) and form the unified interface of the WINNER system. There are three MAC modes within the concept:

- **FDD cellular MAC**
- **TDD cellular MAC**

- **MAC for peer-to-peer transmission** designed using the TDD physical layer mode.

The combinations of PHY and MAC modes thus define three WINNER system modes. Parameterizations within modes provide further flexibility and adaptability. Both PLMs use generalized multi-carrier (GMC) transmission, which includes CP-OFDM (cyclic-prefix orthogonal frequency division multiple access) and serial modulation as special cases.

3.3 Chunk, Slot, Frame, and Super-frame Definitions

The basic time-frequency resource unit in OFDM links is denoted as a chunk. It consists of a rectangular time-frequency area that comprises a number of subsequent OFDM symbols and a number of adjacent subcarriers. A chunk contains payload symbols and pilot symbols. It also contain control symbols that are placed within the chunks to minimize feedback delay (in-chunk control signaling). The number of offered payload bits per chunk depends on the utilized modulation-coding schemes, and on the chunk sizes. Each chunk entity comprises n_{sub} subcarriers and spans a time window of n_{syms} OFDM symbols as shown in Figure 3-2a. In transmission using multiple antennas, the time-frequency resource defined by the chunk may be reused by spatial multiplexing. A *chunk layer* represents the spatial dimension (Figure 3-2 b).

In the FDD physical layer modes, chunks comprise 8 subcarriers by 12 OFDM symbols or $312.5 \text{ kHz} \times 345.6 \text{ } \mu\text{s}$. The complete dimensions are shown in Figure 3-3 and appendix A. The chunks are organized into *frames*. In the TDD mode, each frame consists of a downlink transmission interval followed by an uplink transmission interval, denoted *slots*, or *time-slots*. In FDD, the frame is also split into two slots. The frame duration has been set equal in the two PLMs, to facilitate inter-mode cooperation. With frame duration of $691.2 \text{ } \mu\text{s}$, an FDD frame consists of two chunk time durations, with one chunk per slot. Further details on the frame parameters are provided in Appendix A.

The *super-frame* (SF) is a time-frequency unit that contains pre-specified resources for all transport channels; Figure 3-4 illustrates its preliminary design, comprising of a *preamble* followed by n_f frames. Here $n_f = 8$, resulting in super-frames of approximate duration 5.6 ms. The available number of chunks in the frequency dimension could vary with the geographical location. It is assumed that for the FDD downlink (DL) and uplink (UL) as well as for TDD, there exist frequency bands that are available everywhere. The preamble is transmitted in those commonly available bands. The remainder of the super-frame may use other spectral areas that are available at some locations, or to some operators, but not to others. All of these spectral areas are spanned by one Fast Fourier Transform (FFT) at the receiver and are at present assumed to span at most 100 MHz.

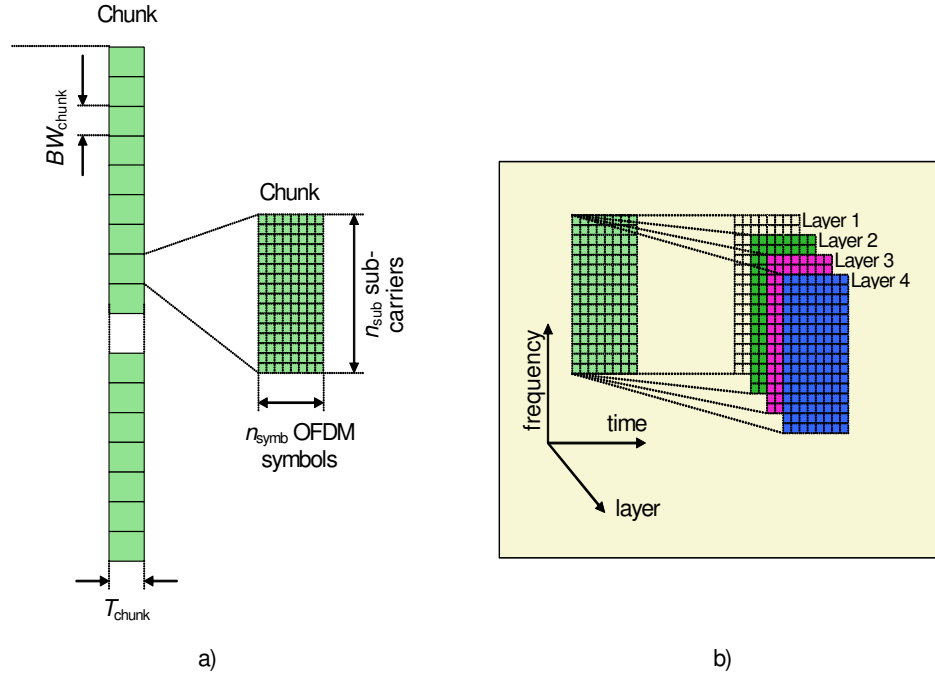


Figure 3-2: a) Multi-carrier downlink physical channel structure. b) Layered time and frequency chunks for multiple input multiple output (MIMO) transmission (taken from [28])

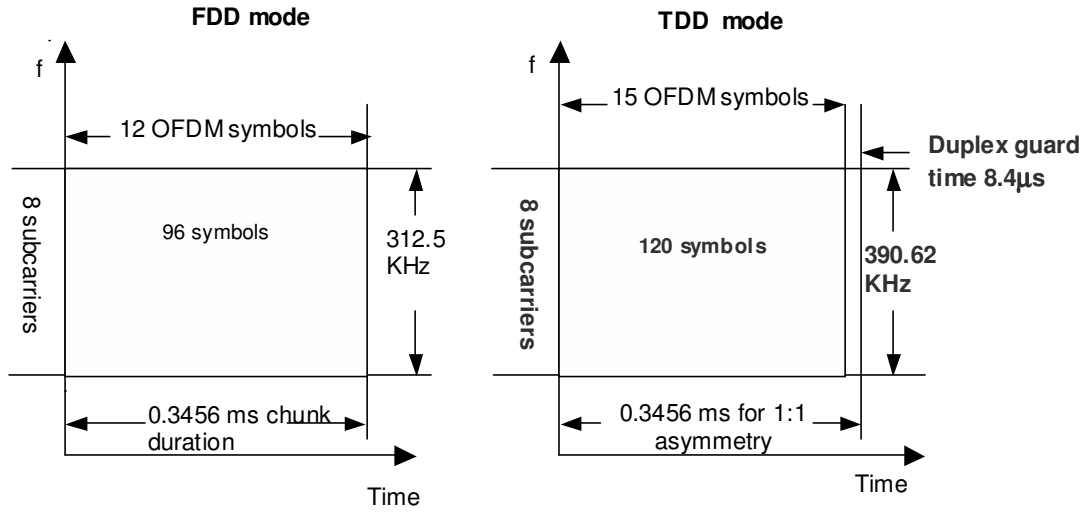


Figure 3-3 Summary of chunk sizes in the two physical layer modes (taken from [28])

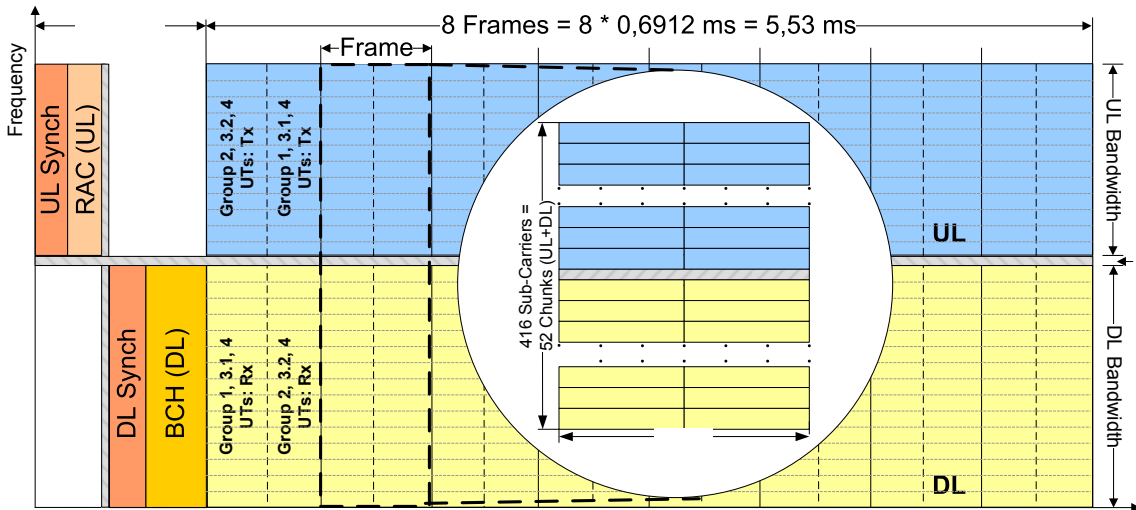


Figure 3-4 WINNER MAC Super Frame structure for the FDD case (taken from [28])

The combination of OFDMA with a TDMA component provides a large amount of multiple access (MA) and resource assignment flexibility with additional complexity required. The granularity of resource assignment is determined by the chunk dimensions (number of subcarriers and OFDM symbols per chunk). Therefore, in frequency-adaptive transmission on

the WINNER downlinks and uplinks, chunk-based TDMA/OFDMA is the primary multiple access scheme of WINNER [28]. The data flows are mapped exclusively onto individual chunks (or chunk layers for MIMO transmission). Further details are provided in Appendix A.

3.4 Resource Scheduling

The resource scheduler determines the resource mapping for the incoming flows at the scheduler. It utilizes two scheduling algorithms:

- I. *Adaptive resource scheduler*
- II. *Non-frequency adaptive resource scheduler*

These algorithms take priorities from the RLC layer into account, as well as the queue levels for each flow.

The *adaptive resource scheduling* and transmission uses predictions of CQI to *utilize the small-scale and frequency-selective variations of the channel* for different terminals. The scheduler assigns a set of chunk layers within the frame to each flow. After scheduling, the resource scheduler buffers (RSB) are drained with bit-level resolution. The bits from each flow are mapped onto the assigned chunk layers. This mapping is exclusive, i.e. several flows do not share a chunk layer. The transmission parameters within each chunk layer are adjusted individually through link adaptation to the frequency-selective channel of the selected user. By selecting the best resources for each flow, multi-user scheduling gains can be realized. The scheduling algorithm should take into account the *channel quality information* of each user in each chunk layer, the RLC flow *transmission requirements/priorities* and the *queue levels*.

Non-frequency adaptive resource scheduling and transmission is instead based on *averaging strategies*. Such transmission schemes are designed to combat and reduce the effect of the variability of the SINR, by interleaving, space-time-frequency coding and diversity combining. Non-frequency adaptive transmission is required when fast channel feedback is unreliable due to for instance a high terminal velocity or a low SINR. The non-frequency

adaptive transmission slowly adapts to the shadow fading, but it averages over the frequency selective (small-scale) fading.

In the baseline assumptions, simple resource scheduling algorithms are used. For frequency-adaptive transmission a proportional-fair scheduling strategy shall be used, such as the score-based scheduler outlined in [29]. A basic Round Robin (RR) scheduler is utilized for non-frequency adaptive transmissions. In both cases a minimum delay of one frame between arrival of a packet in the buffer and its transmission is assumed. Also the CQI information used for the scheduling decision shall be outdated by a minimum of 1 frame.

3.4.1 Score Based (SB) Algorithm

The SB algorithm was proposed by Bonald in [29]. It is a variation of the Proportional fair (PF) algorithm which is the most widely adopted opportunistic scheduling algorithm (patented by Qualcomm Incorporated [42]). A user in the PF is selected in the k^{th} timeslot according to [43, 44]:

$$i^*(t_k) = \arg \max_{1 \leq i \leq N} \left(\frac{r_i(t_k)}{T_i(t_k)} \right), \quad 3-1$$

where $r_i(t_k)$ is the instantaneous rate of user i in time slot k and $T_i(t_k)$ is given by

$$T_i(t_{k+1}) = \begin{cases} \left(1 - \frac{1}{t_c} \right) T_i(t_k) + \frac{1}{t_c} r_i(t_k) & i = i^*(t_k), \\ \left(1 - \frac{1}{t_c} \right) T_i(t_k) & i \neq i^*(t_k), \end{cases} \quad 3-2$$

where t_c [seconds] is a time-window over which T_i is calculated.

In [29], it is shown that the PF scheduler while fair and indeed opportunistic in the ideal case may be unfair and unable to fully exploit multi-user diversity in more realistic cases. For this reason the SB algorithm was proposed. Instead of selecting a user when its transmission rate

is high relative to its own *average throughput*, the SB scheduler selects a user when its transmission rate is high relative to its own *rate statistics* and is given by:

$$i^*(t_k) = \arg \min_{i=1, \dots, N} s_i(t_k), \quad 3-3$$

In the above the score $s_i(t_k)$ of user i at slot k corresponds to the rank of its current transmission rate among the past W values $\{r_i(t_k), r_i(t_{k-1}), \dots, r_i(t_{k-W+1})\}$, where W is the window size. Formally, the score of the selected user i is given by:

$$s_i(t) = 1 + \sum_{l=1}^{W-1} 1_{\{r_i(t_k) < r_i(t_{k-l})\}} + \sum_{l=1}^{W-1} 1_{\{r_i(t_k) = r_i(t_{k-l})\}} X_l, \quad 3-4$$

where $\{X_l\}$ are i.i.d random variables on $\{0,1\}$ with $P_r(X=0) = P_r(X=1) = 0.5$

SB was initially proposed for single carrier systems. It has been adapted to the WINNER for multicarrier systems. It is assumed that each time the SB algorithm is invoked, the SINR for each user for each chunk in the base station coverage area is known at the BS. The SB scheduler schedules the user i in chunk j who has the best score. The score is calculated based on the current rank of the user's SINR, $\gamma_{ij}(0)$ among its past W values of SINR in window $\{\gamma_{ij^*}(-1), \gamma_{ij^*}(-2), \dots, \gamma_{ij^*}(-W-1)\}$, where $\gamma_{ij^*}(-1)$ is the SINR value of user i in the past scheduled chunk j^* and $j^* \in \{all\ chunks\}$. The corresponding score for the user i in chunk j will be given by

$$s_{ij} = 1 + \sum_{l=1}^{W-1} 1_{\{\gamma_{ij}(0) < \gamma_{ij^*}(-l)\}} + \sum_{l=1}^{W-1} 1_{\{\gamma_{ij}(0) = \gamma_{ij^*}(-l)\}} X_l. \quad 3-5$$

In the next section, the rational behind choosing the window size is discussed in detail and by simulation, the optimal window size for the WINNER network is presented.

3.4.2 SB Parameter Selection

As mentioned in the previous section, the performance of the SB algorithm depends on the window size which tracks the past W SINR values on the transmitted chunks for each user. A larger window size corresponds to more fairness as the algorithm can track the variations of the user over a longer period.

If a user has been suffering from bad channel condition in the past, it will have transmitted using chunks with lower SINR. When this deprived user has the opportunity to transmit on chunks with higher SINR (with reference to the past W chunks transmitted), it will receive a higher score thus giving it a priority in utilizing those chunks. Similarly, a user receiving comparatively better channel conditions during the past W values will receive a lower score relative to its own channel statistics.

By simulation (shown in Figure 3-5, Figure 3-6, and Figure 3-7), the optimum window size has been shown, keeping into consideration that larger the window size, higher the complexity. We compare five window sizes of $W = [1, 10, 100, 1000, 3000]$ in terms of throughput, packets dropped and queuing delay. We observe that the optimal window size is that of $W=100$.

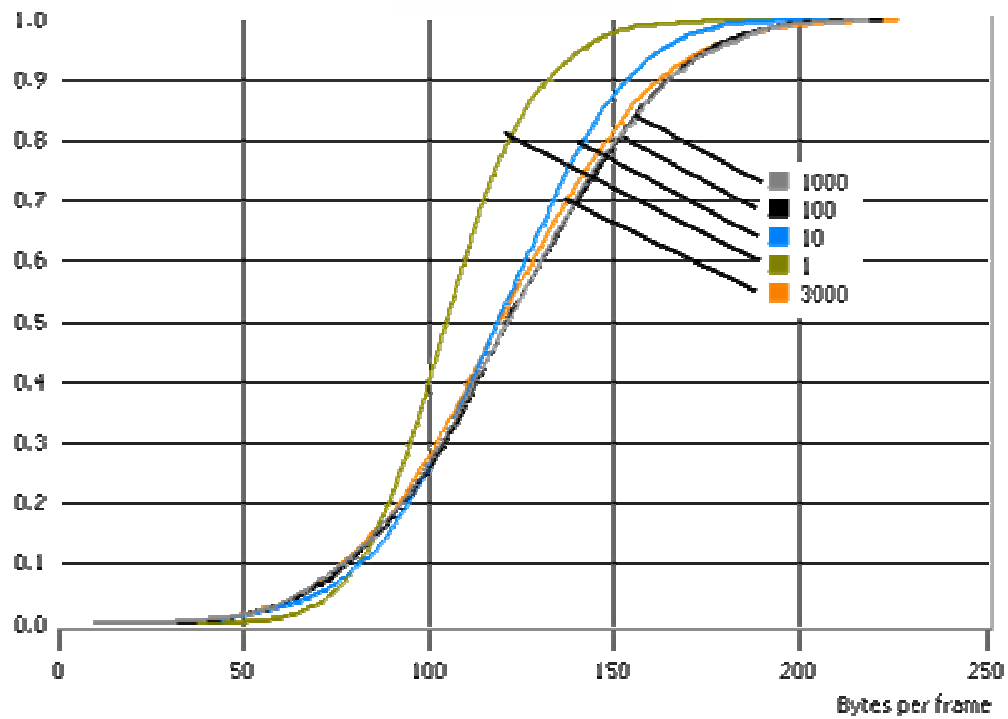


Figure 3-5 CDF of average user throughput for SB scheduler for different window sizes

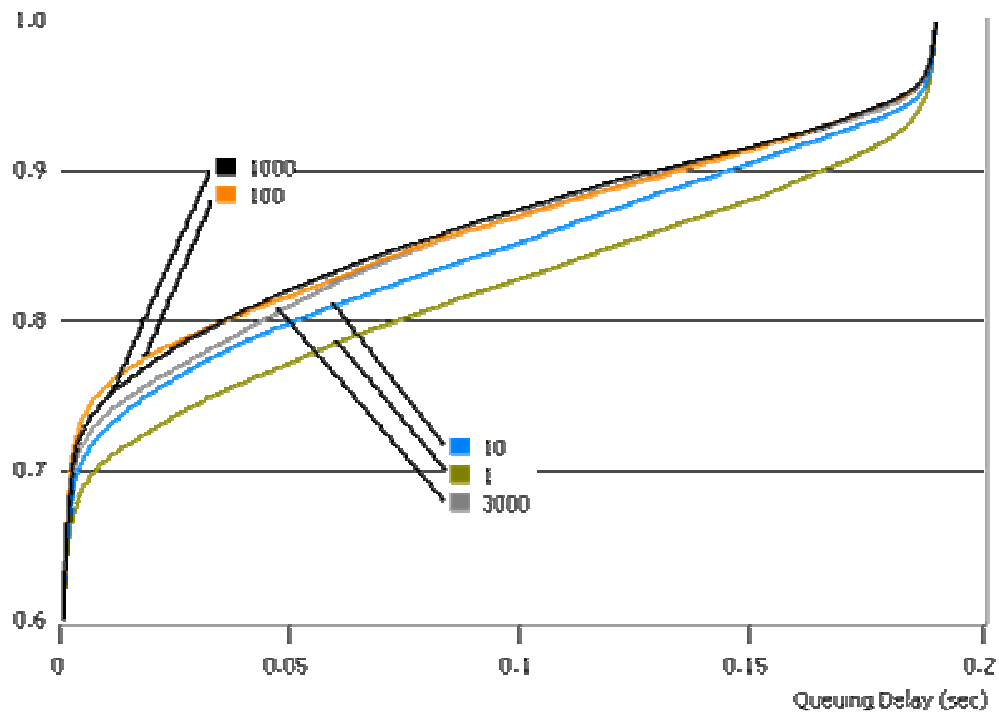


Figure 3-6 CDF of average user queuing delay for SB scheduler for different window sizes

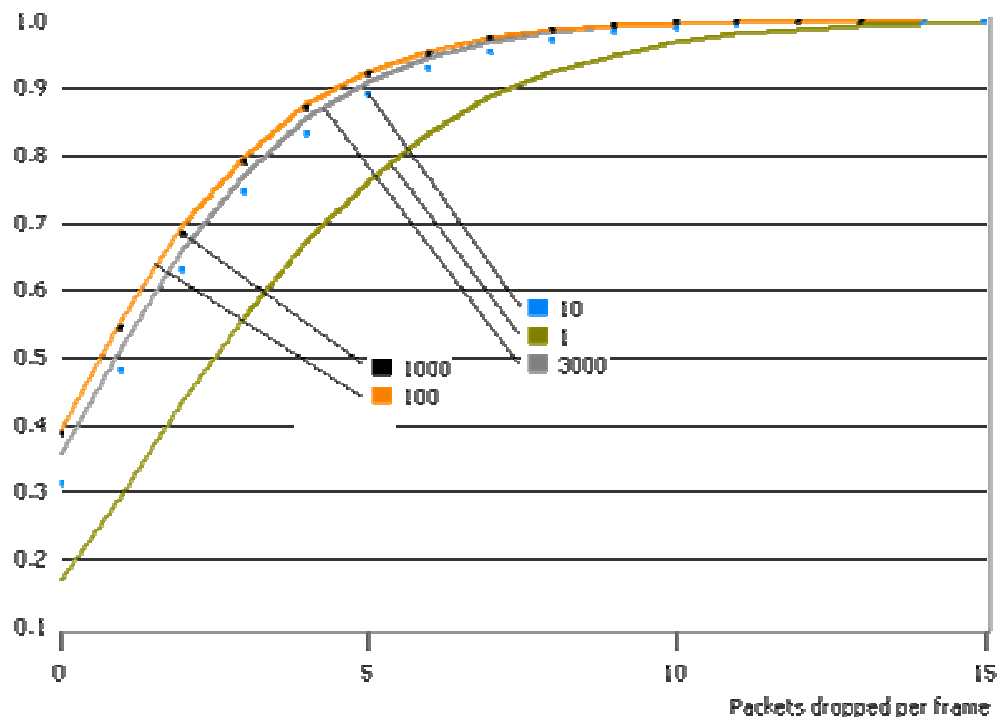


Figure 3-7 CDF of packets dropped per frame for SB scheduler for different window sizes

Chapter 4

Adaptive Token Bank Fair Queuing Algorithm

4.1 Introduction to ATBFQ

An overview of the scheduling operation is shown in Figure 4-1. It highlights how the scheduling operation involving the per-flow queuing (PFQ) of multiple service classes and the scheduling decisions based on feedback, prioritization and other QoS parameters take place. This is shown in the context of the WINNER scheme. As the access method being used is OFDMA, therefore in each scheduling interval, there are many chunks to be scheduled (depending on the number of subcarriers and the OFDMA time symbol length). The ATBFQ is modified to take advantage of this fine granularity of resource units.

