RATE SCHEDULING FOR HSDPA IN UMTS

Farhan Hameed

Master Thesis

Computer Engineering

Reg # E 3584 D

DEGREE PROJECT

In Computer Engineering

Programme International Masters in Computer Engineering	Reg. No:	Extent 30 ECTS
Name of student Farhan Hameed	Year-Month-Day	
Supervisor Ernst Nordström	Examiner Prof. Mark Dougherty	
Company/Department Department of Economics and Social Sciences, Dalarna University, Sweden	Supervisor at Department Ernst Nordström	
Title Rate scheduling for HSDPA in UMTS		

Högskolan Dalarna
Röda vägen 3,78188
Borlänge

Phone: +46 (0)23778000
Fax: +46 (0)23778050
URL: http://www.du.se

Abstract

The introduction of a new technology High Speed Downlink Packet Access (HSDPA) in the Release 5 of the 3GPP specifications raises the question about its performance capabilities. HSDPA is a promising technology which gives theoretical rates up to 14.4 Mbits. The main objective of this thesis is to discuss the system level performance of HSDPA

Mainly the thesis exploration focuses on the Packet Scheduler because it is the central entity of the HSDPA design. Due to its function, the Packet Scheduler has a direct impact on the HSDPA system performance. Similarly, it also determines the end user performance, and more specifically the relative performance between the users in the cell.

The thesis analyzes several Packet Scheduling algorithms that can optimize the trade-off between system capacity and end user performance for the traffic classes targeted in this thesis.

The performance evaluation of the algorithms in the HSDPA system are carried out under computer aided simulations that are assessed under realistic conditions to predict the results as precise on the algorithms efficiency. The simulation of the HSDPA system and the algorithms are coded in C/C++ language.

Abbervations

2G – Second Generation

3G – Third Generation

3GPP – Third Generation Partnership Project

4G – Fourth Generation

8-PSK – Octagonal Phase Shift Keying

ANSI – American National Standards Institute

bps – bits per second

BSC – Base Station Controller

BTS – Base Transceiving Station

C/I – Carrier to Interference Ratio

CDMA – Code Division Multiple Access

dB – Decibel

EDGE – Enhanced Data Rates for GSM Evolution

EGPRS – Enhanced General Packet Radio Service

ERP – Enterprise Resource Planning

FDD – Frequency Division Duplex

FTP – File Transfer Protocol

Gbps – Gigabits Per Second

GGSN – Gateway GPRS Support Node

GHz — Gigahertz

GPRS – General Packet Radio Service

HARQ – Hybrid Automatic Repeat Request

HSDPA – High Speed Downlink Packet Access

HS-PDSCH - High Speed Physical Downlink Shared Channels

HSPA – High Speed Packet Access (HSDPA with HSUPA)

HSPA+ – HSPA Evolution

HSUPA – High Speed Uplink Packet Access

IEEE – Institute of Electrical and Electronic Engineers

RAB – Radio Access Bearer

RAN – Radio Access Network

RF – Radio Frequency

RNC – Radio Network Controller

SGSN – Serving GPRS Support Node

SMS – Short Message Service

SNR – Signal to Noise Ratio

TDMA – Time Division Multiple Access

IP – Internet Protocol

IR – Incremental Redundancy

ISP – Internet Service Provider

ITU – International Telecommunications Union

kHz — Kilohertz

MAC – Medium Access Control

Mcps – Megachips Per Second

MCS – Modulation and Coding Scheme

MHz – Megahertz

MIMO – Multiple Input Multiple Output

MSC – Mobile Switching Center

OFDM – Orthogonal Frequency Division Multiplexing

PHY – Physical Layer

PDN – Packet Data Network

PDU - Protocol Data Unit

QAM – Quadrature Amplitude Modulation

TD-CDMA – Time Division Code Division Multiple Access

TIA/EIA – Telecommunications Industry Association/Electronics Industry Association

TTI – Transmission Time Interval

UMTS – Universal Mobile Telecommunications System

UTRAN – UMTS Terrestrial Radio Access Network

VPN – Virtual Private Network

WCDMA – Wideband CDMA

WiMAX – Worldwide Interoperability for Microwave Access

Table of Contents

1. INTRODUCTION

THESIS ENVIROMENT

This thesis is part of a research project entitled Traffic engineering in Future Internet Domains (TEFID) at Department of Economics and Social Sciences at Dalarna University. The simulated algorithms for investigation in this thesis are developed using C programming language under WIN XP/Linux based environment. The results are shown in the form of line graphs produced by Microsoft Excel.

BACKGROUND

Mobile networks have seen tremendous development in the last few decades starting from the first generation up to the evolution of the fourth generation. The cellular networks are set apart in categories from each other by the word generation. Each of these generations is distinct from the other based on the capacities and services they provide.
This thesis is related to the third generation mobile systems technologies. Particularly deals with the UMTS network. A brief introduction to the UMTS networks is provided.

UMTS networks

UMTS is a step into the third generation mobile networks. It deals with ever increasing demand for higher data rates for mobile and internet applications in the mobile communication world.
UMTS which is also referred as WCDMA is foreseen as the successor to GSM technology. Because GSM was so successfully implemented in Europe and worldwide the UMTS Core network was based on the evolved Core network of GSM. This could be seen in the first release of UMTS (3GPP Release 99), also the UMTS core network is supposed to support both UMTS (UTRAN) and GSM (GSM BSS) radio access networks.