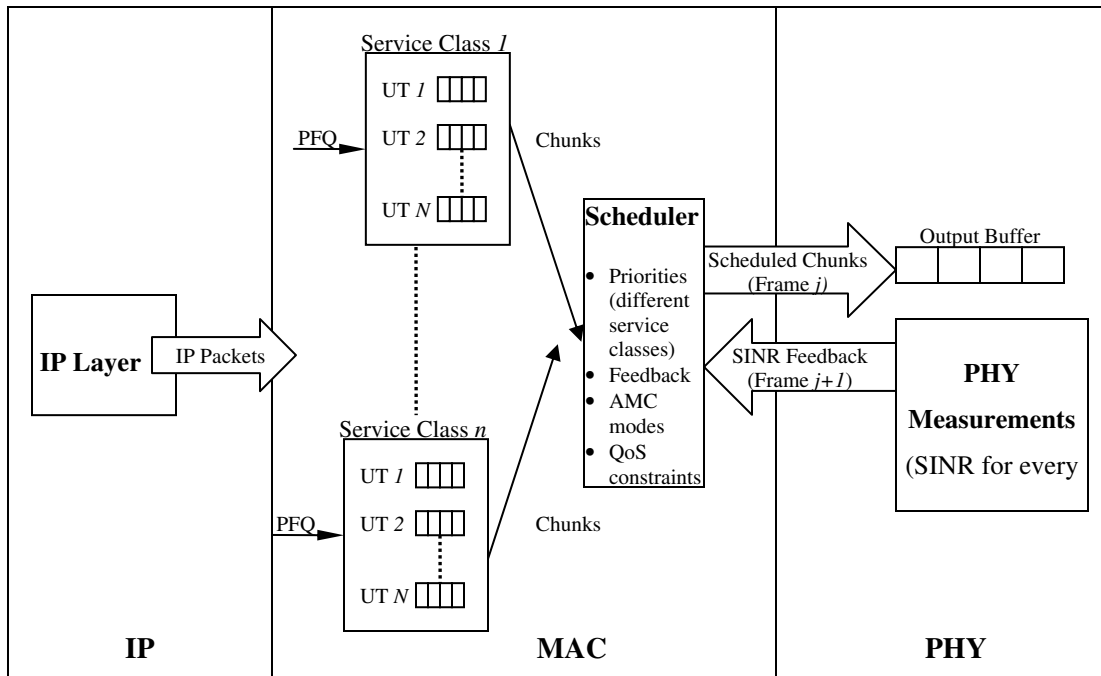


Figure 4-1 Overview of the ATBFQ scheduling operation

Each time a packet is generated, we check to see whether it belongs to an already existing flow. If it belongs to a new flow, then ATBFQ parameters are initialized. Based on the service class type, a debt limit, burst credit, and the creditable threshold are set. The value for these parameters varies from one service class to another. The packets are then queued in subqueues in a manner such that each subqueue belongs to a particular flow. According to the WINNER terminology, this queuing structure is called the service level cache (SLC) cache.

The operation of the scheduler is shown by the following flowchart shown in . This can be summarized by the following functions which are executed each time the scheduler is invoked at the beginning of the frame.

1. At the scheduler, information is retrieved from the higher RLC layer about all the active users using the *getActiveUsers()* function. An active user is defined as a backlogged queue which has packets waiting to be served.
2. Based on this list of active users, a priority is calculated given by the following priority index:

$$P_i = \frac{E_i}{r_i}.$$

The *highestBorrowPriority()* function is called to calculate this for all active users N_{UT} . This function then returns the user i with the highest priority in the k^{th} frame given by:

$$i^*(t_k) = \arg \max_{1 \leq i \leq N_{act}} (P_i). \quad 4-1$$

where N_{act} is the number of active users in the current scheduling frame.

3. Using the *borrowBudget()* function, a certain budget is calculated for the user i based upon the amount of tokens it has contributed to the bank and the debt limit it has incurred from the previous rounds of scheduling.
4. Once the budget is calculated and if it is less than the number of tokens in the bank, resources are allocated to the user i using the *maxSINR()* function. This is the second

level of scheduling and this deals with allocation of chunk resources to the selected user i . This allocation is based on the maximum SINR principle where the chunk j with the best SINR is given to the selected user i^* [41]:

$$j^*(t_k) = \arg \max_{1 \leq j \leq N_{chunks}} (\gamma_{i^*j}(t_k)), \quad 4-2$$

where γ_{ij} is the SINR of the selected user i in chunk j and N_{chunks} is the number of available chunks in the k^{th} frame.

5. The *resourceMap()* function determines the amount of bits that can be mapped to the chunk depending on the type of modulation and coding used.
6. Each time a chunk resource is allocated, the *updateCounter()* function is called. This function updates the bank, the counter E_i and the allocated budget.

The selected user i gets to transmit as long as:

- its queue is backlogged or,
- the allocated budget is less than the total bank size and more than the number of bits that can be supported for the smallest modulation and coding scheme (for this work, this is BPSK rate 1/2).

If either of these conditions is not satisfied, then the user is classified as non-active. A new priority is calculated on the updated active users and steps 1-6 are repeated. This procedure is carried till

- there are no chunk resources available or,
- there are no active users.

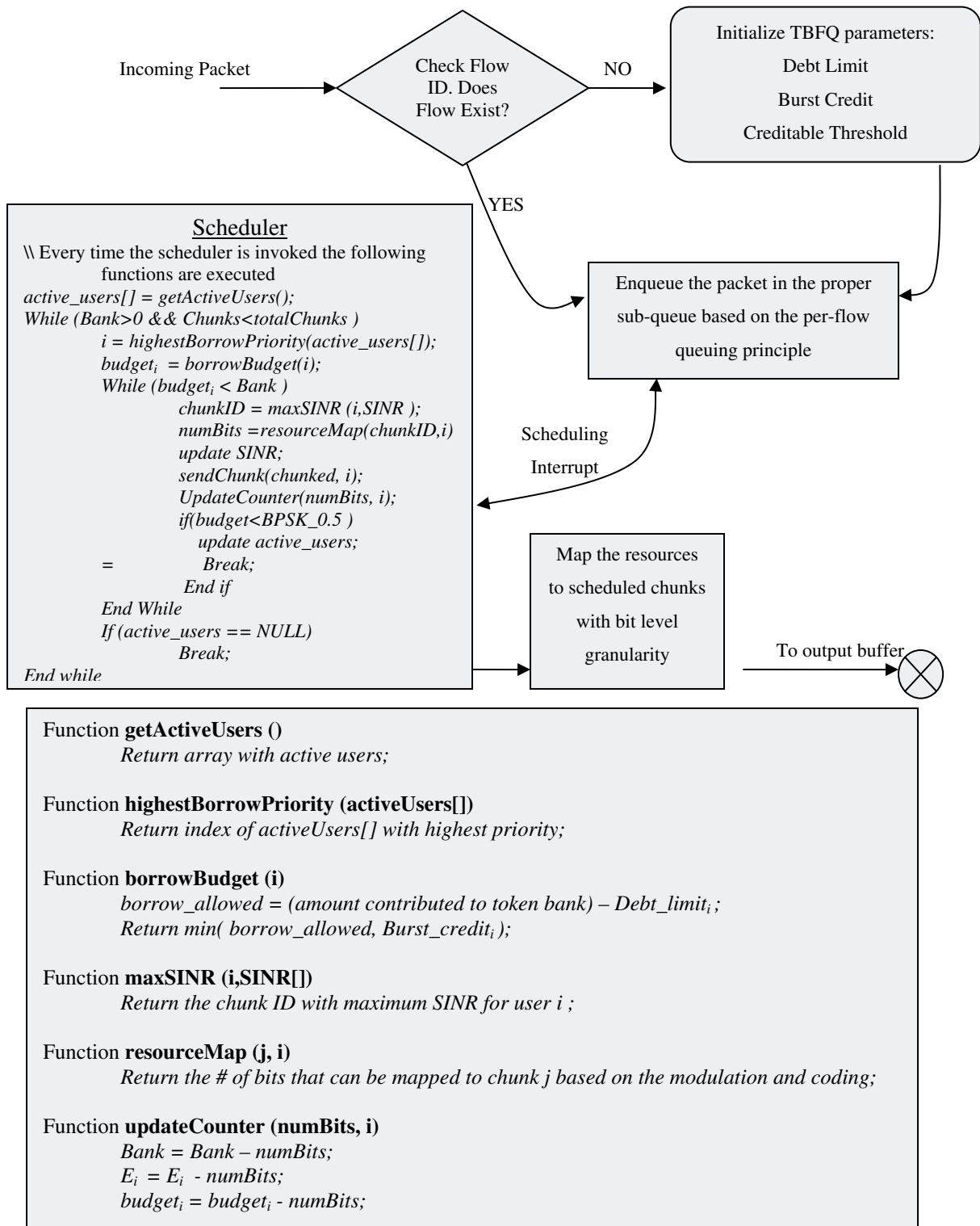


Figure 4-2 Pseudo-code for ATBFQ

4.2 ATBFQ Parameter Selection

The performance of the ATBFQ scheduler depends on its parameters that define the debt limit, the burst credit, and the token generation rate. The token generation rate is detrimental to the extent to which the burstiness in the user traffic can be accommodated. A user in its burst mode transmits more data in a short interval of time than its actual statistics and hence requires more resources in order to maintain a certain QoS level. Thus the token generation rate is set to be considerably large. For simulations shown here, this has been taken to be three times the packet arrival rate. The debt limit ensures the extent to which a user can borrow from the bank. It also acts as a measure to prevent malicious users (users transmitting at higher transmission rates) from borrowing extensively.

The burst credit (BC_i) determines the amount of bits priority user i can receive in a frame. In the generic TBFQ algorithm, this quantity was a fixed measure for the duration of the simulation. In this thesis, this has been modified to be adaptive. Through simulation it was observed that for low loading cases, a higher value for BC_i is optimal as shown in Table 4-1. On the other hand, for high loading conditions, a lower value for BC_i is sufficient as it makes use of multiuser diversity as shown in Table 4-2. It is further shown that this can be achieved for both the low and high loading conditions by calculating the BC_i for the priority user in an adaptive manner. This adaptive value depends on the past spectral efficiency, the number of available chunks, the number of OFDM symbols in each chunk and the number of active users in that particular frame. It can be formulated as follows:

$$BC_i(\text{bits}) = \frac{(\text{user efficiency}_i \text{ in bpsec/Hz}) * (\text{chunk resource in Hz-sec}) * (\text{available chunks})}{\text{active users in the current frame}}$$

4-3

<i>Burst Credit (BC)</i>	<i>Queuing Delay (sec)</i>	<i>Packets Dropped (per frame)</i>	<i>Throughput (Byte per frame)</i>	<i>Spectral Efficiency (bits/sec/Hz)</i>
BC = 1000	0.055044	4.359845	815.347619	2.3744
BC = 5000	0.017409	0.761943	1473.325065	2.046488
BC = 10000	0.015329	0.415673	1546.56654	1.97897
<i>Adaptive BC</i>	0.011811	0.301975	1551.041998	2.336033

Table 4-1 Burst credit for ATBFQ for low loading (8 users)

<i>Burst Credit (BC)</i>	<i>Queuing Delay (sec)</i>	<i>Packets Dropped (per frame)</i>	<i>Throughput (Byte per frame)</i>	<i>Spectral Efficiency (bits/sec/Hz)</i>
BC = 1000	0.043819	3.192160	2299.372600	2.084589
BC = 5000	0.035570	3.983687	2094.002875	1.875760
BC = 10000	0.035521	4.002754	2090.401007	1.872067
<i>Adaptive BC</i>	0.037458	2.005188	2497.095882	2.298360

Table 4-2 Burst credit for ATBFQ for high loading (20 users)

Chapter 5

System Level Simulation Model

5.1 Introduction

When analyzing or assessing the performance of a radio network it is typically not sufficient to study the performance of a single radio link. Instead, one would like to assess the overall network performance, accounting for that several pairs of communicating nodes must share a common radio resource. For instance, in a cellular network it must be considered that the resources in a cell are shared among all user terminals associated with the cell and it is further of great importance to account for interference from neighbouring cells. Multicell evaluations of cellular networks are often performed by computer simulations, referred to as system-level simulations.

System-level simulations include, among others, deployment models, user behaviour models, and channel models. They further comprise models of the network functionality and the radio network algorithms. The deployment models consider parameters such as the base station density, the base station equipment, and the base station antenna positions (below or above rooftops, etc.). User behaviour models include models for the user position, the user mobility, and the traffic generated by the user (in uplink and downlink). Channel models model the radio channel of desired and interfering links.

Furthermore, in a system level simulation, typically there is no explicit modelling of physical layer procedures such as modulation and coding. Instead, less complex link performance models are used to estimate the performance of single links. Such a link performance model is often referred to as a link-to-system interface. The link-to-system interface needs as input a measure of the radio link quality and delivers as output an estimate of the packet error probability. Often, the SINR is used as a measure of the radio link quality, which means that system level simulations must include the calculation of the received SINR. Here the *actual*

value interfaces, where the SINR depends on the fast fading realizations of the channel [30], and *average value* interfaces, where channel quality metrics are an average of the instantaneous SINRs over the fast fading can be distinguished from each other.

This chapter describes different methods and models that may be used in multi-cell system level simulations. The channel model is discussed in view of the WINNER requirements. It discusses the central cell option to set-up a multi-cell system simulation that accounts for interference from neighbouring cells. A method describing the link to system interface and how to achieve the corresponding adaptive modulation and coding curves is described. The traffic model and the scheduling models are also discussed in detail.

5.2 Channel Model

As transmit signal propagates from transmitter to receiver the signal strength weakens. Propagation or channel models characterize the varying nature of wireless channel which has been one of the most challenging tasks of designing a radio system. Channel models are used to predict the average received signal strength for a given transmitter-receiver (T-R) separation which is called *large-scale path-loss* model, and as well as the fluctuations of the signal strength around the average for a particular location which is termed as *small-scale fading* or *fading* model.

Large-scale path-loss depends on the distance between the transmitter and receiver as well as the operating frequency and is therefore modeled in a deterministic fashion for a given T-R distance and frequency. But, in reality, this loss is not constant, and the variations depend on the objects surrounding the receiver and the terrain of the transmission path. This location depended variation of large-scale path-loss is known as *shadowing*. Therefore, a wireless channel is characterized by three different attenuating effects *large-scale path-loss*, *fading* and *shadowing*.

The WINNER system is being developed to cater to different types of propagation scenarios. The main difference between them exists due to their diverse environments. Channel model parameters were defined for seven propagation scenarios, namely indoor small office (A1), urban micro-cell (B1), indoor (B3), stationary feeder (B5), suburban (C1), urban macro-cell (C2), and rural macro-cell (D1). These models are described in [31]. The work done in this study is based on the urban-macro-cell (C2) model.

In the following sections, the C2 channel model is described based on the *large-scale path-loss*, *fading* and *shadowing* models.

5.2.1 Large Scale Path Loss Model

Large-scale path-loss models are empirical models established from extensive field measurements in different terrain conditions such as urban, suburban and rural. The general expression for path-loss for a T-R separation d is given by [34],

$$PL[dB] = PL_{fs} + 10n \log_{10}\left(\frac{d}{d_o}\right), \quad 5-1$$

where large-scale path-loss PL is expressed in dB, n is propagation exponent, and PL_{fs} is free space path-loss at reference distance d_o . PL_{fs} is dependent on operating wavelength λ and is given by,

$$PL_{fs} = 20 \log_{10}\left(\frac{4\pi d_o}{\lambda}\right). \quad 5-2$$

In this study we have used suburban model (as in [32]) presented for fixed broadband wireless access networks. Since the simulation network is implemented in 3.7 GHz operating region and the roof-top antenna height is considered 1.5 meters, the path-loss is adjusted [33]. The resulting large-scale path-loss including correction factors is given by

$$PL = 38.4 + 35.0 \log_{10}(d) [dB],$$

with $50m < d < 5km$.

The large-scale path-loss is given by free space path-loss when the T-R separation is less than the reference distance. In this thesis, a terrain with moderate to low density trees is chosen so that with sufficiently high BS antenna, the resulting propagation constant n becomes 3.5.

5.2.2 Shadowing

Measurements established that the large-scale path-loss such as given in (4.3) for a particular T-R separation is random and distributed log-normally around the mean value described by the path-loss formula. Therefore, path-loss including the shadowing would be,

$$PL[dB] = \begin{cases} 38.4 + 35.0 \log_{10}(d) + X_{\sigma} & d \geq d_o, \\ 20 \log_{10}(\frac{4\pi d_o}{\lambda}) + X_{\sigma} & d < d_o. \end{cases} \quad 5-3$$

In the above, X_{σ} is a Gaussian random variable with a mean of 0 and a standard deviation of σ dB. In this study we have considered independent lognormal random variables with a standard deviation of 8 dB for shadowing [33].

5.2.3 Fading

Fading is the variation of the received signal resulted from the constructive or destructive addition of multiple versions of the transmitted signal each having followed a different transmission path. This fluctuation is experienced over a short period of time and is, therefore, denoted as *small-scale fading* or *fading*. Fading is dependent on the speed of the receiver in case of mobile terminal. In *mobile wireless* systems, fading mainly depends on the speed of the users as well as the movements of the objects surrounding the receiver. Doppler frequency is a measure of the rate of changes in fading. Time-correlated flat Rayleigh fading with Doppler frequency (f_m) of 255.433 Hz has been considered in this study,

$$f_m = \frac{vf_c \cos \theta}{c} = \frac{19.4 \times 3.95 \times 10^9}{3 \times 10^8} = 255.433 \text{ Hz}.$$

5-4

Time and frequency correlated Rayleigh channel samples obtained from power delay profile for WINNER wide area scenario are used to generate the channel fading dataset. The plot of the instantaneous fading power (in dB) is illustrated in Figure 5-1. It should be mentioned here that the mean value of the fading power is unity in the linear scale.