Network Architecture

In UMTS network architecture, the major difference from the previous GSM evolved GPRS network is the introduction of the UTRAN (the UMTS Radio Access Network). This employs the CDMA technology for the air interface referred as Wideband-CDMA. This change basically facilitates in the transmission of voice, video and data services on the same network. The Core Network (CN) remained unchanged, but with some upgrades in software to adjust for the UMTS upgrading.

In UMTS the mobile equipment known as User Equipment (UE), is connected to the NodeB over interface “Uu”. NodeB also known as WCDMA Base Station (WBTS) is the termination point between the transmission network (UTRAN) and the air interface. It is a network entity that supports a single cell, or if in sectored sites could cover more then one cell. NodeB is responsible to provide all the required signal processing functions to support the WCDMA air interface and this is where most of the complexity arises. The NodeBs are the equivalent of BTS in GSM.
Several WBTSs are handled by a single Radio network controller (RNC) over interface “Iub”.

RNC is the nucleus of the new access network (UTRAN); it is the replacement of BSC in GSM. Network operation judgments are undertaken at this controller; to facilitate in its work it has a high speed packet switch at its center that can support a reasonable throughput of traffic. The RNC is connected to the Core Network (CN) through the interface “Iu”. One feature not found in previous GSM networks is the capability of supporting interconnections between two RNCs, this is made possible by the introduction of interface “Iur”. This enables the RNC to be fully aware and handle the Radio Resource Management (RRM) all by it self, eliminating the burden from the Core network. Most of the decision making process is software based, which is expected to have a high processing capacity.

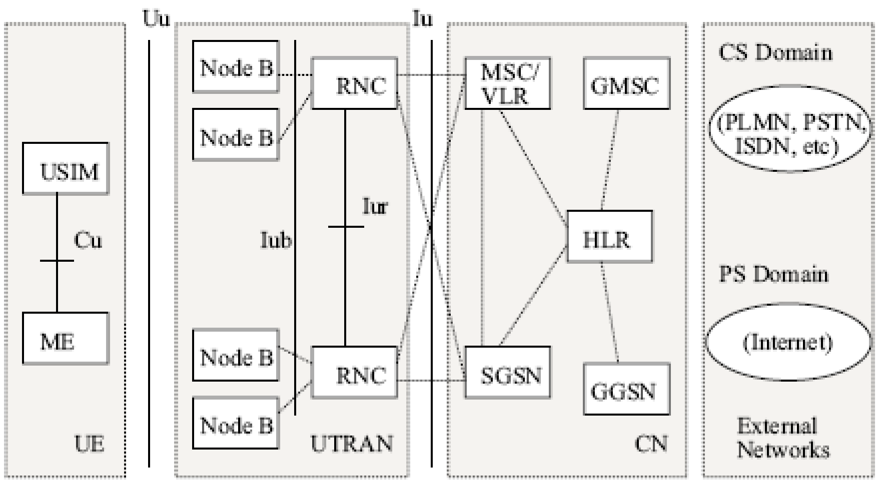


Figure 1: UMTS Architecture [1]

In the first UMTS release R99 mostly the Core network was not touched in regard to the introduction of the UTRAN from the previous 2G evolved GSM CN, except for software modifications and upgrading were implemented to support the new Access network (UTRAN). While in the later releases R4 and more there were recommendations also for the alteration of the CN for the bearing of some features.

Air interface

The W–CDMA technology in the Late 90’s was chosen to be the multiple access technique for the third-generation mobile telephone system in Europe. In other words W–CDMA was chosen as the air interface for UTRAN. The term WCDMA also refers to one of the International Telecommunications Union's IMT-2000 leading standards for 3G cellular network.
W-CDMA has been developed into a complete set of specifications, a detailed protocol that defines how a mobile phone communicates with the tower, how signals are modulated, how datagrams are structured, and system interfaces are specified allowing free competition on technology elements.

The key operational features of the W-CDMA radio interface are summarized below [2]

- Radio channels are 5MHz wide.
- Chip rate of 3.84 Mcps
- Supports two basic modes of duplex, frequency division and time division. Current systems use frequency division, one frequency for uplink and one for downlink. For time division, FOMA uses sixteen slots per radio frame, where as UMTS uses 15 slots per radio frame.
- Employs coherent detection on uplink and downlink based on the use of pilot symbols.
- Supports inter-cell asynchronous operation.
- Variable mission on a 10 ms frame basis.
- Multicode transmission.
- Adaptive power control based on SIR (Signal-to-Interference Ratio).
- Multiuser detection and smart antennas can be used to increase capacity and coverage.
- Multiple types of handoff between different cells including soft handoff, softer handoff and hard handoff.

WCDMA Logical Channels

Three categories of channels have been defined in UMTS in order to keep effective control multiplexing and de-multiplexing: logical channels, transport channels and physical channels WCDMA basically follows the ITU Recommendation M.1035 in the definition of logical channels.
Some examples of these three types of channels are given below.

- Logical Channels: Common Control Channel (CCCH), Dedicated Control Channel (DCCH), Common Traffic Channel (CTCH), Dedicated Traffic Channel (DTCH).
- Transport Channels: Forward Access Channel (FACH), Random Access Channel (RACH), Dedicated Channel (DCH), Broadcast Channel (BCH), Downlink Share Channel (DSCH).
- Physical Channels: Dedicated Physicals Data Channel (DPDCH), Dedicated Physical Control Channel (DPCCH), Physical Random Access Channel (PRACH).

The transport channel and the logical channels exist between the UE and the RNC via the Node B, whereas the physical channels only exist between the UE and the Node B. Further information about the channels of UMTS can be found in [3]
WCDMA and CDMA2000 systems do support packet data but the design attitude still primal in a way that the system resources such as power, code and data rate are optimized to voice services.
Since late 99 system designers realized that the main wireless data applications would be Internet protocol (IP) related, thus optimum packet data performance was the primary goal for the system designers to accomplish. With the design philosophy change, some new technologies appeared such as adaptive modulation and coding, hybrid ARQ, fast scheduling etc. which were all in cooperated in Release 5 of WCDMA named as High Speed Downlink Packet Access (HSDPA) which shall be discussed in detail in the third chapter.

THESIS OBJECTIVE

The introduction of a new technology such as HSDPA in the Release 5 of the 3GPP specifications raises the question about its performance capabilities.
The main objective of this thesis is to discuss the system level performance of HSDPA

Mainly the thesis investigation will concentrate on the Packet Scheduler because it is the central entity of the HSDPA design. Due to its function, the Packet Scheduler has a direct impact on the HSDPA system performance. Similarly, it also determines the end user performance, and more specifically the relative performance between the users in the cell. The thesis analyzes several Packet Scheduling algorithms that can optimize the trade-off between system capacity and end user performance for the traffic classes targeted in this thesis such as Streaming (Multimedia), Interactive/Background (data).
The performance evaluation of the algorithms in the HSDPA system are carried out under computer aided simulations that are assessed under realistic conditions to predict the results as precise on the algorithms efficiency.

LIMITATIONS

- Only one user to be scheduled or served in one time slot.
- The User to be scheduled is assumed to be at a stationary position.
- The simulation is configured for a smaller version (scale) of the realistic network due to huge computational times.
- The scheduling schemes are simulated to work in a centralized manner at the node B.
- The numbers of fading channels are quantized into five states so as to avoid complexity in computation.

OUTLINE OF DISSERTATION

The thesis report is organized as follows:

Chapter 1: Gives a short introduction and outlines the objectives of this Master thesis.

Chapter 2: presents the problem description of the most relevant QoS attributes of the network under study, which allows identifying the QoS demands imposed on the conveying networks. The chapter also brings up the questions to which the thesis gives answers.

Chapter 3: provides a general overview of the HSDPA technology that is required to achieve a full comprehension of the HSDPA investigations carried out in this Master thesis. This chapter also provides an overview on the Packet Scheduling entity of HSDPA, which further on in the following chapters are used to better understand this aspect.

Chapter 4 & 5 : these chapters' gives a description of the Traffic model and the channel model, the ON/OFF Model and the Finite State Markov Model respectively, used in the simulation of the network and details of implementation of the models.

Chapter 6 : Describes in the chapter the scheduling techniques chosen for analysis of the HSDPA network, along with the simulation of these techniques and their performance. It also provides with the performance parameters used for the analysis of the scheduling techniques.

Chapter 7 : deals with the simulation mainly the detailed description of the simulation system on the whole, along with the experiments scenarios and their results and the analysis of the results.

.

Chapter 8: draws the main conclusions of this Ph.D. investigation and discusses future research topics.

The references followed with the the simulator Code are included as well.
The next chapter continues with the thesis.

2. PROBLEM DEFINITATION

Problem Statement

Our problem involves a scheduler (Base station) placed in a cell with different users scattered randomly in the area around the scheduler. On the downlink each users wishing to transmit data from a single base station to many mobile destinations

In the network assuming characteristics such as

- Number of users to be served or scheduled is $N \ ,i \in \{1,2,3,...N\}$
- A list of Modulation and Coding scheme $M_j, j \in \{1,2,3,..M\}$
- Each user having variable channel condition at different time slots C_k^i where 'i' is the particular user and 'k' being the particular discrete time slot. $C_k = \{C_k^1, C_k^2, C_k^2, ...C_k^N\}$, Where Ck is the set of all channel states

Due to limited resources the users 'i' competes for the radio resource at each time slot 'k'. One user scheduled per one Transmission time interval (Time slot).

Scheduling Decision

The User with better channel conditions is scheduled, in order to maximize the overall network throughput.

$$i_k = \Phi(C_k)$$

Furthermore the scheduled user is assigned a modulation and Coding scheme 'j' that would optimize the data rate for its channel conditions at that time slot. $\psi(C_k^i) = j$

Quality Consideration

Keeping the Quality standards QoS in tact while scheduling is a problem, as the users with not so good channels might constantly get starved.

The scheduler should function in a way as to keeping the average throughput '$S_i(k)$' of a user 'i' up till time 'k' above a minimum specified threshold 'D'. i.e $S_i(k) \geq D$.

In Order to capture the Quality standard, the system efficiency and fairness between users should be balanced. This could be implemented by providing higher priority to the users with low performance.

Phone: +46 (0)23778000
Fax: +46 (0)23778050
URL: http://www.du.se

QUESTIONS for INVESTIGATION

Answers to the following questions shall be given in this thesis

1. Under what circumstances the theoretical optimal throughput of 14.4 Mbps value is obtained?
2. How close to the optimum do the 3 algorithms get?
3. How do the scheduling algorithms compare in performance to each other?
4. Describe the complexity of the scheduling algorithms studied?
5. What scheduling techniques performed better then the others, in different conditions?
6. The scheduling algorithm that give the best fairness in comparison to the others?
7. Mention the commonalities between the three algorithms used?