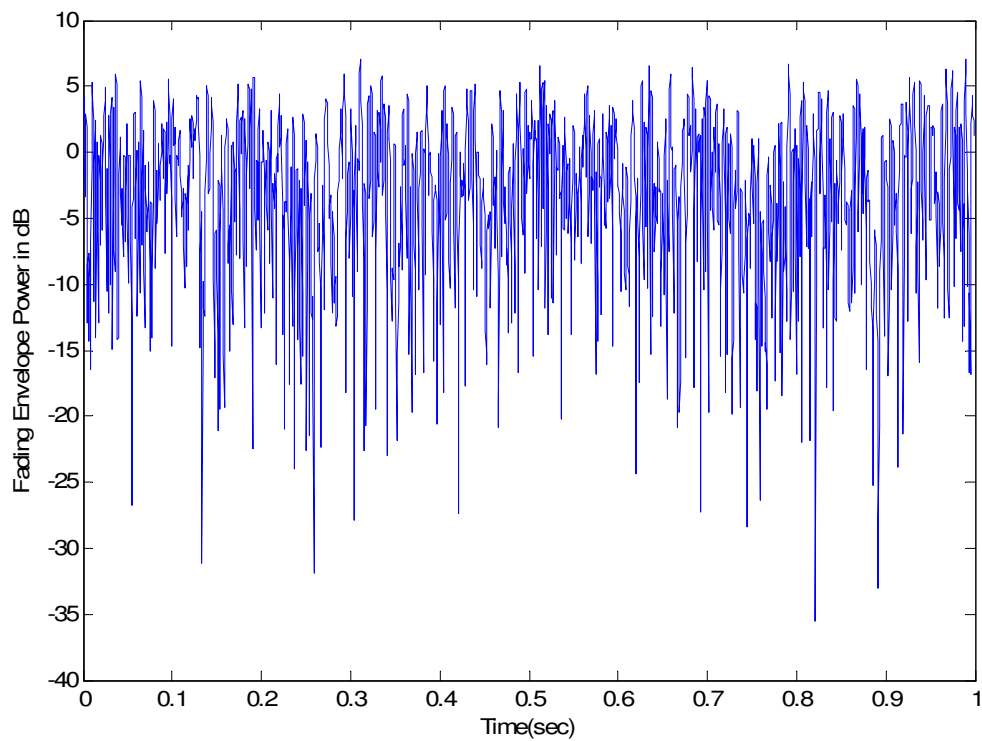


Figure 5-1 Instantaneous power of the fading envelope (shown for 1 sec)

Coherence time T_c is a statistical measure of the time duration over which the received signals have strong amplitude correlation and it is the time domain dual of the Doppler spread. A rule of thumb expression for time coherence can be expressed by the following relation [34],

$$T_c = \frac{0.423}{f_m} = 1.656 \text{ msec}.$$

Using maximum Doppler frequency of 255.433 Hz, T_c would be 1.656 msec. Figure 5-2 shows the fade duration of the fading dataset that are used in this study. It is apparent that the fade durations are larger than T_c and therefore the samples are correlated over the coherence time.

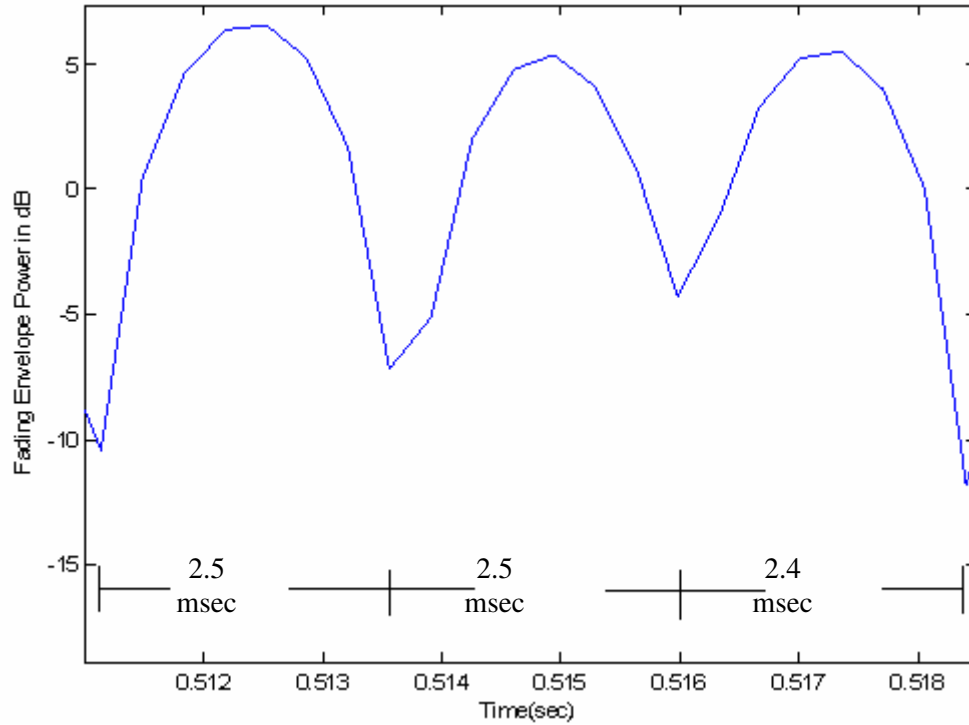


Figure 5-2 Time correlation of the fading samples (zoomed version of Figure 4-1)

5.3 Background Noise Model

Thermal noise is the source of background noise at the receiver. Average thermal noise power is related to the Boltzman's constant C , ambient temperature T in degree Kelvin and channel bandwidth B in Hz as shown below:

$$P_N = 10 \log_{10}(CT_k B) + NF(dB). \quad 5-5$$

With a noise figure (NF) of 7 dB the average noise power is – 144 dBW. This is an additive white Gaussian noise (AWGN) that must be considered while calculating received signal's SINR.

5.4 Adaptive Modulation and Coding

In the baseline design for WINNER, a rate-compatible punctured block low-density parity check code (BLDPCC) of mother code rate 1/2 is used for the transmission of information data. Code rates of 2/3 and 3/4 are obtained by puncturing, and combined with different modulation alphabets– Binary Phase Shift Keying (BPSK), Quadrature Phase Shift Keying (QPSK), 16-bit Quadrature Amplitude Modulation (16-QAM), and 64-QAM. When plotting the throughput versus SINR for these combinations of modulation and code rate, some of the combinations become obsolete, since they don't contribute to the hull curve. Thus a baseline modulation and coding scheme for AMC consists of the following combinations:

AMC	1	2	3	4	5	6	7	8	9	10
Mod.	BPSK		QPSK			16-QAM			64-QAM	
r_c	1/2	2/3	1/2	2/3	3/4	1/2	2/3	3/4	2/3	3/4

Table 5-1 Baseline modulation and coding scheme for adaptive modulation and coding

The corresponding hull curves for the B-LDPCC (FEC block size of 1728 bits) and using 10% Block Error Rate (BLER) as switching criterion are shown for the FDD case in Figure 5-3. This is based on the initial transmissions, i.e. retransmissions are not included. Overhead includes in-chunk pilots and control symbols, as well as the super-frame pre-amble. Table 5-2 illustrates the various chunk sizes with the corresponding modulation and coding rate combinations. A detailed description for the derivation of the LS interface is provided in the Appendix B.

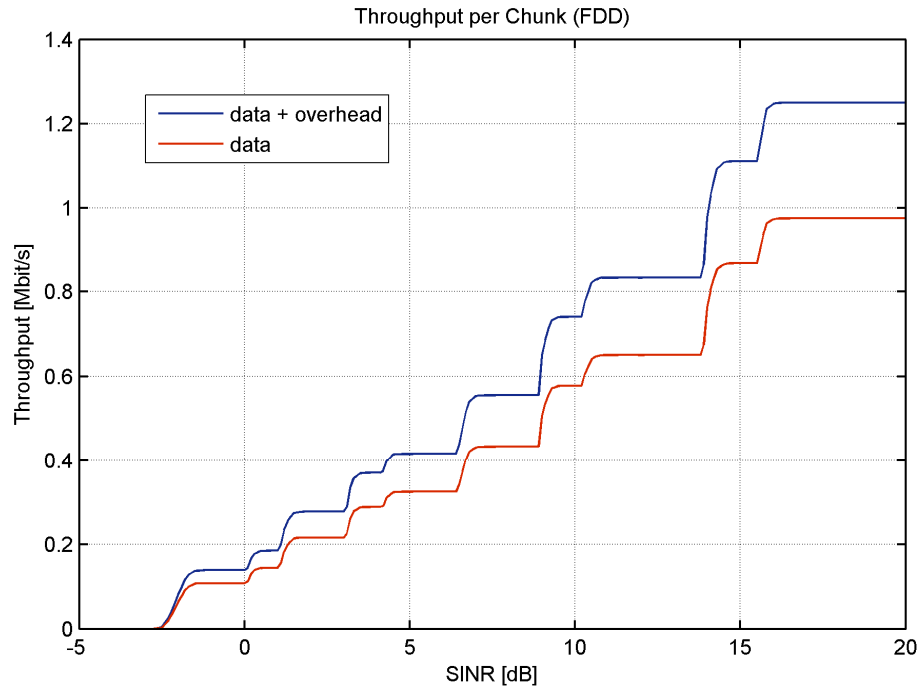


Figure 5-3 Throughput per chunk versus SINR for the baseline AMC shown for the hull curves

AMC Scheme	SINR (dB)	Chunk Size in bits
BPSK 1/2	$0.2311 \geq \text{SINR} > -1.7$	48
BPSK 2/3	$1.231 \geq \text{SINR} > 0.2311$	72
QPSK 1/2	$3.245 \geq \text{SINR} > 1.231$	96
QPSK 2/3	$4.242 \geq \text{SINR} > 3.245$	128
QPSK 3/4	$6.686 \geq \text{SINR} > 4.242$	144
16QAM 1/2	$9.079 \geq \text{SINR} > 6.686$	192
16QAM 2/3	$10.33 \geq \text{SINR} > 9.079$	256
16QAM 3/4	$14.08 \geq \text{SINR} > 10.33$	288
64QAM 2/3	$15.6 \geq \text{SINR} > 14.08$	384
64QAM 3/4	$\text{SINR} > 15.6$	432

Table 5-2 AMC mode for information block-size of 1728 bits

5.5 Interference Model

In practice, when analyzing radio networks by means of computer simulation, it is only possible to model a limited number of cells. That is, the considered system layout is limited and the cells in the outer parts of the modelled network are not surrounded by cells on all sides. The interference situation in these cells differs significantly from the interference experienced in the central part of the network where cells are surrounded by interfering cells on all sides. The performance of such a network is hence not representative of a large real-world network in which basically all cells experience interference from all directions.

There exists two popular techniques to account for the impact of surrounding cells and inter-cell interference, namely the central-cell technique and the wrap-around technique. For this study the central cell technique is used. Both these two techniques are described in the next section.

5.5.1 *Wrap-around Model*

The use of wrap-around, described such as the one described in [36] is one way to overcome the limitation of a limited system. With wrap-around the cell layout is folded such that cells on the right side of the network are connected with cells on the left side and, similarly, cells in the upper part of the network get connected to cells in the lower part. The created area may be seen as borderless, but with a finite surface, and it may be visualized as a torus [36].

5.5.2 *Central Cell Model*

In the central cell technique, the system statistics are collected only in the central part of the simulated multi-cell layout. This is typically the central cell in an omni-directional layout, or the three sectors of the central cell in a tri-sectored layout (as considered in this study). Here, only the system functions of the central cell and the associated terminals are simulated in detail, while the remaining cells are accounted for via simplified models.

When compared to a multi-cell simulation where UTs are generated in multiple cells, e.g. using the wrap-around technique, the central cell technique allows considerable savings in computation resources, as only a reduced number of links needs to be monitored and managed at the same time. From the simulation duration perspective, however, the central-cell technique may be equivalent to a multi-cell simulation using the wrap-around technique. In effect, many independent snapshots of UT locations are needed in the central cell in order to gather results from enough UTs to obtain reliable statistics, whereas a reduced set of snapshots may be sufficient with the wrap-around technique due to the high number of UTs monitored simultaneously in all the simulated cells.

The central cell technique is particularly suited to the downlink, as it avoids the explicit simulation of the UTs in the neighbouring cells. Indeed, the interference situation can be generated accurately by simulating the transmitted signal from the neighbouring base stations only.

In the downlink case, the central cell technique requires models in order to provide an accurate and realistic behaviour of the signals transmitted by interfering base stations. While some simplified models are common to the central cell and the wrap-around techniques, such as those required to reproduce the frequency selectivity and spatial properties of the interferers' channels, additional specific models are needed for the central cell technique due to the absence of an explicit simulation of the UT behaviour. These models essentially concern the following aspects, which are addressed in more details in the next two sections

- The base station activity variation in time and frequency, due to traffic models and scheduling and
- The directional properties of the interfering signal.

The cellular layout of the assumed model is shown in Figure 5-4. Each cell in the network has three sectors. As described above, we only consider the effect of interference on the sector of interest in the central cell (denoted as sector 1 in BS=1). For this purpose the

interference from the first tier will be taken into account. We also assume a frequency reuse factor of one in each sector. A mathematical model for the interference is shown in Appendix D.

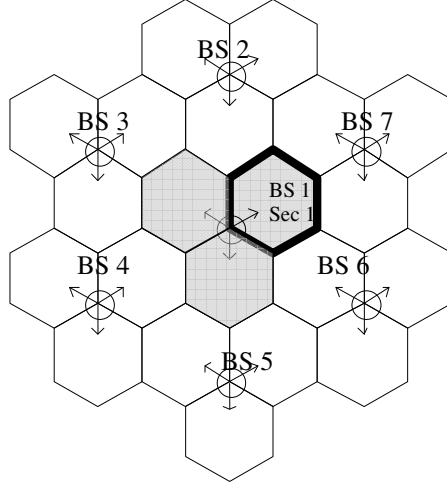


Figure 5-4 Network layout under study

The SINR obtained for chunk j of user i at the output of any linear detector can be expressed under the following generic expression

$$SINR_{i,j} = \frac{P_{\text{signal } i,j}^{1,1}}{(P_{\text{inter } i,j} + P_{\text{intra } i,j}) + P_{\text{noise } i,j}}$$

where $P_{\text{signal } i,j}^{1,1}$ denotes the desired signal power in sector 1 in BS 1, and $P_{\text{noise},i}$ the noise power.

For the given layout in Figure 5-4, $P_{\text{intra},i,j}$, the power of intra-cell interference, and $P_{\text{inter},i,j}$ the power of inter-cell interference are given by the following expressions,

$$P_{\text{intra } i,j} = \sum_{s=2}^3 I_j^{b=1,s} X_I \quad \text{and ,}$$

$$P_{\text{inter } i,j} = \sum_{b=2}^7 \sum_{s=1}^3 I_j^{b,s} X_I,$$

where $I_j^{b,s}$ is the interference power for chunk j from sector s in BS b . X_I has a binary value defined by

$$X_I = \begin{cases} 1 & x \leq AF \\ 0 & x > AF \end{cases}$$

where x is a uniform random variable defined over $[0,1]$, and AF (activity factor) is defined as a probability for a particular interfering link to be active. For example, AF of 1 denotes a high level of interference where all the links are being interfered at (100% interference). In this case, for a link of interest in sector 1 in BS 1, the interference will comprise of 18 (6 BS x 3 sectors) inter-cell and 2 intra-cell links.

5.6 Traffic Model

Traffic models control the generation of data packets in system level simulations. Depending on which service or application that is used the traffic patterns look differently. Examples of applications are e.g. voice, web surfing, video streaming, and file download. Given the used service type, packets are generated with certain (statistical) characteristics, including packet size and packet inter-arrival time. Session times may, in a similar way, be statistically modelled

Real-time video streaming traffic is used for the purpose of this study. Two *Interrupted Renewal Process* (2IRP) sources are superimposed to model user's video traffic in the downlink transmission as indicated in [37]. In each of these sources, modeled traffic is bursty in nature as both have separate ON and OFF distributions. During the OFF state the IRP process does not generate packets. During the ON state, packets are generated with exponentially distributed inter-arrival time. The mean dwell time or sojourn time in ON or OFF state is Pareto distributed. The model is shown in Figure 5-6 . The IRP process is derived from *Interrupted Poisson Process* (IPP). The difference between IPP and IRP is that while the ON and OFF time for IPP is exponential distributed, they are Pareto in IRP.

To find out the generated data rate of the IRP process, we begin with a state diagram of a basic IRP process as shown in Figure 5-5. Let us assume that $\mu_{on \rightarrow off}$ is the state transition rate from ON to OFF and $\mu_{off \rightarrow on}$ is the rate from OFF to ON. The mean dwell time on the ON state is given by

$$T_{ON} = \frac{1}{\mu_{on \rightarrow off}} \quad 5-6$$

while the mean dwell time on the OFF state is

$$T_{OFF} = \frac{1}{\mu_{off \rightarrow on}} \quad 5-7$$

The probability being ON state would be

$$P_{r_{ON}} = \frac{T_{ON}}{T_{OFF} + T_{ON}} = \frac{\mu_{off \rightarrow on}}{\mu_{on \rightarrow off} + \mu_{off \rightarrow on}} \quad 5-8$$

And the probability being OFF state would

$$P_{r_{OFF}} = \frac{T_{OFF}}{T_{OFF} + T_{ON}} = \frac{\mu_{on \rightarrow off}}{\mu_{on \rightarrow off} + \mu_{off \rightarrow on}} \quad 5-9$$

If, during the ON state process, the source generates Λ packets per second on average, then the mean number of generated packets will be $\Lambda \times P_{r_{ON}}$ packets per second. The Pareto distribution parameters for OFF and ON distributions are given in Table 3.1. If a random variable X is Pareto distributed, the probability density function of X is given by,

$$f_X(x) = \frac{\beta \rho^\beta}{x^{\beta+1}} \quad 5-10$$

where $\rho \leq x < \infty$.

The parameters ρ and β are called location and shape parameters, respectively. The location parameter $\rho = 1$ is considered in this model. The shape parameter β is related to the mean value as $E[x] = \beta (\beta - 1)$ for $\beta > 1$. From the parameters in , it can be found that for IRP1, the mean dwell time on ON and OFF states are

$$T_{ON} = \frac{\beta_{11}}{\beta_{11} - 1} = 8.143 \text{ sec} \quad 5-11$$

and

$$T_{OFF} = \frac{\beta_{21}}{\beta_{21} - 1} = 5.545 \text{ sec} \quad 5-12$$

Therefore, for IRP1 the probability being ON state,

$$P_{ON} = \frac{T_{ON}}{T_{OFF} + T_{ON}} = \frac{8.143}{8.143 + 5.545} = 0.595 \quad 5-13$$

Similarly, for IRP2 the probability being ON state is given by 0.384. With these probabilities and packet arrival (exponential) rates shown in the table, IRP1 and IRP2 generate

$$\Lambda_1 \times P_{ON} = 0.595 \times 1123.8 = 668.49 \text{ packets/sec}$$

$$\Lambda_2 \times P_{ON} = 0.384 \times 1547.5 = 594.51 \text{ packets/sec}$$

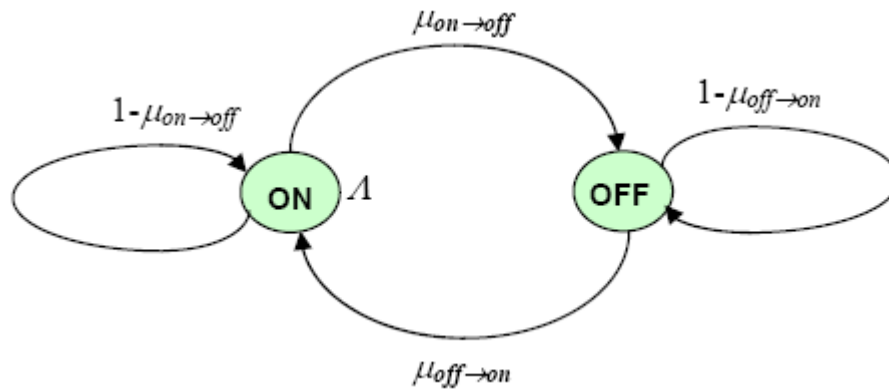


Figure 5-5 State transition diagram of a 2IRP Process

Source_i	Λ_i (pkts/sec)	β_{1i}	β_{2i}	Average pkts/sec
IRP#1	1123.80	1.14	1.22	668.49
IRP#2	1547.50	1.54	1.28	594.51
2IRP Average				1263.00

Table 5-3 Traffic Model Parameters of the Video Stream

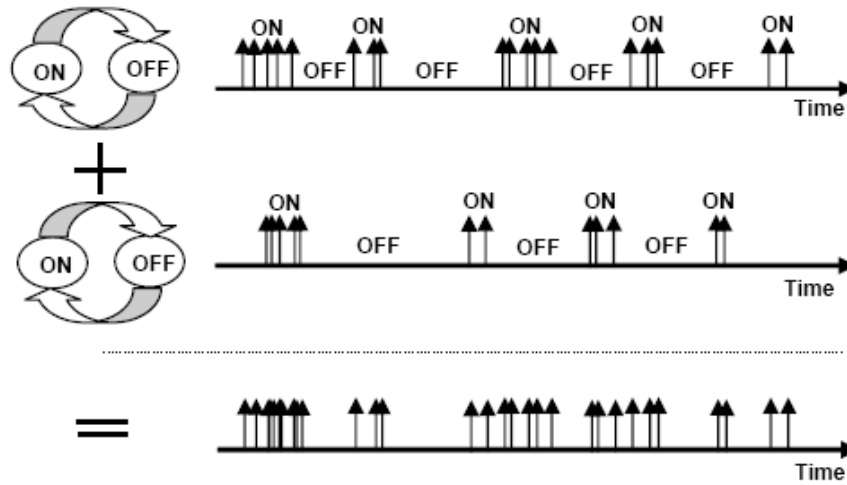


Figure 5-6 2IRP video traffic model (Taken from [37])

This is a standard model of video traffic often used in fixed broadband wireless access networks [38,39]. The average packet rate of one 2IRP generator is 1263.8 packets per second. The length of packets is assumed to be variable and is uniformly distributed between 176 to 200 bytes. Therefore, the average data rate for each user is 1.92 Mbps. The derivation of this is shown in Appendix C.

5.7 System View of the Simulation Model

The system diagram of the simulation model is shown in Figure 5-8. The WINNER model can support multiple service classes such as video, voice, FTP, web, and messaging.

Although in the simulation model shown here, only one type of traffic service class i.e. video streaming has been implemented, but the provisioning is there to implement more for future work. All the traffic generated in the system is organized in the following manner:

- I. Traffic from each service class is en-queued in a separate queue,
- II. Within each queue, each user is queued in a separate sub-queue.

As described earlier, the scheduler is invoked in the WINNER concept at the beginning of every frame. This is at a time scale of every 0.6912 msec. In each of these scheduling intervals, there are 96 chunks available to the scheduler (due to the OFDM time/frequency resource unit). In the WINNER context, there are two level of scheduling as shown in Figure 5-7.

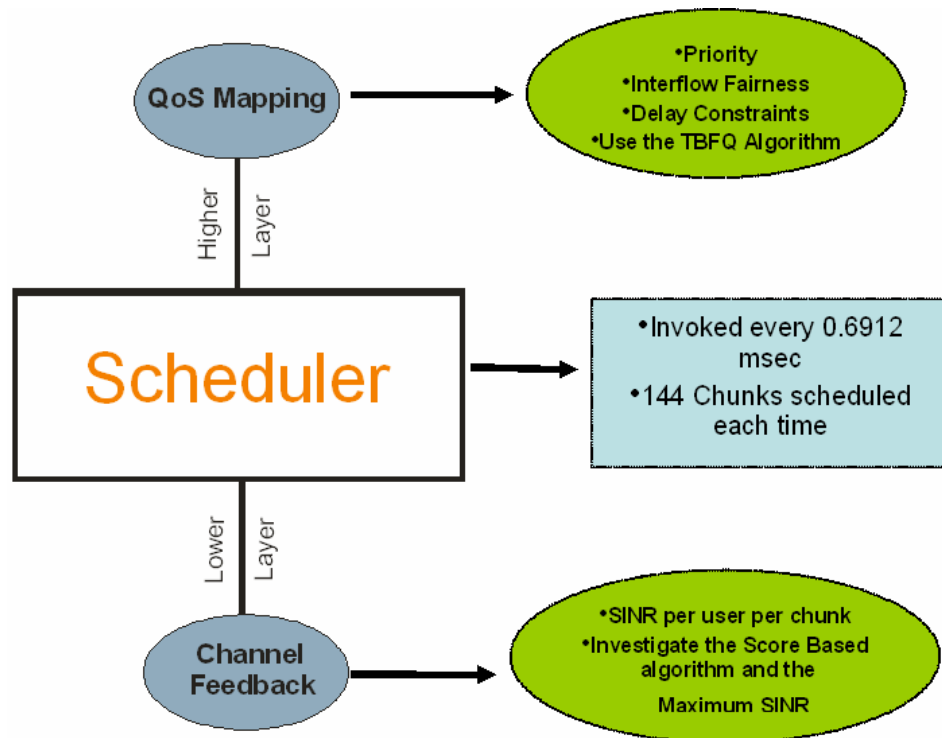


Figure 5-7 Two levels of Scheduling

The first level of scheduling takes into account the traffic queue levels and the QoS issues and then schedules the users accordingly. The factors taken into account are mainly:

- I. Priority
- II. Interflow Fairness
- III. Delay Constraints

The second level of scheduling deals with the allocation of chunks to the users selected by the first level of scheduling. This decision is based on the SINR feedback provided to the BS by all the UT for every chunk. In the considered model, ATBFQ is used at the first level. At the second level we use the Maximum SINR criteria for resource allocation. This block diagram for this operation is shown in the following figure.

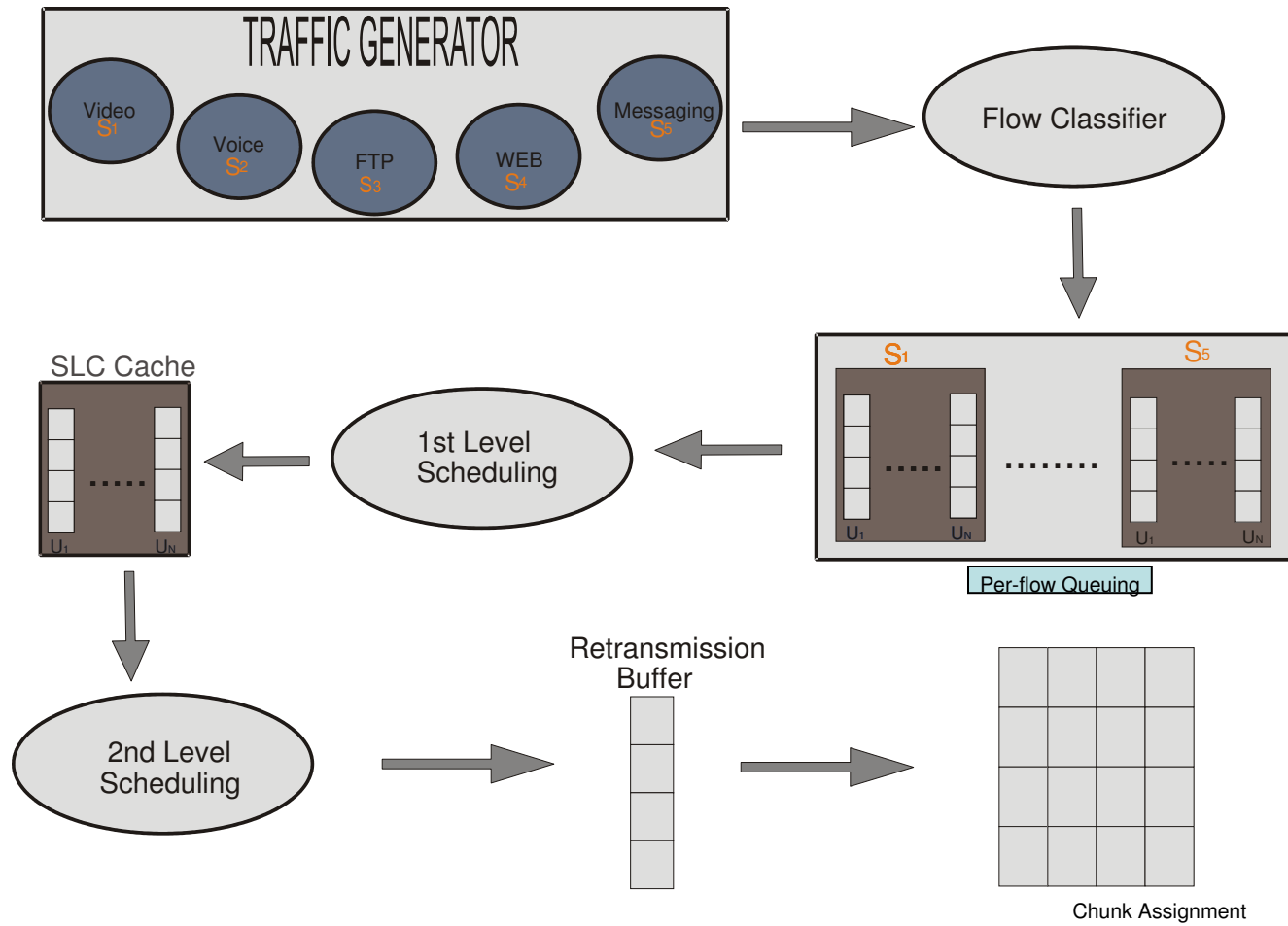


Figure 5-8 System level view of the simulation setup

5.8 Summary of the Simulation Assumptions

The simulations were carried out using a combination of both MATLAB and OPNET simulators. MATLAB is used to generate the fading dataset for the UTs in for desired as well as interferer links from the first tier. Shadowing and the antenna gain are also simulated in MATLAB. This information is provided to OPNET which models the scheduler located at the central BS. Traffic models for video streaming are also implemented in OPNET. Network activity is monitored for duration of 60 sec in varying loading and interference conditions. The table below summarizes the simulation parameters.

Parameter	Used Value/Model
Scenario	Wide Area DL(Frequency Adaptive)
Channel model	C2
Sector Tx antenna	120 ⁰ directional with baseline antenna pattern [40]
UT receive antenna	Omni-directional
Inter-site distance	1000 meters
Signal bandwidth	15 MHz (i.e., 48 chunks which is 1/3 rd of the baseline assumptions)
Frequency reuse	1 (i.e., 48 chunks available at each BS/sector)
Mobility	70 km/hr
Sector Tx power	46 dBm
Scheduler	Adaptive Token Bank Fair Queuing, Score Based , Round Robin
Interference model	Brute force Method (Central cell is considered with interference from the 1 st Tier)
Antenna configuration	SISO
Coding	B-LDPC
AMC modes	BPSK (rate 1/2 and 2/3), QPSK (rate 1/2, 2/3, and 3/4), 16QAM (rate 1/2, 2/3, and 3/4), and 64QAM (rate 2/3 and 3/4)
AMC thresholds	With FEC block of 1728 bits and 10% BLER
Frame duration	0.6912 ms (This is also the scheduling interval)
Service class	Video
Traffic model	1.9Mbps 2IRP model for MPEG video
Packet size	188 Bytes
Packet drop criterion	Delay>0.19 sec

Table 5-4 Summary of simulation assumptions

Chapter 6

Simulation Results

In this chapter, the ATBFQ scheduling scheme is investigated in the presence of interference at the system level. The results of the scheduling decisions are shown under two scenarios:

- Performance under varied network loading
- Performance under varied interference activity factors

The results of the proposed scheme are also compared to that of the reference SB algorithm and the traditional Round robin (RR) technique.

6.1 Assessment Criteria

It is necessary to identify the criteria which show the performance as well as the technical interpretations in light of the above mentioned scenarios.

Some of the common assessment criteria are:

- CDF of the user throughput
- CDF of the chunk SINR
- CDF of average user queuing delay
- The average sector throughput
- Performance of the cell edge users
- Packets dropped

CDF of user throughput is an important performance indicator, because it can be used for deriving the standard performance measure criteria like fairness, average throughput and spectrum efficiency.

The SINR measure is defined as the SINR after receiver processing but prior to decoding and geometrically averaged over the subcarriers in a chunk. In other words, this SINR represents the average channel quality on the chunk. Since networks' providers are often interested in maximizing average performance per cell or sector, the average cell/sector throughput is also an important performance criterion. Delay distribution is also an important performance criterion for end-to-end (E2E) performance evaluations, since delay experienced by a user has a great impact on user's satisfaction, especially in the case of real-time services.

Queuing delay is defined as the time interval from when the packet enters the transmission queue to when the packet is transmitted. Since ARQ is not considered in this study, therefore, the delays due to retransmission are not considered. CDF of the delay is important for scheduling investigations, since scheduling can make a trade-off between throughput and delay. Therefore, in the case of scheduling simulations, the delay CDF along with throughput CDF is also needed to evaluate the performance.

Another important performance indicator for real time traffic is the number of packets dropped. Although a scheme might achieve higher throughput, but that could be at the expense of resulting deprived users. In such a case, users will suffer with more packet dropping. Therefore packet dropping ratio is also a good indicator of fairness.

6.2 Spectral Efficiency

Figure 6-1 shows the comparison of spectral efficiency of the ATBFQ algorithm with that of SB (window size of 100) and RR algorithms when the number of users is varying. Two levels of interference AF's (0.5 and 0.7) are shown. It is seen that for low loading levels, ATBFQ outperforms SB and RR. This difference is due to the opportunistic nature of SB as it tries to take advantage of multiuser diversity which is not available at low loading levels. This becomes more evident at medium to high loading when SB outperforms ATBFQ and RR when there are more users. It tries to maximize the spectral efficiency by selecting users with the best channel conditions whereas ATBFQ tries to maintain fairness by distributing resources to deprived users

with bad channel conditions. This will become more evident when the performance results for queuing delay and packet drop will be shown in later sections.

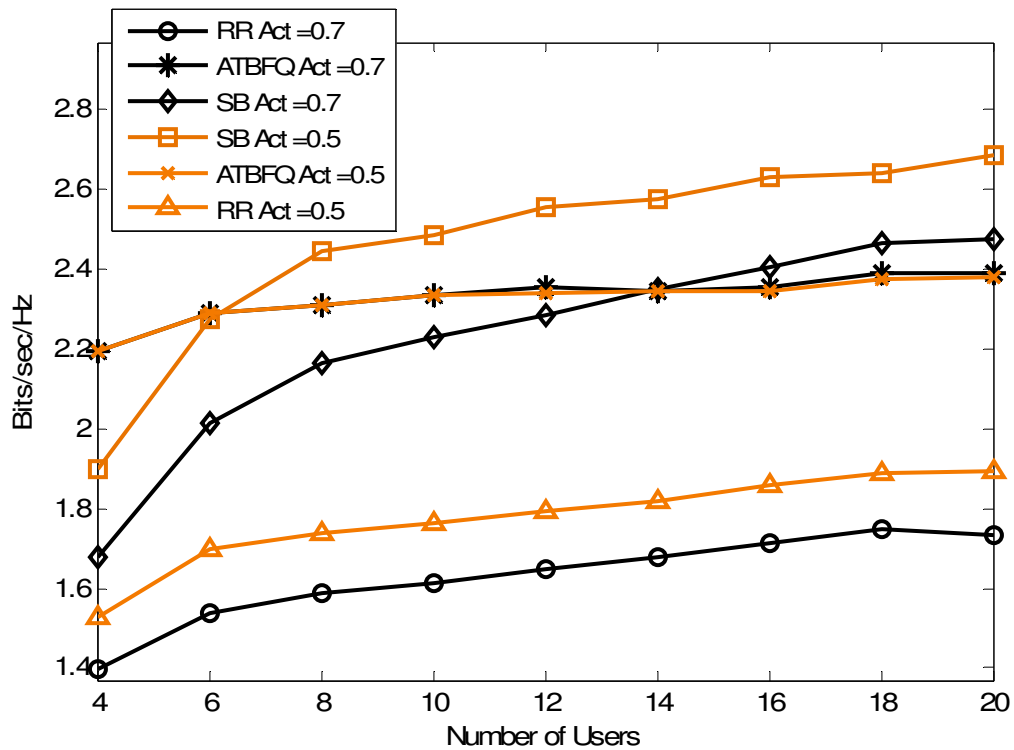


Figure 6-1 Average spectral efficiency vs. number of users

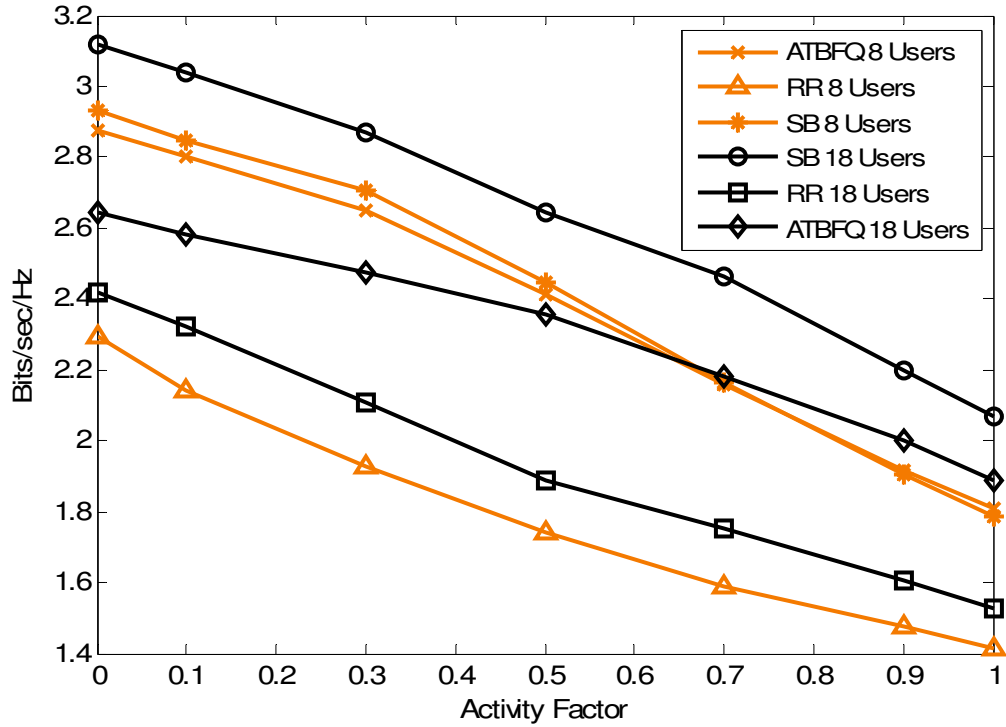


Figure 6-2 Average spectral efficiency vs. activity factor for low and high loading

Figure 6-2 shows the spectral efficiency vs. varying interference activity factors for low and high loading cases. Again the same trend is observed i.e. SB outperforms ATBFQ at low to medium interference whereas the performance is similar at high interference.

6.3 Queuing Delay

Figure 6-3 shows the average user queuing delay vs. number of users. ATBFQ is compared to SB and the RR schemes at two different AFs of 0.5 and 0.7. At an AF=0.5, the performance of ATBFQ is much better than SB. It is seen that average user queuing delay is almost constant till 14 users for ATBFQ and then tends to increase exponentially. If the AF is increased to 0.7, the performance of ATBFQ still outperforms SB. At a lower AF, the users experience better channel conditions. Hence fewer chunks are utilised to transmit data as compared to a higher AF. This is the reason why RR performs better than SB at a lower loading.

The same trend is observed in Figure 6-4. This shows the average user queuing delay vs. different AFs. Two different cases are shown here. We see the performance in the case of low loading (8 users) and high loading (20 users). The trend for both these cases is the same in the presence of low to medium interference i.e. ATBFQ outperforms SB. But at high interference SB outperforms ATBFQ. This is explained by the fact that ATBFQ attempts to maintain fairness among users and due to the high interference conditions, the average user queuing delay increases.