The next chapter describes HSDPA.

3. HSDPA

As discussed in the introductory chapter, the recent third generation standardization and related technology development reveal the need of the high-speed packet data of wireless internet. The WCDMA system in the Release 99 does fulfill the general requirements of voice and data services by providing data transmission rates up to 2 Mbps.
With the introduction of Release-5 of the specifications in the spring of 2002, WCDMA packet data support was further enhanced to provide peak data rates in the order of 10 Mbps together with lower round-trip delays and increased capacity provide a further boost for wireless data access.
HSDPA can give a theoretical maximum channel rate of 14.4 MBits this should be possible with a channel with no fading. In this case 4/4, 16 QAM, and 15 codes can be used. In a real network, fading exists. This means that the channel can be a state where the channel capacity is less than maximum.

Introduction

The UMTS Release-5 encloses a new set of features known collectively as HSDPA. A new transport channel targeting packet data transmissions is introduced in the release-5, the high speed DSCH (HS-DSCH), which can be seen as a continued evolution of the DSCH transport channel. The HS-DSCH channel supports three principles: fast link adaptation, Hybrid ARQ (HARQ), and fast scheduling which help to achieve the requirements of shorter delay and high throughput,
These three principles rely on rapid adaptation to changing radio conditions or in other words faster link adaptation; hence the corresponding functionality is placed in the Node B instead of the RNC for quick response.

Architecture of HSDPA System

HSDPA uses the same network infrastructure as that of the WCDMA/UMTS discussed earlier in the introductory chapter. In order to accommodate the new features and high data rate capabilities HSDPA provides, a new medium access layer called MAC-hs introduced in the Node B. Moreover some additional control channels have also been introduced to achieve the desired functionality.

MAC-hs

A specialized MAC high speed (MAC-hs) entity with enhanced control functionalities has been set-up on top of the physical layer in both, UE and Node B. This layer provides HARQ mechanisms and fast scheduling, facilitating the efficient usage of the radio resources in adaptation to the instantaneous channel conditions and network load.
The new relocated MAC-hs layer at the Node B facilitates fast scheduling by avoiding the latency involved when MAC-hs is placed at the RNC
The modified protocol architecture [4] effecting different protocols layers is show in the figure below.

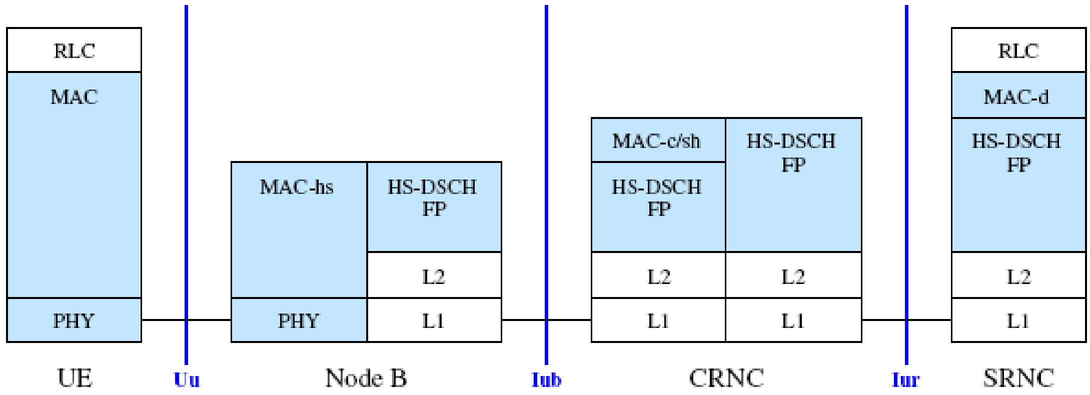


Figure 2 HSDPA protocol architecture, modified parts highlighted

HSDPA Channel structure

To implement the HSDPA features, three new channels are introduced in the physical layer specification.

High-speed Downlink Shared Channel (HS-DSCH)

HS-DSCH carries the user data in the downlink direction, with the peak rate reaching up to 10 Mbps range. It is easy to understand that HS-DSCH can only be applied on packet switch domain, for HSDPA is a packet-based data service.
HS-DSCH has specific characteristics some of them are listed below.

- Reduced Delay: The TTI has been defined to be 2ms (three slots) to achieve a short round trip delay for operations between the terminal and Node B for retransmissions.
- Higher Peak Data-Rate: Adding a higher order modulation scheme, 16 QAM, as well as lower encoding redundancy has increased the instantaneous peak data rate.
- Higher Capacity: with the utilization of 16 QAM modulation along with the already in use QPSK modulation in previous releases allows higher capacity up to 10 Mbps

Also the other two Channels introduced are defined below

- High-speed Shared Control Channel (HS-SCCH) carries the necessary physical layer control information to enable decoding of the data on HS-DSCH and to perform the possible physical layer combining of the data sent on HS-DSCH in the case of retransmission of an erroneous packet.
- Uplink High-Speed Dedicated Physical Control Channel (HS-DPCCH) carries the necessary control information in the uplink, namely, ARQ acknowledgements (both positive and negative ones) and downlink quality feedback information.

Adaptive modulation and coding (AMC)

As discussed in [5] [6], the benefits of adapting a wireless system especially a CDMA based system, to the changing channel conditions are well known. Techniques such as fast power control found in WCDMA were disadvantageous in a sense that intercellular interference over the downlink increased.