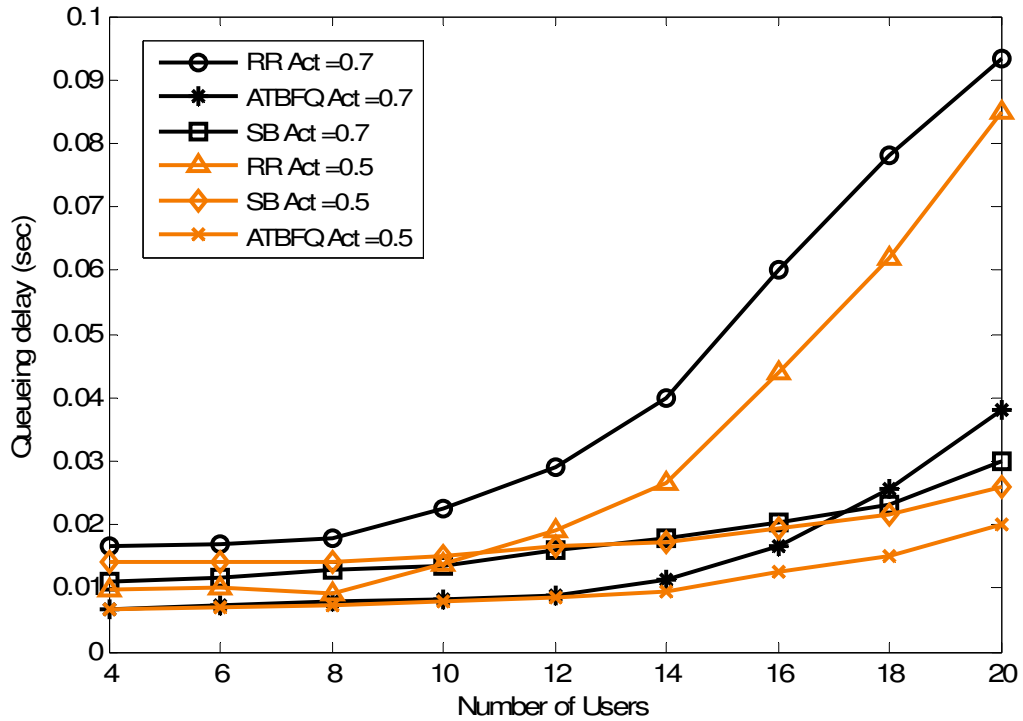


Figure 6-3 Average user queuing delay for medium and high interference scenarios

Figure 6-5 and Figure 6-6 show the CDF of the average user queuing delay for low (8 users) and high (20 users) loading respectively. The performance of ATBFQ is compared to that of SB and RR at an AF=0.7. For low loading values it is shown that RR outperforms both the ATBFQ and SB. We observe for high loading that although the performance degrades for both ATBFQ and SB as expected, SB slightly outperforms ATBFQ. This is explained due to the fact

that this delay is only considered for transmitted packets (the delay statistics for the dropped packets are not taken into account while calculating the delay CDF's). Thus as ATBFQ tries to maintain fairness at high loading, it incurs high queuing delays.

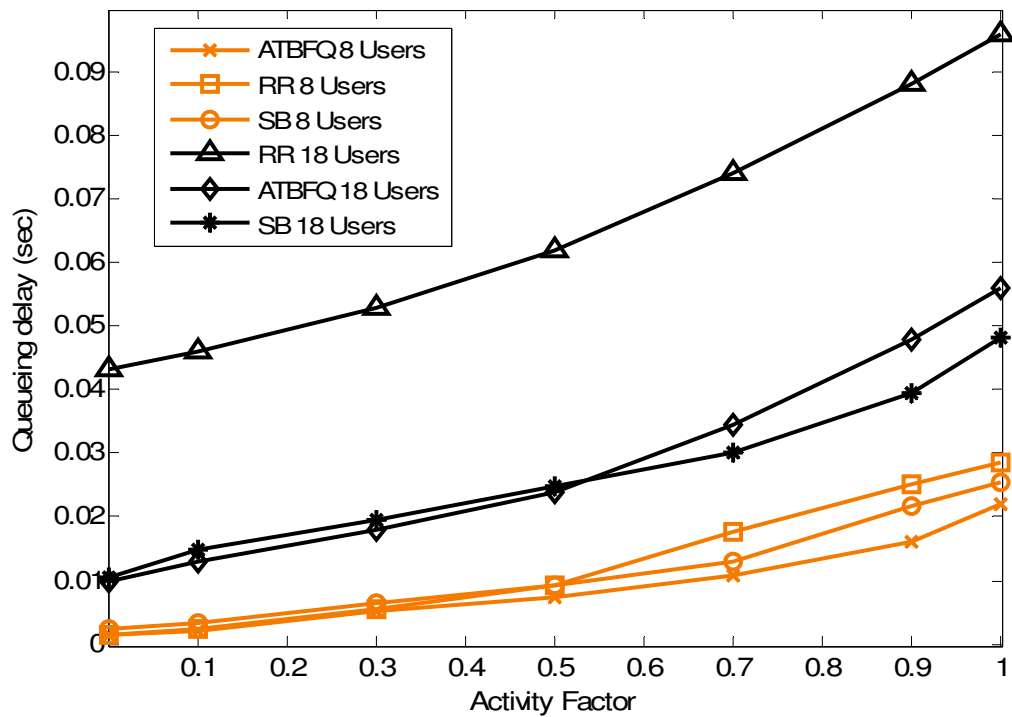


Figure 6-4 Average queuing delay vs. different activity factors for low and high loading

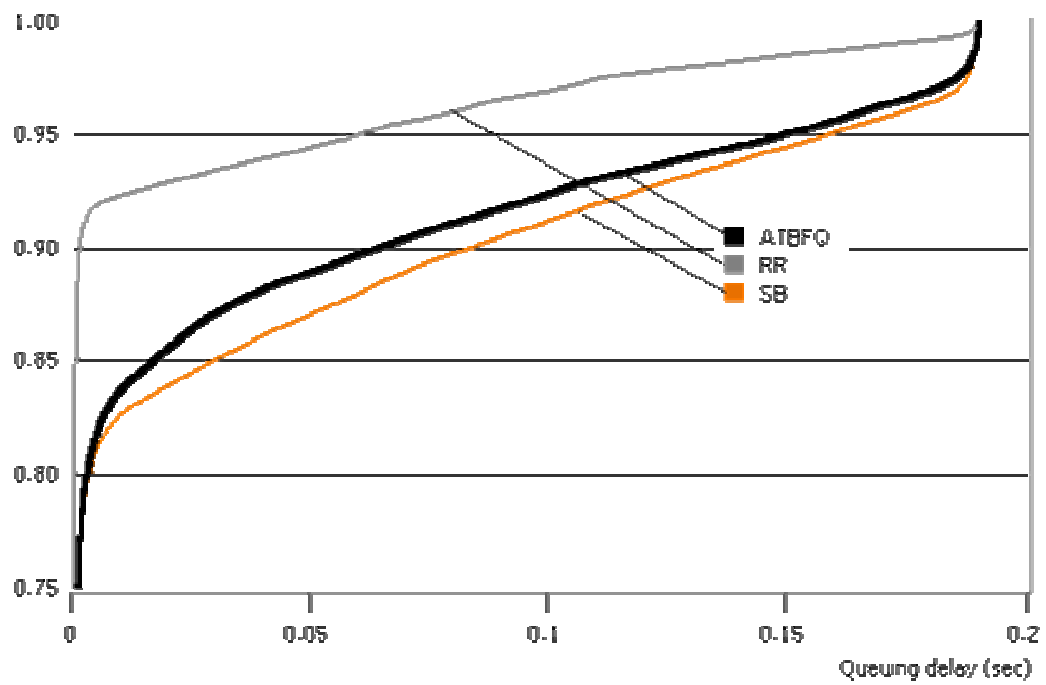


Figure 6-5 CDF of average queuing delay for low loading (8 users)

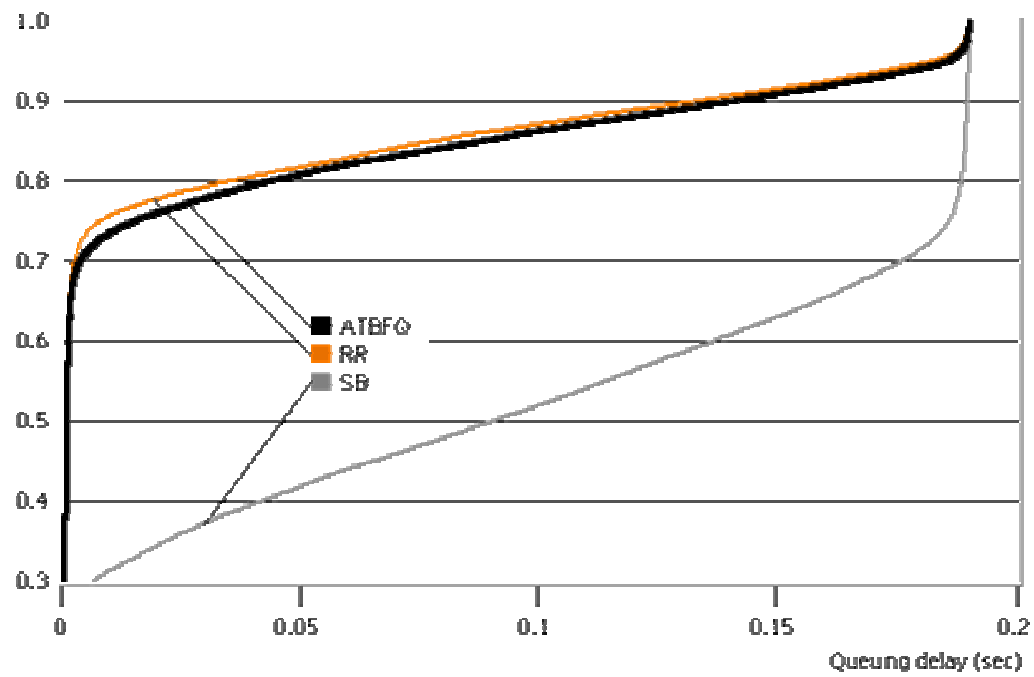


Figure 6-6 CDF of average user queuing delay for high loading (20 Users)

6.4 Packets Dropped

As the traffic type considered for this study is video streaming, therefore packets are dropped if they face a queuing delay of more than 0.19 seconds. Figure 6-7 shows average packets dropped per frame vs. number of users. This is shown for two different AF's of 0.5 and 0.7. At an AF=0.5 we observe the average packets dropped per frame for ATBFQ is less than SB. The curve is almost constant till 14 users and then increases exponentially with the number of users. The same trend is observed for SB but with more packets dropped. At higher interference (shown by an AF =0.7), we observe the same trend for ATBFQ as compared to the lower interference between lower to medium loading. But at higher loading values, the number of packets dropped per frame is comparable for both ATBFQ and SB. The performance for the RR is also shown. It can be seen that for low loading and a lower interference AF, it outperforms SB in terms of packets dropped.

This trend is again visible in Figure 6-8 which shows the average packets dropped per frame vs. different AFs for low and high loading values. Figure 6-9 and Figure 6-10 show the CDF of the packets dropped per frame for low and high loading, respectively. These curves again indicate the opportunistic nature of SB as it only tends to favor the users with good channel conditions. That's why a higher drop rate even when there is low loading is observed.

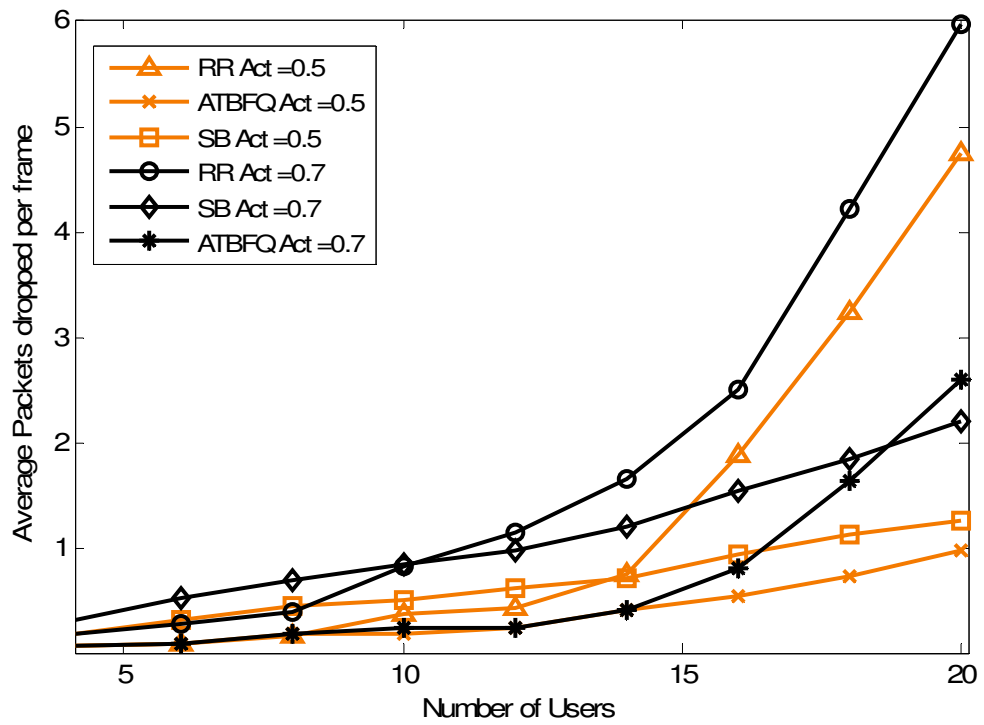


Figure 6-7 Average packets dropped vs. number of users

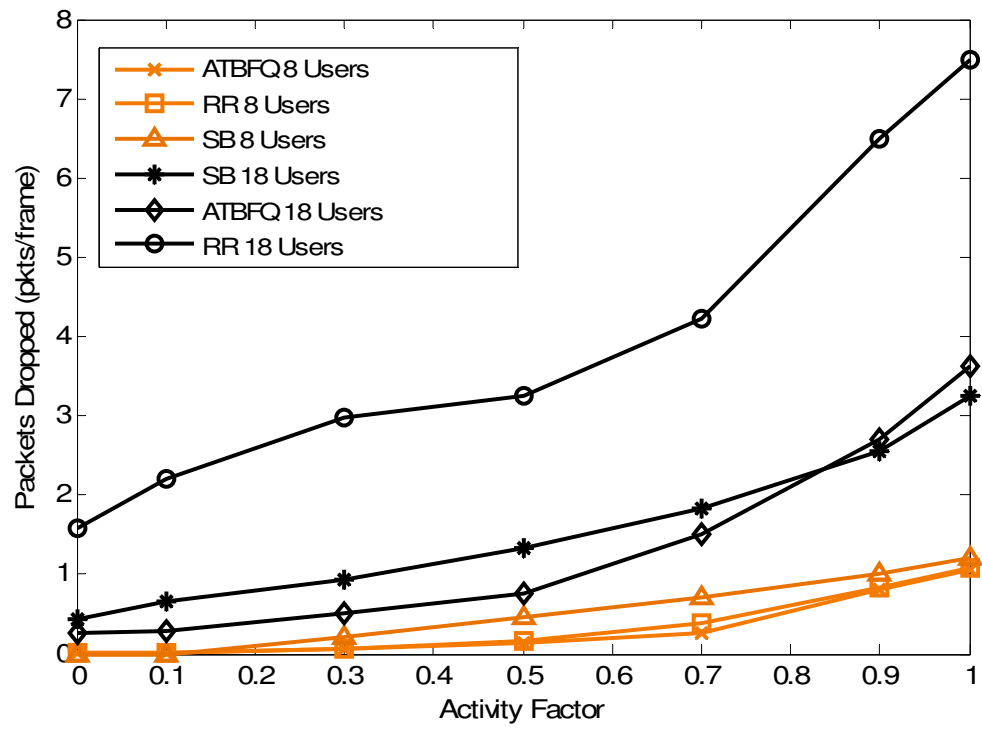


Figure 6-8 Average packets dropped vs. different activity factor for low and high loading

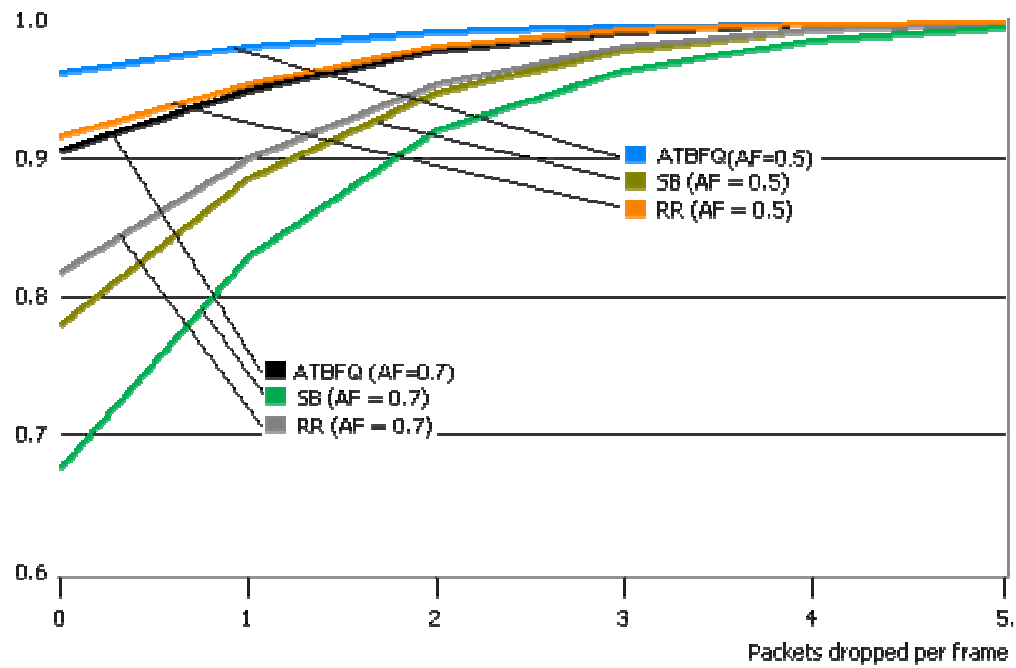


Figure 6-9 CDF of packets dropped per frame for 8 users at AF=0.5, 0.7

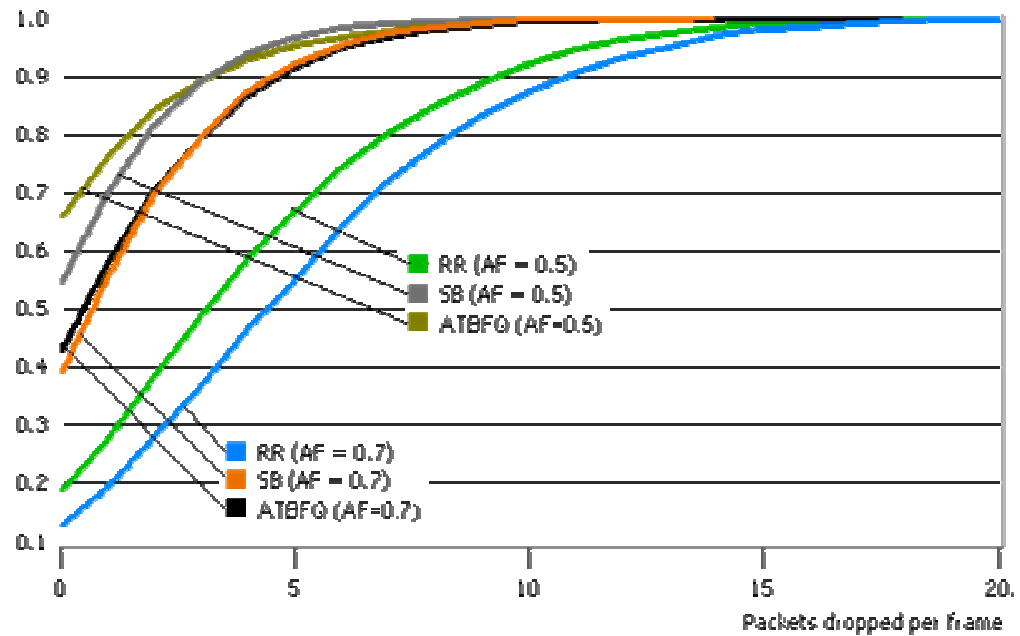


Figure 6-10 CDF of packets dropped per frame for 20 users at AF=0.5, 0.7

6.5 SINR

Figure 6-11 and Figure 6-12 compare the CDF of the scheduled SINR for ATBFQ, RR and SB for 8 and 20 users, respectively at an AF of 0.7. At low loading it is observed that ATBFQ outperforms SB and RR. Since maximum SINR method is used to assign chunks to the selected users in ATBFQ, the chunks with the highest SINR's are allocated to the scheduled users. At low loading, SB cannot take advantage of multiuser diversity and thus suffers. In the case of high loading values, it is seen that SB outperforms ATBFQ as it takes advantage of multiuser diversity.

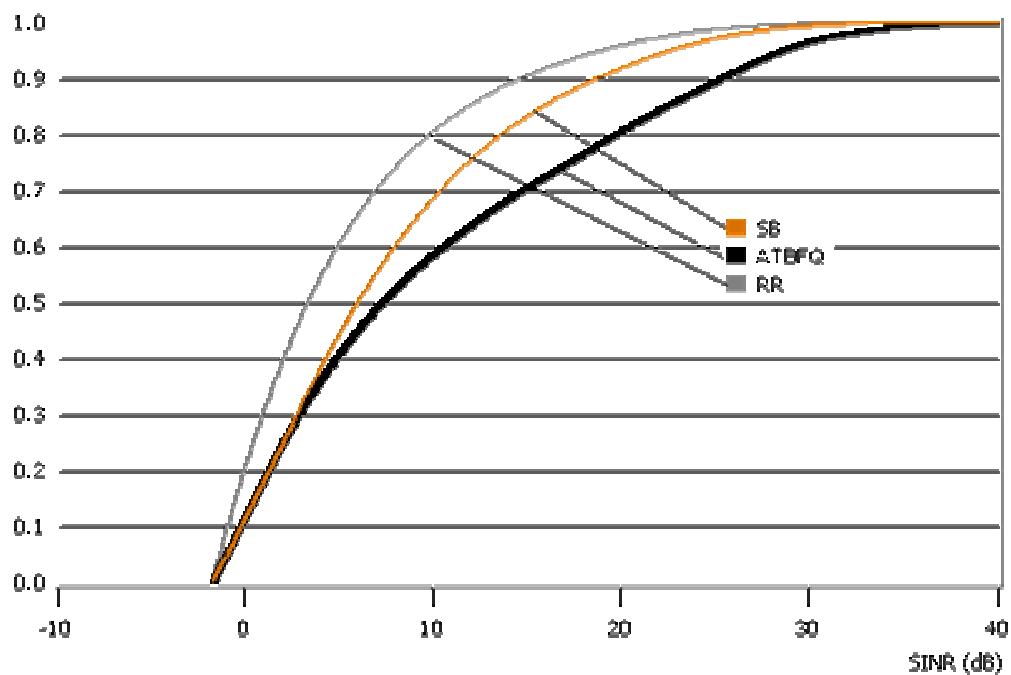


Figure 6-11 CDF of SINR for 8 users on scheduled chunks

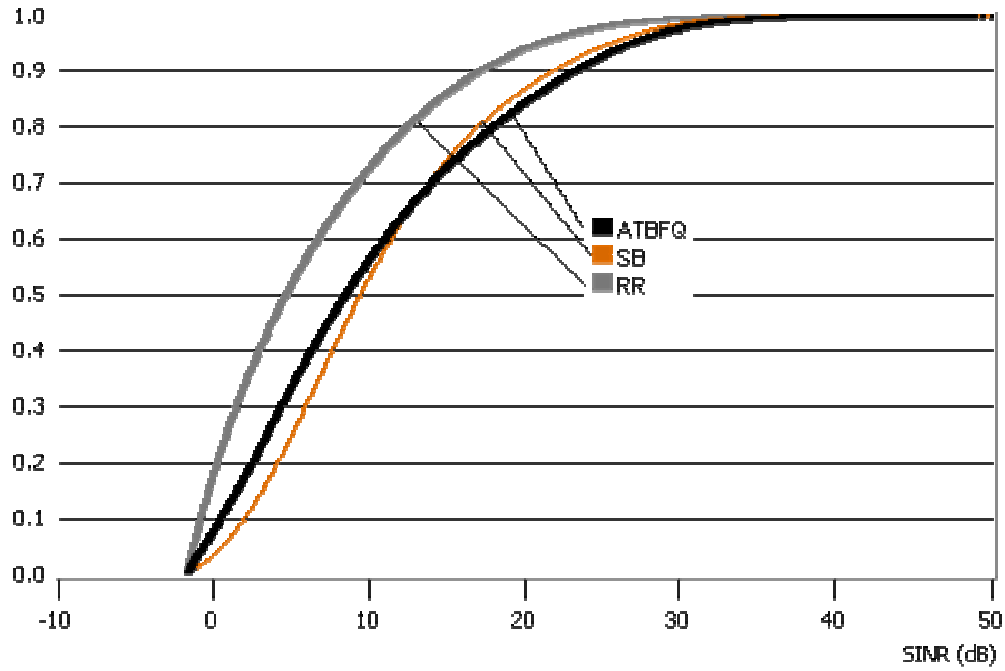


Figure 6-12 CDF of SINR for 20 users on scheduled chunks

6.6 Throughput

Figure 6-13 and Figure 6-14 show the CDF of average user throughput (bytes per frame) for 8 and 20 users, respectively. ATBFQ has higher average user throughput for lower loading cases whereas SB achieves higher average throughput at higher loading values. For the high loading also it is observed that the curve for ATBFQ has a steeper slope. This usually indicates fairness as more users are serviced with similar throughput. This is not the case for SB.

The sector throughput with an AF of 0.7 is seen in Figure 6-15. ATBFQ performs better at lower-medium loading levels. But as the number of users approaches 20, SB achieves higher throughput. This is because SB is opportunistic in nature whereas ATBFQ tries to maintain fairness. As the number of users increase, SB takes advantage of the multi-user diversity to achieve higher throughput.

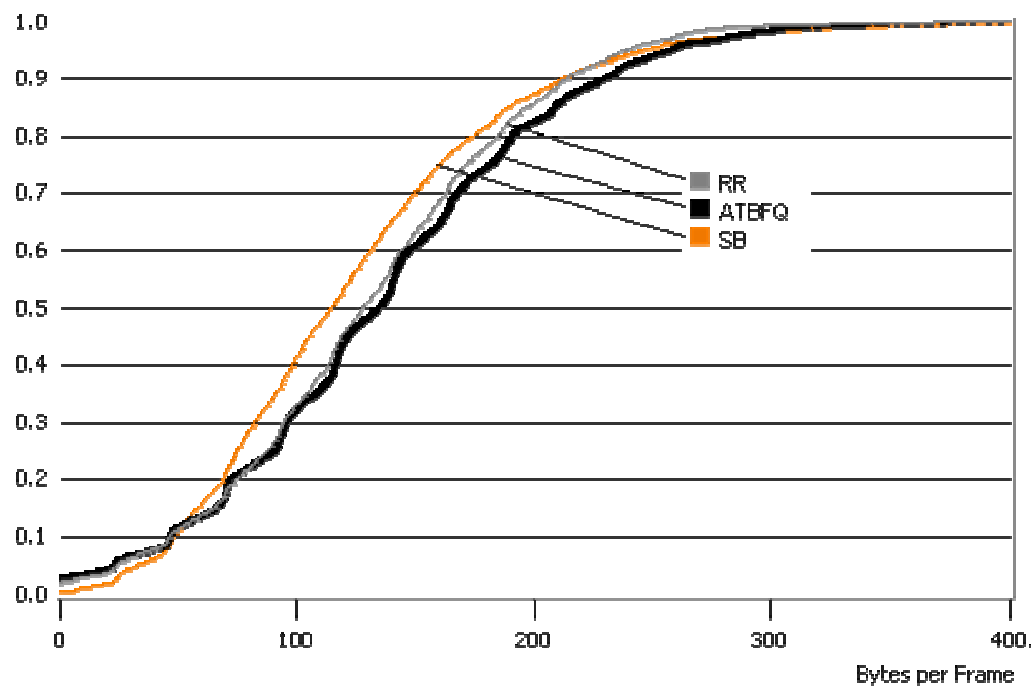


Figure 6-13 CDF of throughput (bytes per frame per sector) for 8 users

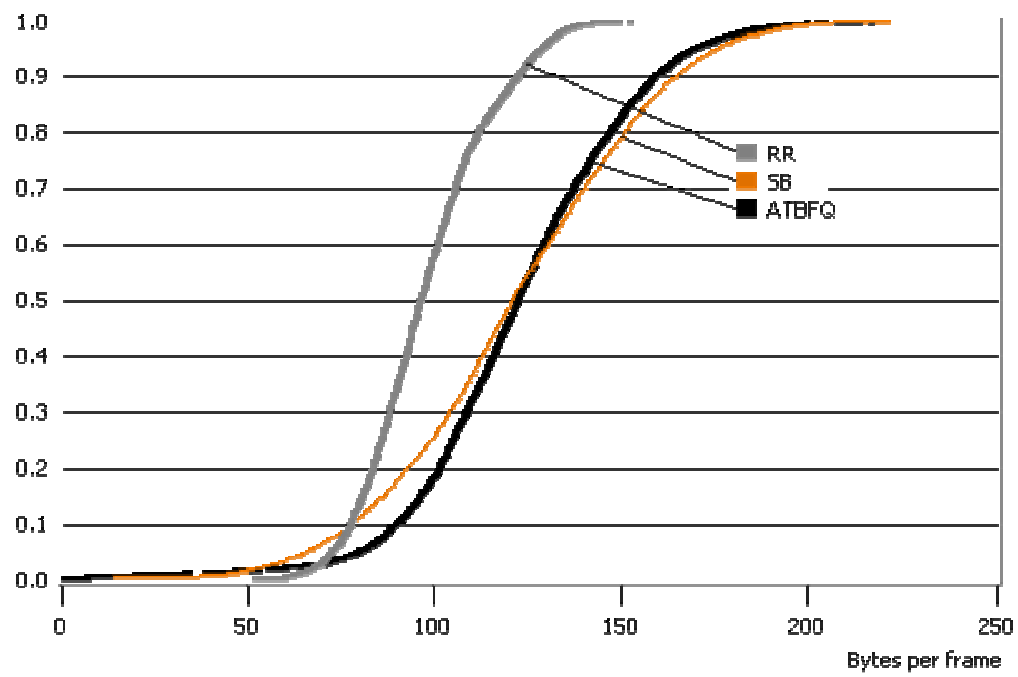


Figure 6-14 CDF of throughput (bytes per frame per sector) for 20 users

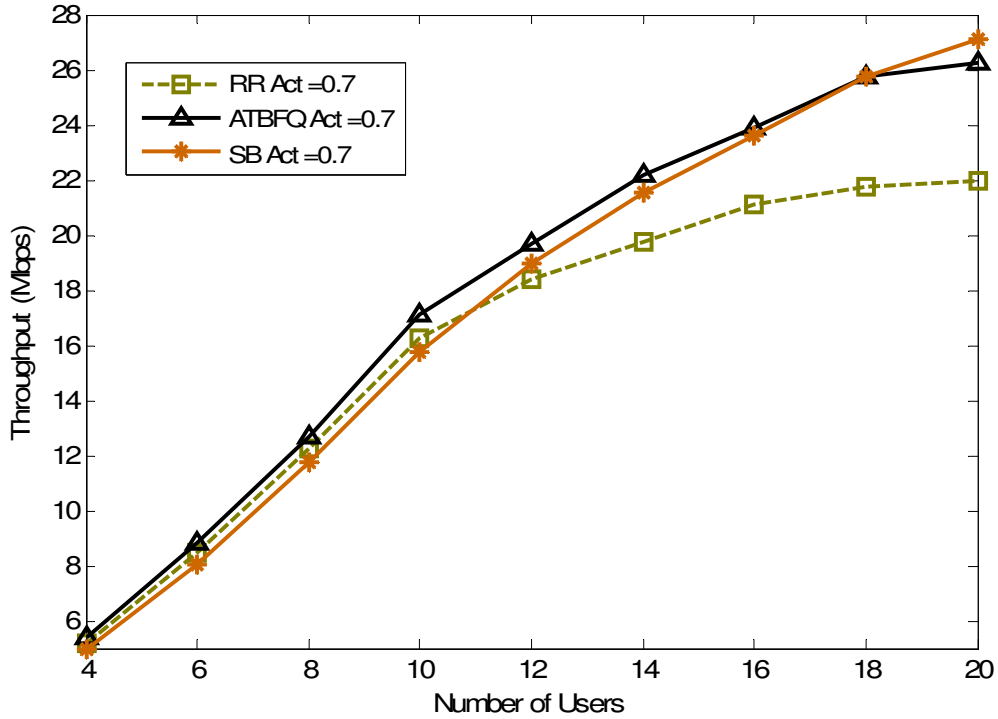


Figure 6-15 Total sector throughput vs. number of users

6.7 Performance vs. Distance

Figure 6-16 shows the packet transmit ratio (defined as the packets transmitted/total packets generated) vs. distance from BS for 20 users. It can be observed that as the distance increases, the packet transmit ratio for SB decreases i.e. the number of dropped packets increases. This can be further visualized by the fitted curves for both algorithms which show their respective trends with the varying distance. As SB tries to maximize the throughput, the cell edge users are affected and suffer packet losses. ATBFQ on the other hand is fair in nature and tries to offer better rates to the cell edge users. If a cell edge user is suffering from bad channel conditions, ATBFQ gives it priority to transmit in the next scheduling interval. By assigning priorities in such a manner, ATBFQ also keeps track of the queue levels and tries to maintain constant queuing delay for the cell edge users as shown in Figure 6-17. The same cannot be said for SB as the average user queuing delay increases exponentially with the distance.

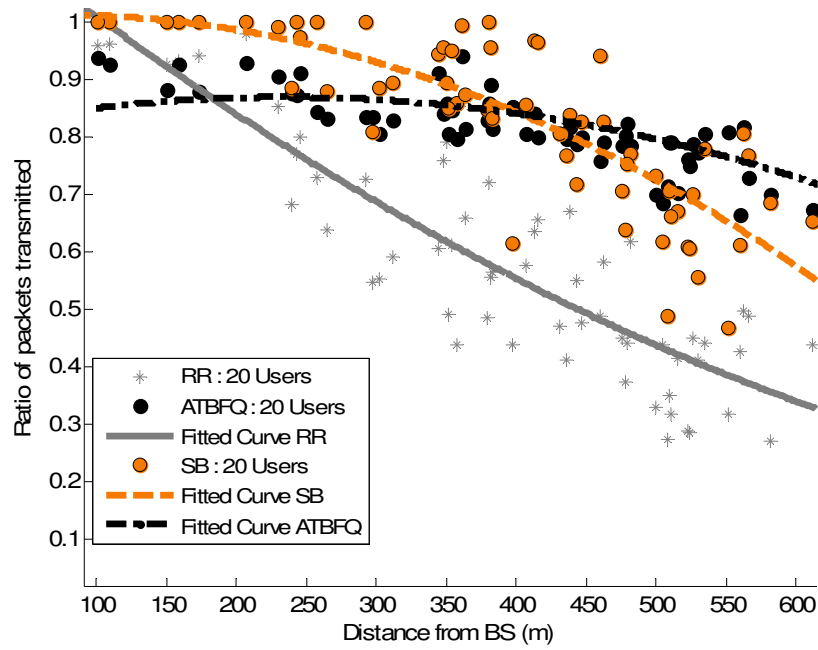


Figure 6-16 Ratio of packets transmitted vs. distance from BS

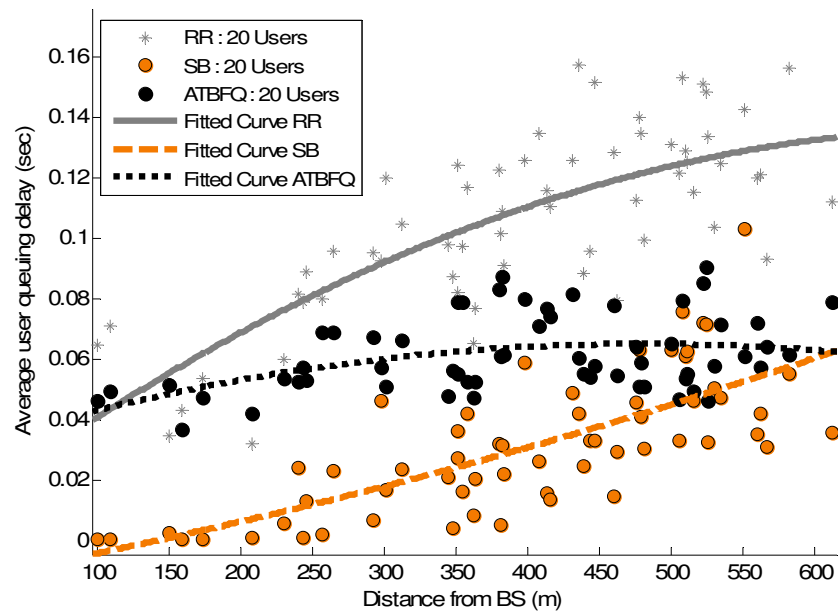


Figure 6-17 Average queuing delay per user vs. distance from BS

To further emphasize the advantage ATBFQ provides to cell edge users, the performance of two users is compared; one close to the BS (150 m) and the other at the cell edge (527 m). For convenience these users are called 1 and 2 respectively. The performance of both these users is studied for low and high loading cases for the ATBFQ, SB and RR algorithms. In Table 6-1, we look at the low loading case. The average queuing delay, average packets dropped and the average throughput are compared. It is observed that the performance of User 1 is comparable for all three algorithms. But in the case of User 2, the packets dropped increases whereas the throughput decreases for both the SB and RR algorithms. The delay performance of User 2 can be further visualized by observing the CDF of the average queuing delay for all three algorithms in Figure 6-18.

	ATBFQ			SB			RR		
	<i>Queuing Delay (ms)</i>	<i>Avg. Packets Dropped</i>	<i>Avg. Throughput (Mbps)</i>	<i>Queuing Delay (ms)</i>	<i>Avg. Packets Dropped</i>	<i>Avg. Throughput (Mbps)</i>	<i>Queuing Delay (ms)</i>	<i>Avg. Packets Dropped</i>	<i>Avg. Throughput (Mbps)</i>
User1 (150 m)	0.6	0	1.9	0.8	0	1.9	0.6	0	1.9
User2 (552 m)	28	0.1101	1.69	55	0.31	1.32	51	0.1918	1.52

Table 6-1 Comparison of ATBFQ, SB and the RR algorithms for a low loading scenario

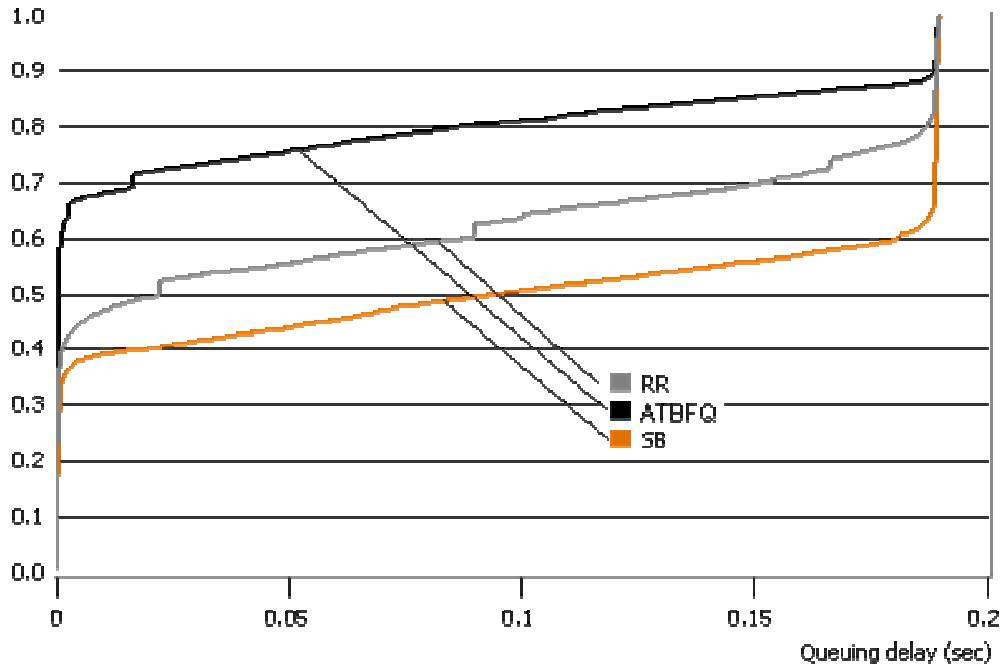


Figure 6-18 CDF of average queuing delay for cell edge User 2 for low loading case

Next the performance of both User 1 and User 2 is compared in a high loading scenario in Table 6-2. It is observed that for User 1, the performance is comparable for all three algorithms with SB faring the best. But for User 2, ATBFQ outperforms both SB and RR. We observe a throughput increase of a factor of 30% over SB and 140% over RR. The average queuing delay is also less for ATBFQ. We can further visualize the delay performance in Figure 6-19 and Figure 6-20 for User 1 and User 2, respectively.