The principle of AMC is to change the modulation and coding format (transport format) in accordance with instantaneous variations in the channel conditions, subject to system restrictions. AMC extends the systems ability to adapt to good channel conditions. Channel conditions should be estimated by feedback from the receiver.
For a system with AMC, user in favorable position or experiencing "up-fade" typically would be assigned higher order modulation with higher code rate (e.g. 64 QAM with r = ¾ turbo codes). On the other hand, user close to cell boundary, are assigned lower order modulation with lower code rates (e.g. QPSK with r = ½ turbo codes). This shifts the picture to rate control rather then power control for wireless data. Further detailed explanation for the selection of the modulation and coding rate at each transmission frame are discussed in [7].

Hybrid Automatic Repeat reQuest (HARQ)

In the Link adaptation process, AMC suffers degradation. This is because Firstly AMC provides limited precision in data rate selection, i.e. the channel quality often estimates a data rate between two subsequent MCSs. Second the channel quality it self can be estimated with some probabilities of error, due to the difference between time of measurement and the time of rate selection and also due to measurement errors.
The HARQ technique here helps to adjust the coding rates precisely, and thus improves the link adaptation accuracy and the efficiency of the channel utilization
In HARQ scheme, the corrupted packet is not discarded but stored in the buffer of the receiver instead. When the retransmitted packet is received, it will be combined with the previous transmission of the same information bits, this process is called soft combining. The combined signal is then put to decode, if again fail in decoding, further retransmissions (up to a preset number defined by the system) will occur and is soft combined until the packet is decoded successfully.
The soft combining process of HARQ increases the possibility of a successful decoding of the information bits, therefore increases the transmit efficiency.
There are two types of HARQ schemes defined in the 3GPP specifications: namely Incremental Redundancy and Chase Combining

Packet Scheduling

Packet Scheduling aims at maximizing system throughput while satisfying the QoS requirements of users. The scheduler exploits the multi-user diversity to increase the system throughput. This idea is based on the fact that good channel conditions allow for higher data rates by using a higher-order modulation and coding schemes. Scheduling is applied mainly based on channel conditions to exploit AMC and HARQ to their maximum potential, and should also concern the amount of data waiting for transit and the priorities of services at the same time.

HSDPA Packet Scheduler Process

At every TTI every UE sends a report Channel Quality Indicator (CQI) to Node-B. The CQI contains information about the instantaneous channel quality of the user; the report also mentions in it the MCS and channel codes UE expects. The user (UE) is able to measure its current channel conditions by measuring the power of the received signal from the Node B. Therefore, users with good channel conditions enjoy potentially higher supportable data rates by using higher modulation and coding rates, whereas users with bad channel conditions will experience lower data rates instead of adjusting their transmission power.

Scheduling Algorithms in HSDPA

The pace of the scheduling process divides the packet scheduling methods into two main groups namely Fast Scheduling method and Slow Scheduling methods.

Slow Scheduling Methods

Scheduling algorithms that base their scheduling decisions on the average user's signal quality (or that do not use any user's performance metric at all).
Slow scheduling methods comprise the following algorithms:

- Average C/I (Avg. CI): This scheduling algorithm serves in every TTI the user with largest average C/I with backlogged data to be transmitted. The default averaging window length for the average C/I computation is usually 100ms.
- Round Robin (RR): In this scheme, the users are served in a cyclic order ignoring the channel quality conditions. This method outstands.

Fast Scheduling Algorithm

Scheduling algorithms utilizing the channel conditions of users need to make decisions every TTI to better exploit fast variation of channel conditions and are therefore called fast scheduling algorithms.
Since real-time applications have different QoS constraints than non-real-time applications, the design of scheduling algorithms for real-time applications should be different from that for non-real-time applications. Therefore, scheduling algorithms can be classified into two groups:

Non-Real-time (NRT) methods:

NRT applications do not require strict QoS guarantees, as these applications are suited for data traffic (i.e., interactive and background). The time shared nature of the HSDPA channels design are very well suited for these algorithms.

- *Maximum Carrier-to-Interface Ratio (Max CIR)* : This algorithm [8] tends to maximize the system throughput by serving, in every TTI, the user with the best channel quality). It can be seen that this algorithm provides high system throughput since only those with high current supportable data rates get served. However, this algorithm has an obvious drawback in that it ignores those users with bad channel conditions, which may lead to starvation.
- *Proportional Fairness (PF):* The PF algorithm [9] tries to increase the degree of fairness among users by selecting those with the largest relative channel quality. Relative channel quality is the instantaneous channel quality condition of the user divided by its current average throughput. Therefore, this algorithm considers not only those users with good channel conditions but also those with low average throughputs by giving them higher priority.

Real-time (RT) methods:

Streaming applications impose strict constraints on the network in order to satisfy their QoS requirements.

RT Packet scheduling algorithm tend to be quite complicated as these must be able to guarantee QoS requirements for streaming users as well as exploiting information about their instantaneous channel conditions in its scheduling decisions. Guaranteeing the QoS requirements of streaming users is a challenging task, especially when the traffic load in the cell is high.

- *Opportunistic Algorithm*: opportunistic algorithm for scheduling HSDPA users is a RT Scheduling algorithm. It works by selecting modulation/coding and multi-code schemes that exploit channel and buffer variations to increase the probability of uninterrupted media play-out. The scheduling problem of providing uninterrupted play-out is transformed to a feasibility problem that considers two sets of stochastic Quality-of-Service (QoS) constraints: stability constraints and robustness constraints.

In this thesis the performance of the three Fast scheduling methods used in HSDPA is tested and compared in chapter 6.
The next chapter introduces the Traffic model and the simulation to the traffic model used in the thesis.

4. TRAFFIC MODEL

A data network like HSDPA has different characteristics from a traditional voice network in many aspects. In data network, the traffic volume for downlink is much higher than that for uplink. Also, there are different kinds of services such as HTTP, WAP, VoIP, real time multimedia traffic, and so on, which have their own requirements of delay and loss rate. Data traffic is bursty on the whole.

Performance of a network requires excellent traffic models that have the ability to capture the statistical characteristics of the actual traffic on the network.
The model used here in the experiment for the analysis of the traffic is the ON-OFF source model which shows the characteristics of bursty data traffic.

ON/OFF SOURCE MODEL

To model the arrival of the network traffic consider the following

- N different ON-OFF sources.
- The sources are statistically identical and independent.
- Each of the sources is in one of the two states, ON state or OFF state.
- In ON state the source generates traffic, while it is silent in the OFF state.
- The time between the two states, the transition time is expected to follow exponential distribution.

The Queue of Size M Mbits is shared by the N Sources served by a constant rate C Mbps.

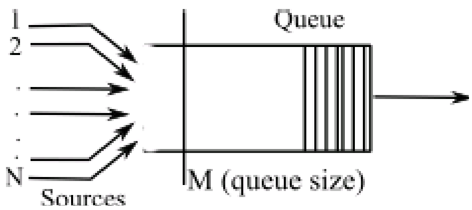


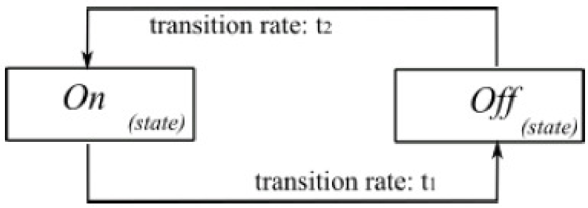
Fig:

Simulation Model

The traffic model is described by the four parameters below

- Number of Sources N.
- Transition probability from state ON to OFF state ´$t_1 = \frac{1}{t_{ON}}$ [s]` , where ´$t_{ON}$` is the average time spent in ON state

- Transition probability from state OFF to ON state ´ $t_2 = \frac{1}{t_{OFF}}$ [s] ` , where ´ $t_{OFF}$ ` is the average time spent in OFF state
- Peak rate in the ON state R[Mbps]



The number of users (sources) range from 1 to 10 in the experiments. At each point in time the users are in one of the two states Either ON or OFF state. The On state representing burst of data, while the OFF state means no data burst.
The peak rate R[Mbps] of the data burst in the ON state, depends on type of source it is (i.e. voice or multimedia).
The total time spent in the ON state is known as the ON period, similarly the time by a source in the OFF state is known as the OFF period.
The source hence can be modeled by a two state irreducible continuous Markov chain

$$X(t), t > 0$$

The time to the next state ON/OFF are exponentially distributed, and show the property of memory-less ness. The time is simulated by the help of a RAND function which produces a Generator that depends on the type of state the source is associated with. The expression is given by the equation

$$time\_to\_next\_state = -\log(RAND)\frac{-\log(RAND)}{transition\_rate}$$

Where, transition rate is either t1 or t2.

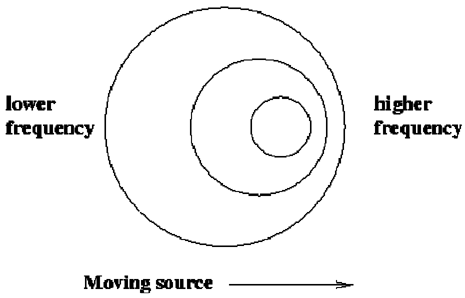
The next chapter gives details on the Channel model used in the simulations.

5. CHANNEL MODEL

The transmitter in the wireless network produces the signal and sends it over the propagation channel towards the receiver. The signal that emerges from the channel is corrupted (i.e. Fading), but it does contain the transmitted signal. Communication system design begins with detailing the channel model
Simulation of the traffic channel is modeled and used to anticipate the behavior of a propagation channel and check out how the channel affects the transmitted signals in the experiments or simulations.

FADING

Fading as the name suggests refers to the distortion that a carrier-modulated telecommunication signal experiences over certain propagation media10.
Two factors contributing to signal fading, multipath fading: is where superposition of multiple copies of the signal are seen by the receiver, due to the reflectors (obstacles) present in the path of the signal from transmitter to receiver. This would result in either destructive or constructive interference in the overall signal power.
The second factor is Doppler Effect: the user's movement towards or away from the base station causes a shift in the frequency of the signal transmitted along each signal path. This corresponds to different rates of change in phase.



Two types of Doppler Effect Slow vs. Fast fading, slow fading is found when the signal shows correlated behaviors in the change of the fading magnitude over a period of time, while in fast fading there is not any correlation found. Here the amplitude and phase change imposed by the channel varies considerably over the period of use.

And also multipath fading can be characterized in two types Flat vs. Frequency-selective; Flat fading has the characteristics of experiencing correlated fading on all frequencies of the signals. While frequency selective fading as the name proposes shows uncorrelated fading behavior for the spectral components of the transmitted signal

MARKOV MODEL for FLAT FADING

Unlike WCDMA R99, the transmission power is fixed and SNR is directly used to measure the channel quality and capacity at the receiver, here the fading process can be seen as the process that controls the transmission capacity of the system, i.e. as the amount of fading increases the available capacity decreases. Hence, each Fading state (Channel state) is associated with a capacity value (Modulation and coding scheme).