	ATBFQ			SB			RR		
	Queuing Delay (ms)	Packets dropped per frame	Throughput (Mbps)	Queuing Delay (ms)	Packets dropped per frame	Throughput (Mbps)	Queuing Delay (ms)	Packets dropped per frame	Throughput (Mbps)
User1 (150 m)	17.6	0.044	1.816	5.1	0.003	1.89	35.6	0.05775	1.79
User2 (552 m)	57.6	0.169	1.5	85.2	0.3821	1.17	129	0.65	0.66

Table 6-2 Comparison of ATBFQ, SB and the RR algorithms for a high loading scenario

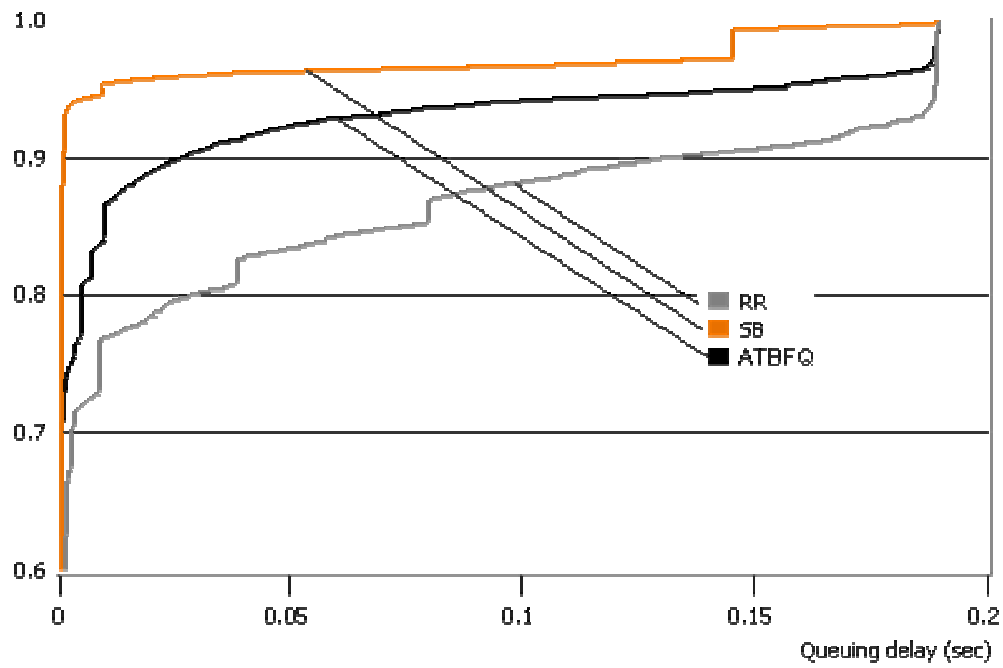


Figure 6-19 CDF of average queuing delay for User1 in a high loading scenario

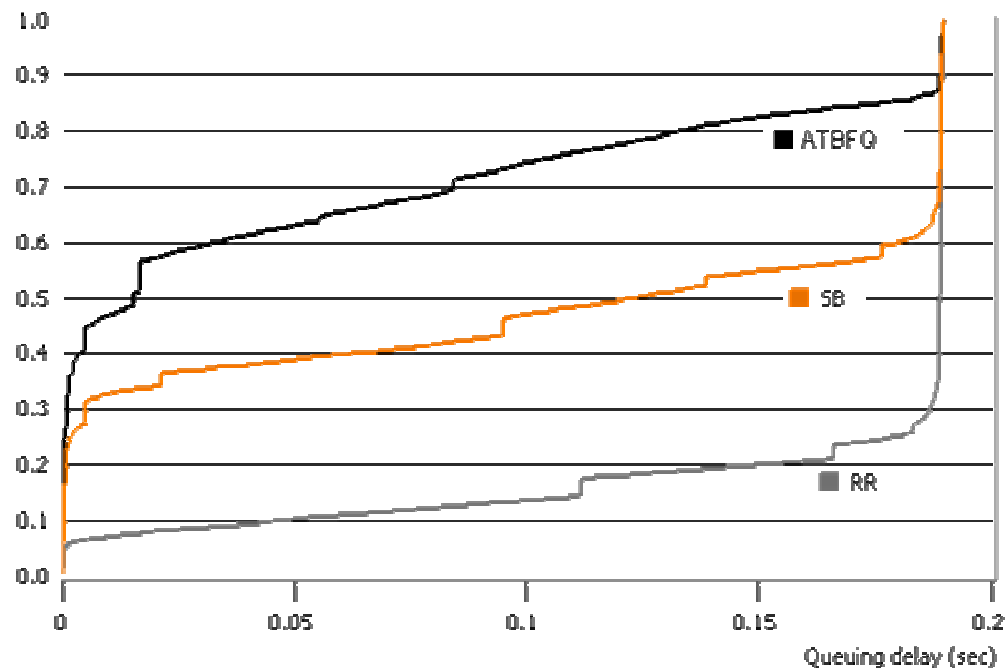


Figure 6-20 CDF of average queuing delay for User 2 in a high loading scenario

6.8 Fairness Analysis

The performance of wireless networking systems is hard to judge solely based on performance metrics such throughput or spectral efficiency. This is mainly due to the fact that there are scheduling algorithms which work to achieve better performance for “favored users” or users receiving good channel conditions. Under such scheduling schemes, the users close to the cell edge (cell edge users) suffer tremendously as most of the bandwidth is scheduled to users closer to the BS. In such circumstances, the overall throughput of the system is maximized but the fairness amongst all users is greatly affected. Therefore it is essential to design a performance metric which is a good indicator of the fairness amongst the users. One such index is the fairness index proposed by Jain in [46]. This fairness index is bounded between zero and unity and has been used in recent papers [47, 48]. If a system allocates resources to n contending users such that an i^{th} user receives an allocation x_i , then this fairness index $f_I(x)$ is given by:

$$f_I(x) = \frac{\left[\sum_{i=1}^n x_i \right]^2}{n \sum_{i=1}^n x_i^2}, \quad 6-1$$

where $x_i \geq 0$. This index measures the equality of user allocation x . If all users get the same amount, i.e. x_i s are equal, then the fairness index is 1 and the system is 100% fair. As the disparity increases, the fairness decreases and only a select few of the users are allocated resources with the fairness index close to 0.

The fairness index is calculated based on the metric ‘ x ’ and will vary with different metrics considered. Example of such metrics could be throughput, delay or power. In this thesis this metric is considered to be the throughput. The fairness is calculated in two different cases depending on the timescale during which the fairness is calculated. These two cases are called short term fairness and long term fairness. The metric ‘ x ’ is modified to take into account the queue size also and is given by

$$x_i = \frac{\text{total_bits_transmitted_in_current_frame_for_user_i}}{\text{total_bits_in_queue_for_user_i}}. \quad 6-2$$

The reason for normalizing the throughput is to remove ambiguity as the throughput metric on its own does not provide an insight into the overall fairness of the system when using modeled traffic. A user might have limited data in its queues in which case it will have less data to transmit in the frame. Similarly, another user might get to transmit considerable amount of data in a frame, simply because its queues contain more packets. Therefore if the fairness is calculated solely based on the throughput, then such users will skew the overall fairness statistics.

Using the quantity ' x_i ' given in Eq. 5-3, the fairness at the frame level can be calculated. This method is called the short term fairness because of the minute frame scale at which the fairness is calculated. In Figure 6-21, the average short term fairness is shown for the low loading cases for time duration of 25 sec. It is observed that ATBFQ is fair to 83% of the users on the average in every frame. For the high loading case shown in Figure 6-22, we also observe that ATBFQ is fairer than both RR and ATBFQ at all times.

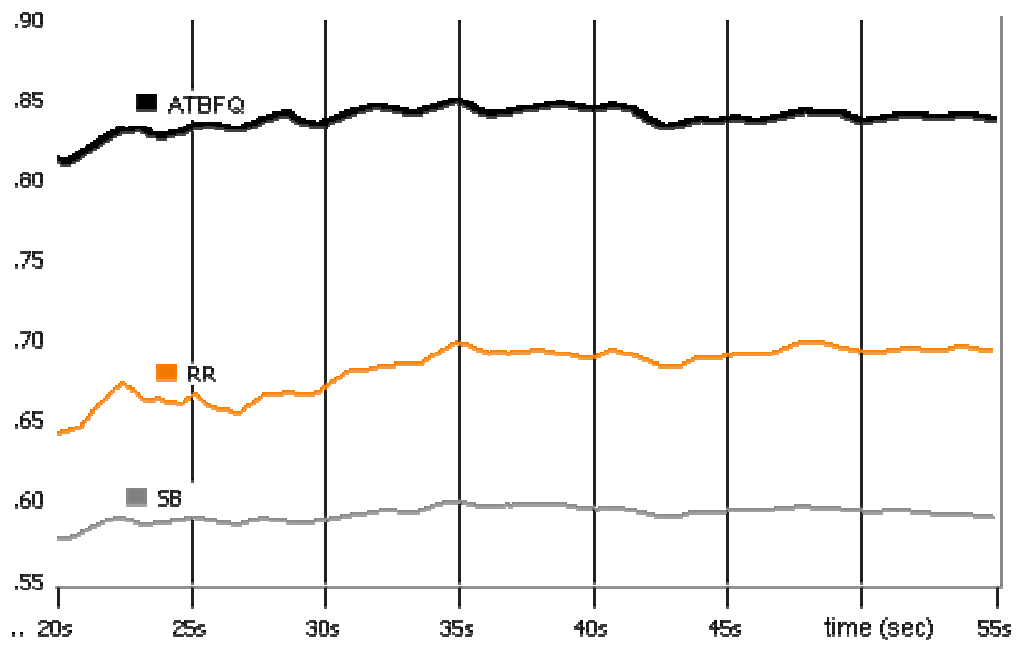


Figure 6-21 Average short term fairness for low loading case (shown over 35 sec of simulation time)

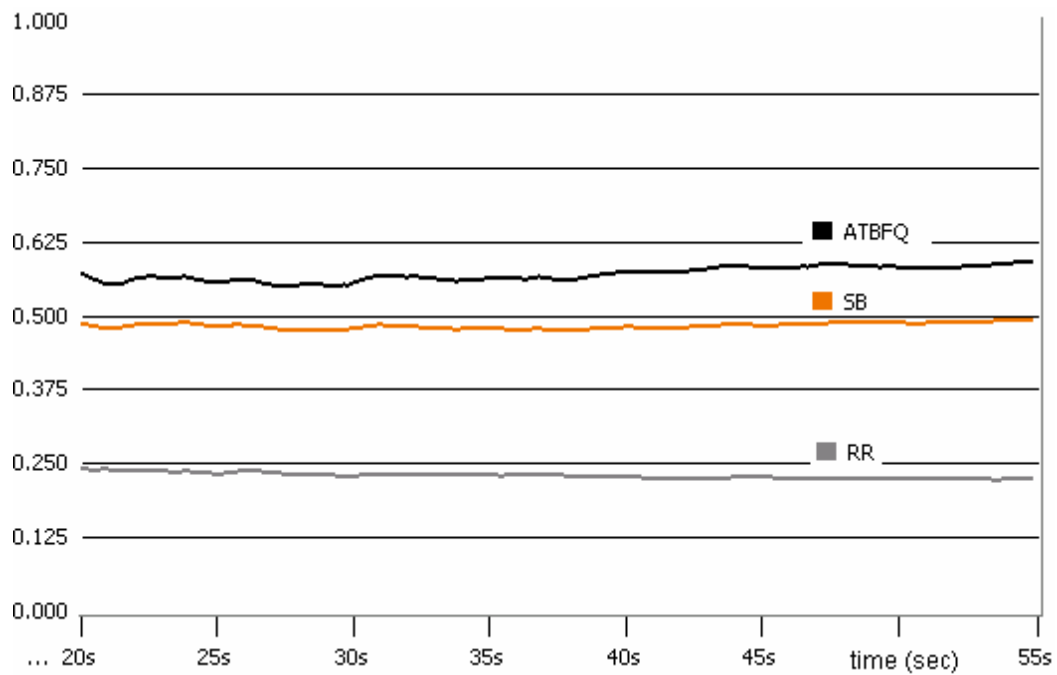


Figure 6-22 Average short term fairness for high loading

The long term fairness is calculated on the metric x_i given by:

$$x_i = \left[\frac{TP_i}{Q_i} \right]_{(t_1, t_2)},$$

where TP_i is the transmitted throughput for user i in the time interval $[t_1, t_2]$ in bits and Q_i is the total queue size for user i given by $Q_i = Q_i^{t_2} - Q_i^{t_1}$ where $Q_i^{t_2}$ and $Q_i^{t_1}$ are the queue sizes of user i at time instant t_1 and t_2 respectively. We chose $t_2 - t_1 = 11.06$ ms which is equivalent to 16 frames. In Figure 6-23 and Figure 6-24, the CDF plots for the long term fairness are shown for low and high loading cases respectively. It is again observed that ATBFQ is more fair to the users as compared to SB.

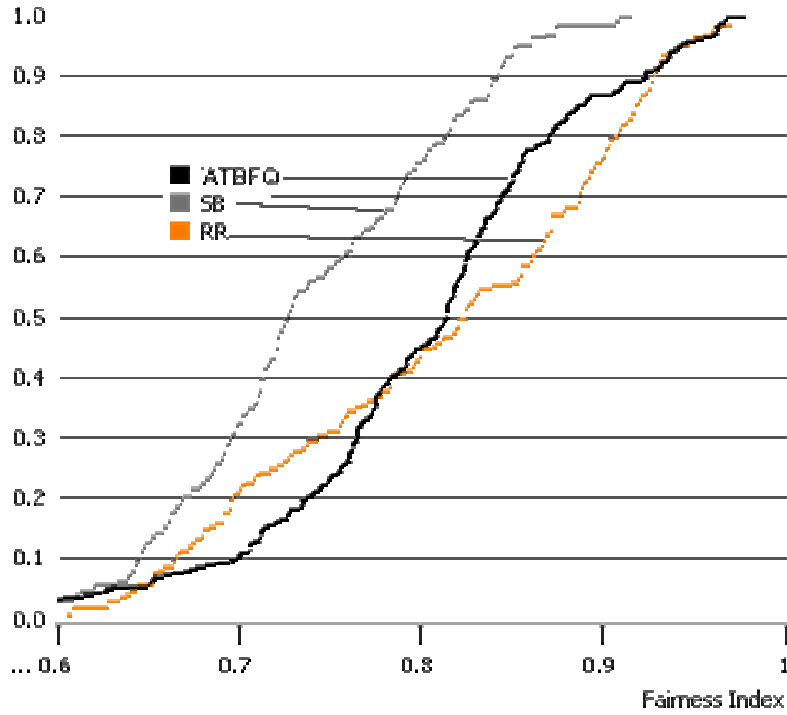


Figure 6-23 CDF for long term fairness for low loading case

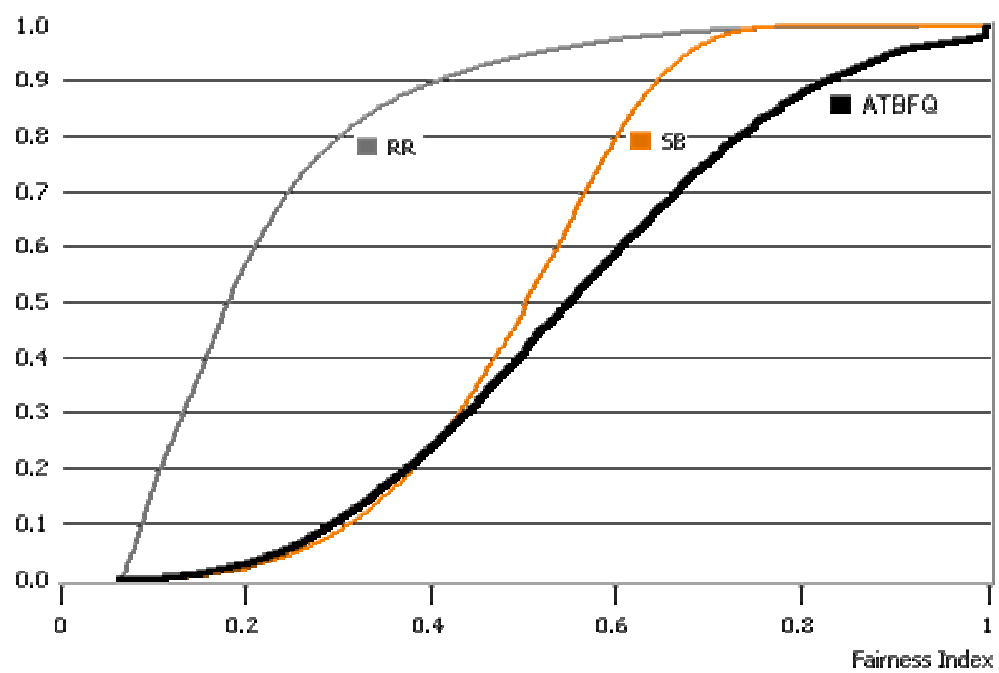


Figure 6-24 CDF for long term fairness for high loading case

Chapter 7

Conclusions and Proposals for Future Work

7.1 Conclusions

To meet the increasing demand for wireless services, achieving high wireless spectral efficiency is becoming increasingly important. In wireless networks, users experience unreliable, location-dependent, and time-varying channel conditions. Opportunistic scheduling exploits the variation of channel conditions to improve spectral efficiency, but this comes at a certain cost of fairness as those users suffering from unfavorable channel conditions suffer considerably. It is thus required to design a scheduling algorithm that not only keeps track of the short term benefits where the channel conditions are taken into account but also maintain long term fairness where minimum QoS requirements are met for all users. Motivated by this vision, the generic TBFQ wireless scheduling algorithm was proposed and developed for single carrier systems.

In this thesis, the TBFQ algorithm has been modified (ATBFQ) to satisfy the requirements for the WINNER system which caters specifically to the growth of Internet data and multimedia applications. It is a queue and channel aware scheduling algorithm which attempts to maintain fairness among all users. The algorithm is studied in the presence of varying interference activity levels and with varying loading conditions. Performance of the ATBFQ is shown with reference to the RR scheduler and the SB scheduler which is the current baseline scheduler in the WINNER system. SB is an opportunistic scheduler belonging to the proportional fair class. It tries to maximize throughput, making use of multiuser diversity while trying to maintain fairness. But this comes at a certain cost as the cell edge users in this scheme suffering

from bad channel conditions are more severely affected. Also, due to the bursty nature of the modeled traffic, such users face higher queuing delays which result in higher packet drops.

Unlike SB, ATBFQ is a credit based scheme which tries to accommodate the burstiness of the users by assigning them more resources in the short term, provided that long term fairness is maintained. For lower to medium loading and at lower interference levels, ATBFQ performs better than SB, as SB cannot take advantage of multiuser diversity. The performance of ATBFQ is better in terms of user throughput, queuing delay, and packet dropping. An improvement is seen also in the overall sector throughput. At high loading ATBFQ still performs better than SB with regards to the queuing delay and packet dropping while the sector throughput is slightly affected. Due to this, SB might be able to accommodate more users but the overall fairness of the users is affected, therefore jeopardizing user experience.

Another advantage that ATBFQ achieves over SB is that it provides fairness to cell edge users. The cost of this fairness is increased average user queuing delay and a slight reduction in the overall sector throughput but this may result in substantial savings in the deployment cost since a fewer number of base stations (BS) will be needed to cover regions.

7.2 Thesis Contribution

We summarize the following contributions achieved throughout the course of this thesis study:

- The generic TBFQ algorithm is modified to accommodate the requirements for a multicarrier system such that can it accommodate many users with widely ranging applications, data rates, and QoS requirements. The scheduling is carried out in both time- and frequency domain in order to take advantage of the channel fluctuations by using various AMC modes.