The system is modeled as Rayleigh model (Flat fading) which is used to model multipath Fading with no direct line-of-sight. The received channel fading amplitude γ in Rayleigh Fading is distributed exponentially with PDF,

$$P(\gamma) = \frac{1}{\gamma_0} \exp\left(-\frac{\gamma}{\gamma_0}\right), \qquad \gamma \geq 0,$$

where, γ_0 is the average SNR.

The Rayleigh distribution system can be modeled by ´m´ Finite-State Markov Channel (FSMC). The state space of a first order Markov chain represented by $S = s_1, s_2, s_3, \ldots, s_m$. The state space 'S' is that of 'm' different channel states with corresponding Signal to noise ratio (SNR).
These discrete SNR thresholds of the network have been obtained by partitioning of the SNR into finite number of intervals, in increasing order represented by $\lambda = [\lambda_1, \lambda_2, \ldots, \lambda_k]$, where $\lambda_1 = 0$ and $\lambda_k = \infty$. The figure shows 'm' different states of the Markov chain.

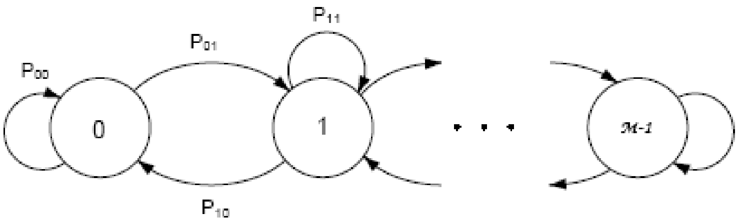


Fig:

The state of the Markov chain can be determined by the transition probabilities 'P_{jk}'; the transition from one state to the other is independent from the previous occurring states.

First-order Markov chain can be defined by its transition probability matrix [11]

$$T = \{P_{j,k}\}, i, j \in 0,1,\ldots,M-1$$

where

$$P_{jk} = P(S_{m+1} = S_k \mid S_m = S_j)$$

Depending on the expected SNR state, different modulation and error-correcting coding rates can be dynamically selected from a set of Modulation and Coding Schemes (MCS). The higher the order of the MCS selected the higher the transmission rate. The SNR is mapped directly into MCS and hence into data rate.

Channel SIMULATION

In the simulation of the Channel model at every next TTI a new Channel state is calculated for each of the N users. This calculation is based on the current channel state and the transition probability matrix for each particular user.
This is a simulation of the Markov chain of finite state, where M number of channel states that are produced as a result of sampling and quantization of the SNR
The following expression is used in the computation of the new channel state for each of the user i.

$$new\_state = \left\langle \begin{array}{ll} state_i & sum + P_{ij} > RAND \vee RAND \geq sum \\ same\_state & else \end{array} \right.$$

where, j number of channel states
P_{ij} is the value from the probability matrix
sum, $sum = sum + P_{ij}$
[9]

The FSMM [12] has done crucial assumption that the state transition can be done only to the adjacent states. It has been seen that the first order model fails to adequately model the autocorrelation function of Gaussian based model as fading becomes faster.
The next chapter focuses on the Scheduling techniques.

6. PACKET SCHEDULING

Wireless data networks such as UMTS HSDPA use downlink scheduling that exploits channel fading to increase the system throughput. As future wireless networks more and more shift towards supporting multimedia and data traffic together, a proper criterion is needed for scheduling that can count various service requirements such as delay, overflow and packet loss.

A good devised scheduling algorithm along with taking into account maximization of the system throughput, should as well keep track about being fair to users. That is, scheduling algorithms should balance the trade-off between maximizing throughput and fairness.

Scheduling Algorithm

Scheduling plays a vital role in the performance of the Network System. Packet scheduling is one of the key design features of HSDPA. A packet scheduler controls the allocation of channels to users within the coverage area of the system by deciding which user should transmit during a given time interval. Based on this feature the system can increase its throughput to a maximum. In this thesis simulation three scheduling algorithms have been used which analyze the HSDPA system capacity each one of which is discussed below.

Opportunistic Scheduling Algorithm

Opportunistic algorithm is a Real-Time algorithm that is used for scheduling of HSDPA users. The algorithm in scheduling of users tends to satisfy the QoS requirements formulated for streaming data in HSDPA system. These QoS constraints have been derived from a discrete-event stochastic model (based on key features of HSDPA system). The quality constraints are presented as a feasible problem for which. The solution to which is given as a practical joint opportunistic user-scheduling and MCM assignment policy

This Opportunistic algorithm exploits channel and buffer variations to increase the probability of uninterrupted media play-out.

Background and Definitions

Discrete event model [13] for a HSDPA system is used here, with these main characteristics:

- N number of Users, each having Channel quality 'c_k' at time slot k.
- Set of modulation and error-correcting coding schemes.' m_k' at time k. Set of spreading codes represented as 'n_k'. These modulation and error-correcting coding schemes are used in the link adaptation process.
- Set of data transfer rate established for users at time k represented by 'r_k', the values of which are dependent on 'm_k' and 'n_k' .
- 'f_i^k' is the instantaneous FER at time 'k' for the user 'i '.
- ' D_i ' is the discharge rate from the UE buffer (also know as play-out rate). '$\overline{D_i}$ 'is the arriving date rate to the BS buffer from the server.