- An adaptive burst credit method for chunk allocation for the ATBFQ scheduler is developed which depends on the user efficiency. Based on this method, chunk resources can be assigned to users depending on the loading conditions. For example, for low loading conditions, the adaptive burst credit can be used to achieve higher throughput whereas in high loading conditions, it can be used to maintain fairness among users.
- The SB scheduler which is the current baseline scheduling scheme in WINNER is modified for multicarrier systems and is used as a benchmark for the proposed ATBFQ scheduling scheme.
- The performance of this modified scheme is studied in the context of the 4G WINNER system. A simulation model for the downlink is built adherent to specifications of this system. The simulation model built consists of the following components:
 - A traffic model which realistically models the burstiness of the video streaming service class,
 - An inter-cell interference model which takes the interference from the first tier of BSs into account,
 - A channel model which accurately depicts the large scale path loss, shadowing and fading for a micro-cell urban environment,
 - To compensate for fast and slow channel variation, a link adaptation technique such as adaptive modulation and coding is employed,
- A fairness index is used to compare the fairness between the proposed scheme and the reference schemes. This index is modified to take into account both the short term and the long term fairness.
- It is also shown how the ATBFQ scheduler can significantly improve the performance for cell edge users in terms of throughput and packets dropped

7.3 Recommendations for Future Research Works

During the course of this research, the following extensions as future research work have been found interesting:

- Future research may consider other service classes such as FTP, VoIP (voice over IP), HTTP etc instead of only video traffic.
- The ATBFQ algorithm may be studied in the context of call admission control (CAC) algorithm that anticipates the user application characteristics and mobility.
- Another potential topic for continuation of this work is to find an empirical formula to find the optimum ATBFQ parameters such as the debt limit, and the token generation rate based on the loading, interference conditions and the traffic service class type.
- It would also be interesting to study the performance and complexity of the modified ATBFQ algorithm in the uplink

References

- [1] Project Presentation, WINNER I deliverable D8.1, March 2004, <https://www.ist-winner.org>.
- [2] S. Verdú, "Wireless bandwidth in the making," *IEEE Commun. Magazine*, vol. 38, no. 7. July 2000, pp. 53-58.
- [3] C.E. Shannon, "A mathematical theory of communication," *Bells Systems Tech Journal*, vol. 27, pp. 623-656, July 1948.
- [4] I. Akyildiz, Y. Altunbasak, F. Fekri, and R. Sivakumar, "AdaptNet: An adaptive protocol suite for the next generation wireless Internet," *IEEE Commun. Magazine*, vol. 42, pp. 128–136, March 2004.
- [5] S. Shakkottai, T. S. Rappaport, and P. C. Karlsson, "Cross-layer design for wireless networks," *IEEE Commun. Magazine*, vol. 41, pp. 74–80, Oct. 2003.
- [6] A. J. Demers, S. Keshav, and S. Shenker, "Analysis and simulation of a fair queueing algorithm," in *Proc. ACM Symposium on Communications Architectures and Protocols (SIGCOMM'89)*, pp. 1–12, Austin, TX, USA, Sept. 1989.
- [7] A. K. Parekh and R. G. Gallager, "A generalized processor sharing approach to flow control in integrated services networks: The single-node case," *IEEE/ACM Trans. on Networking*, vol. 1, pp. 344–357, Mar. 1993.
- [8] A. Stamoulis and G. B. Giannakis, "Packet fair queueing scheduling based on multirate multipath-transparent CDMA for wireless networks," in *Proc. IEEE Joint Conference of the Computer and Communications Societies (INFOCOM'00)*, vol. 3, pp. 1067–1076, Tel-Aviv, Israel, Mar. 2000.
- [9] Q. Liu, X. Wang, and G. B. Giannakis, "A cross-layer scheduling algorithm with QoS support in wireless networks," *IEEE Trans. on Vehicular Technology*, vol. 55, pp. 839–847, May 2006.

- [10] D. Wu and R. Negi, "Effective capacity: A wireless link model for support of quality of service," *IEEE Trans. on Wireless Communications*, vol. 2, pp. 630–643, July 2003.
- [11] S. Borst, "User-level performance of channel-aware scheduling algorithms in wireless data networks," in *Proc. IEEE Joint Conference of the Computer and Communications Societies (INFOCOM'03)*, vol. 1, pp. 321–331, San Francisco, CA, USA, Mar.-Apr. 2003.
- [12] M. Andrews, L. Qian, and A. Stolyar, "Optimal utility based multiuser throughput allocation subject to throughput constraints," in *Proc. IEEE Joint Conference of the IEEE Computer and Communications Societies (INFOCOM'05)*, vol. 4, pp. 2415–2424, Miami, FL, USA, Mar. 2005.
- [13] S. Borst and P. Whiting, "Dynamic channel-sensitive scheduling algorithms for wireless data throughput optimization," *IEEE Trans. on Vehicular Technology*, vol. 52, pp. 569–586, May 2003.
- [14] Y. Cao and V. O. K. Li, "Scheduling algorithms in broadband wireless networks," *Proc. IEEE*, vol. 89, pp. 76–87, Jan. 2001.
- [15] S. Shakkottai and A. L. Stolyar, "Scheduling for multiple flows sharing a time-varying channel: The exponential rule," *American Mathematical Society Translations*, vol. 207, 2002.
- [16] P. Liu, R. Berry, and M. L. Honig, "Delay-sensitive packet scheduling in wireless networks," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC'03)*, vol. 3, pp. 1627–1632, New Orleans, LA, USA, Mar. 2003.
- [17] D. Rajan, A. Sabharwal, and B. Aazhang, "Delay bounded packet scheduling of bursty traffic over wireless channels," *IEEE Trans. on Information Theory*, vol. 50, pp. 125–144, Jan. 2004.
- [18] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, "Providing quality of service over a shared wireless link," *IEEE Communications Magazine*, pp. 150–154, Feb. 2001.

- [19] A. Stamoulis, N. D. Sidiropoulos, and G. B. Giannakis, "Timevarying fair queueing scheduling for multicode CDMA based on dynamic programming," *IEEE Trans. on Wireless Communications*, vol. 3, pp. 512–523, Mar. 2004.
- [20] M. Motani and A. T. Hoang, "An instance of multiuser diversity in wireless networks," in *Proc. IEEE Int. Symposium on Information Theory (ISIT'02)*, pp. 443–448, Lausanne, Switzerland, June-July 2002.
- [21] R. A. Berry and R. G. Gallager, "Communication over fading channels with delay constraints," *IEEE Trans. on Information Theory*, vol. 48, pp. 1135–1149, May 2002.
- [22] W. K. Wong and V. C. M. Leung, "Scheduling for integrated services in next generation broadcast networks," in *Proc. IEEE Wireless Communication and Networking Conference (WCNC)*, New Orleans, LA, Sept. 1999.
- [23] M. Sidi., W.Z. Liu, I. Cidon. and I. Gopal, "Congestion control through input rate regulation," in *Proc. of GLOBECOM '89*, pp 1764-1768, Dallas. TX, 1989.
- [24] C. A. Cooper and K. I. Park, "Toward a broadband congestion control strategy." *IEEE Network Magazine*, pages 18-23. May 1990.
- [25] D.C. Lee, "Effects of leaky bucket parameters on the average queuing delay: Worst case analysis", in *Proc. IEEE Joint Conference of the Computer and Communications Societies (Infocom '94)*, pp. 482-489, Toronto, Canada, June 1994.
- [26] A.V. Aho, J.E. Hopcroft, J.D. Ullman, *The Design and Analysis of Computer Algorithms*. Addison-Wesley, 1974
- [27] *WINNER System Concept Description*, WINNER1 Deliverable D7.6, October 2005, <https://www.ist-winner.org>.
- [28] *Final report on identified RI key technologies, system concept, and their assessment*, WINNER1 Deliverable D2.10, November 2005, <https://www.ist-winner.org>.

- [29] T. Bonald, "A score-based opportunistic scheduler for fading radio channels", in *Proc. European Wireless Conference*, Barcelona, Feb. 2004.
- [30] K. Brueninghaus, D. Astely, "Link Performance Models for System Level Simulations of Broadband Radio Access Systems," *Proc. IEEE PIMRC*, September 2005
- [31] *Final report on link level and system level channel models*, WINNER Deliverable D5.4, November 2005, <https://www.ist-winner.org>.
- [32] European Comission: European cooperation on the field of scientific and technical research (EURO-COST 231): "Digital mobile radio towards future generation systems", Final report, <http://www.lx.it.pt/cost231/>, Bruxelles, 1999
- [33] *Interim channel models*, WINNER II deliverable D1.1.1, February 2007, <https://www.ist-winner.org>.
- [34] T.S. Rappaport, *Wireless Communications: Principles & Practice*, Prentice Hall, 1st Edition, 1996.
- [35] D.J. Young, "The Generation of Correlated Rayleigh Random Variates by Discrete Fourier Transform and Quality Measures for Random Variate Generation", M.Sc. Thesis, Queen's University, Kingston, Canada, 1997.
- [36] J. Zander, S-L. Kim et al. *Radio resource Management for Wireless Networks*, Artech House, 2001.
- [37] C.R. Baugh and Jun Huang, "Traffic model for 802.16 TG3 MAC/PHY simulations", *IEEE 802.16 work-in-progress document # 802.16.3c-01/30r1*, March 2001.
- [38] M.H. Ahmed, H. Yanikomeroglu, S. Mahmoud, and D. Falconer, "Scheduling multimedia traffic in interference-limited broadband wireless access networks", in *Proc. of 5th International Symposium on Wireless Personal Multimedia Communications (WPMC'02)*, Honolulu, Hawaii, USA, October 2002.

- [39] M.H. Ahmed, H. Yanikomeroglu, and S. Mahmoud, “Interference management using packet scheduling in broadband wireless access networks”, *Wiley Wireless Communications and Mobile Computing Journal*, November 2003.
- [40] WINNER II Test scenarios and calibration cases Issue 2, WINNER II Deliverable D6.13.7, December 2006, <https://www.ist-winner.org>.
- [41] R. Knopp and P.A. Humblet, “Information capacity and power control in single cell multiuser communications,” in *Proc. IEEE Int. Conference on Communications (ICC’95)*, pp. 331–335, Seattle, WA, USA , June 1995.
- [42] E. F. Chaponniere, P. J. Black, J. M. Holtzman, and D. N. C. Tse, “Transmitter directed code division multiple access system using path diversity to equitably maximize throughput.” U.S. Patent #6449490, Sept. 2002.
- [43] J. M. Holtzman, “Asymptotic analysis of proportional fair algorithm,” in *Proc. IEEE Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC’01)*, vol. 2, pp. F–33–F–37, San Diego, CA, USA, Sept. 2001.
- [44] H.J. Kushner, P.A. Whiting, “Asymptotic properties of Proportional Fair sharing algorithms”. *Proc. 40th Annual Allerton Conf. Commun. Control Comp.*, 2002.
- [45] W. K. Wong , H. Y. Tang, and V. C. M. Leung, “Token bank fair queuing: a new scheduling algorithm for wireless multimedia services”, *International Journal of Communication Systems*, vol. 17, pp. 591-614, 2004.
- [46] R. Jain, D. Chiu, and W. Hawe, “A quantitative measure of fairness and discrimination for resource allocation in shared computer systems,” DEC Research Report TR-301, Digital Equipment Corporation, Maynard, MA, USA, Sept. 1984.
- [47] H. Sirisena, A. Haider, M. Hassan, and K. Fawlikowski, “Transient fairness of optimized end-to-end window control,” in *Proc. IEEE Global Communications Conference (GLOBECOM’03)*, pp. 3979 – 3983, San Francisco, CA, USA, Dec. 2003.
- [48] G. Berger-Sabbatel, A. Duda, O. Gaudoin, M. Heusse, and F. Rousseau, “Fairness and its impact on delay in 802.11 networks,” in *Proc. of the IEEE Global Communications*

Conference (GLOBECOM'04), pp. 2967–2973, Dallas, TX, USA , Dec. 2004.

Appendix A

Orthogonal Frequency Division Multiple Access (OFDMA) is a multi-user version of the popular OFDM digital modulation scheme. Multiple access is achieved in OFDMA by assigning subsets of subcarriers to individual users. This allows simultaneous low data rate transmission from several users. Based on feedback information about the channel conditions, adaptive user-to-subcarrier assignment can be achieved. If the assignment is done sufficiently fast, this further improves the OFDM robustness to fast fading and cochannel interference, and makes it possible to achieve even better system spectral efficiency. Different number of sub-carriers can be assigned to different users, in view to support differentiated QoS, i.e. to control the data rate and error probability individually for each user. OFDMA can also be seen as an alternative to combining OFDM with time division multiple access (TDMA) or time-domain statistical multiplexing, i.e. packet mode communication.

OFDM Parameters

The OFDM parameters for WINNER are described below.

	Base Coverage Urban	Microcellular	Indoor
Subcarrier distance Δf	39062.5 Hz	48828.125 Hz	
Useful symbol duration T_N	25.6 μ s	20.48 μ s	
Guard interval T_G	3.2 μ s	2.00 μ s	
Total symbol duration	28.8 μ s	22.48 μ s	
used subcarriers	[-576:576] subcarrier 0 unused	[-920:920] subcarrier 0 unused	
Signal bandwidth	2 x 45 MHz	89.84 MHz	
System bandwidth	2 x 50 MHz	100.0 MHz	
FFT bandwidth, sampling rate	80.0 MHz	100.0 MHz	

Table 7-1OFDM/GMC parameters

The frame parameters are provided by the following table.

		Base Coverage Urban
Overall frame length		0.6912 ms
Number of OFDM symbols per frame		24
Chunk layer dimension in symbols x subcarriers		12 x 8 =96
Downlink	Dedicated pilot + control symbols per chunk layer in frequency adaptive transmission	4 + 12 = 16
	Dedicated pilot + control symbols per chunk layer in non-frequency adaptive transmission	8 + 18 = 26
Number of chunks per frame in time and frequency direction		2 x 144
Duplex guard time		0 μ s

Table 7-2 Frame parameters in WINNER

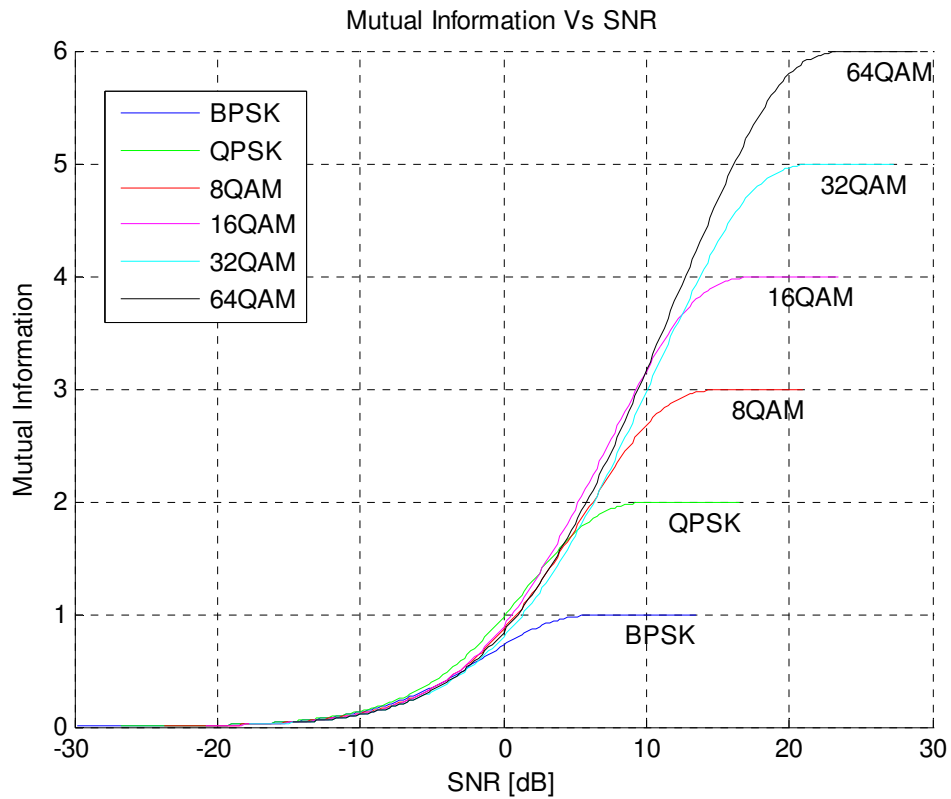
Appendix B

L2S steps (MIESM method)

Input: SINR per subcarrier

Output: BLER

1. Find all subcarriers $[1, \dots, N]$ carrying data symbols of a particular coded block.
2. Calculate the SINR for each subcarrier $n \in [1, \dots, N]$ i.e. SINR_n .
3. Find all $\text{MI}_n = f(\text{SINR}_n)$, $n \in [1, \dots, N]$ from the Mutual Information curve below (the data file producing this curve is available) considering the Modulation scheme being used:



4. Calculate $MI = \frac{1}{N} \sum_{n=1}^N MI_n$
5. Calculate $SINR_{\text{eff}} = f^I(MI)$ from the inverse of Mutual Information curve above.
6. Convert $SINR_{\text{eff}}$ to E_b/N_0 based on Modulation-Coding Scheme being used. To convert, simply use the following table:

Modulation	Coding rate	$E_b/N_0(\text{dB}) - SINR_{\text{eff}}(\text{dB})$
16 QAM	1/2	-3.01
64QAM	1/2	-4.77
BPSK	1/2	+3.01
16 QAM	2/3	-4.26
64QAM	2/3	-6.021
BPSK	2/3	+1.761
16 QAM	3/4	-4.7712
64QAM	3/4	-6.5321
BPSK	3/4	+1.249

7. Based on the size of the particular coded block, Convert E_b/N_0 to BLER using the set of curves such as the following (the data files producing these curves are all available). The BLER curves available have block size ranges from 288 up to 4608 information bits per block. For any other block size, we may use interpolation.

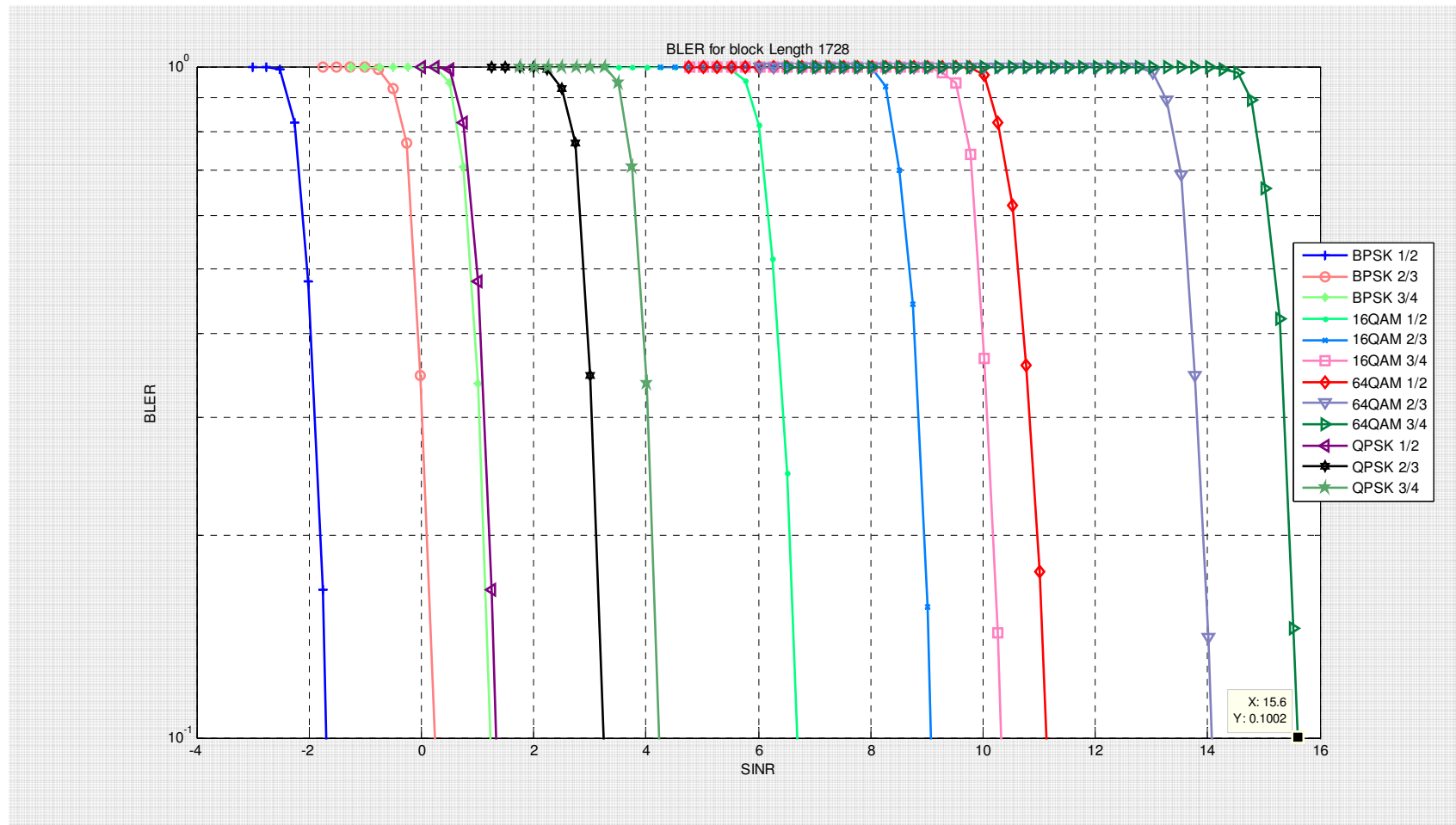


Figure 7-1 BLER vs SINR for Block length of 1728 bits