- λ_k 'is the discharge rate from the BS buffer at time slot 'k' ' next, 'V_k^i' is the number of bits in the BS buffer for user 'i ' at time k slot. While 'U_k^i' is the number of bits in the User Buffer at time k.

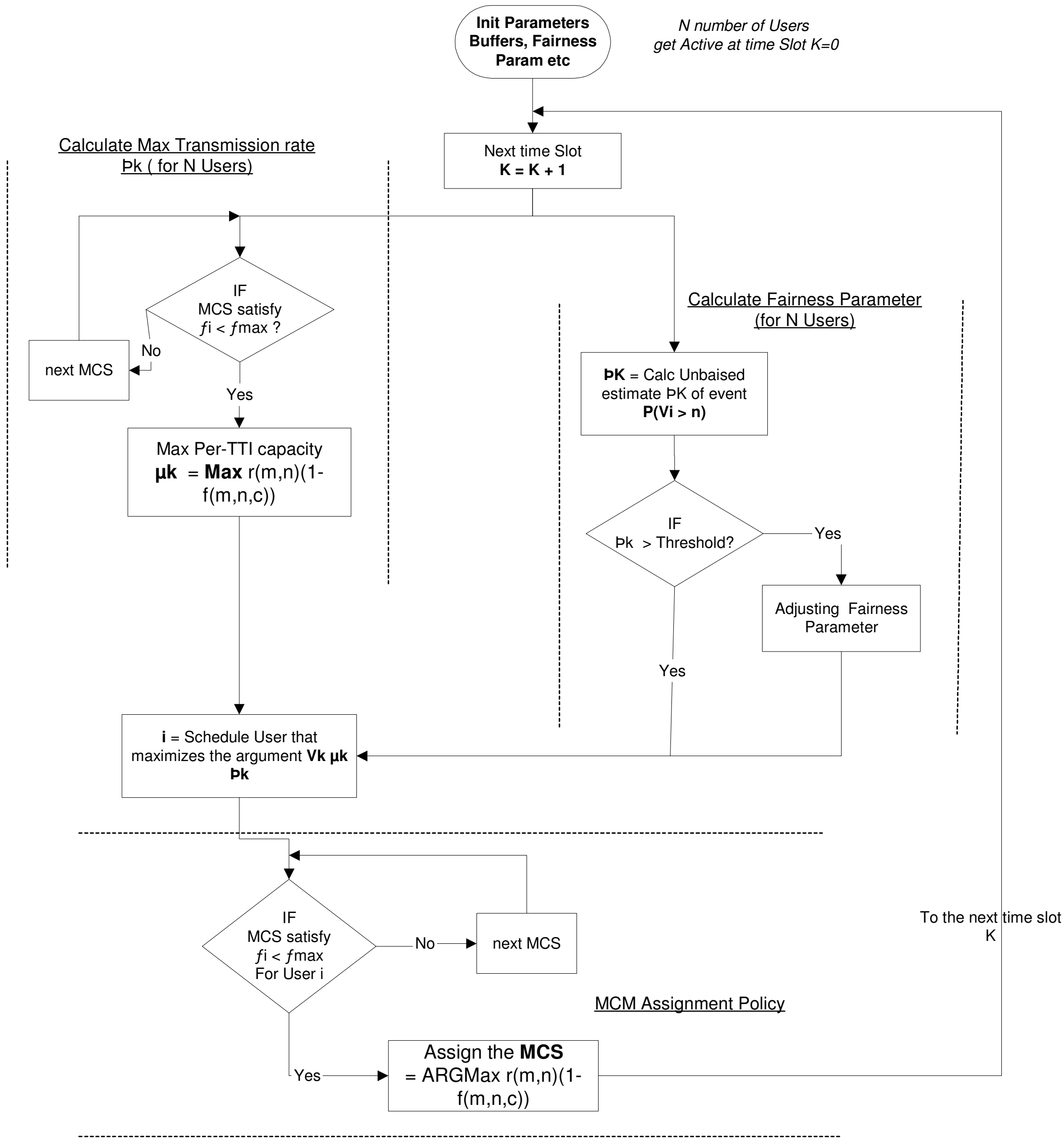
Constraints Two sets of stochastic Quality-of-Service (QoS) constraints: stability constraints and robustness constraints are taken into account while Scheduler is devised.
These constraints make sure that the Users are getting their share of the quality, also the buffers for each user is running smoothly with out interruption.

- *Stability constraints* defined as a queue that it's content do not grow to infinity, so for stable queues the Arriving data rate from server to the BS D_i should match the discharge rate of the BS λ_k in the long run.
$$E(\lambda_k^i) = D_i$$
- *Robustness constraint* is the amount of variation in the UE buffer contents. The Robust quality of service constraint is the probability that the size of the buffer U_i less then the threshold level θ_i for the UE buffer, should be less then the probability threshold of that user δ_i.
$$P(U_i^k < \theta_i) \leq \delta_i$$
- *MAX instantaneous constraint* is the constraint FER should be below a specific level $f_{\max}^i$ for user at each time slot, it is also implemented in this feasible problem for scheduling $f_i^k \leq f_{\max}^i$. This helps to keep a check on the retransmissions which might exceed a maximum level then acceptable for the system.

Feasibility problems are the problems that provide solutions that would settle with in the drafted constraints for the smooth play out of the media files, defining the feasible region of the scheduling problem.

Description

The Discrete event model for streaming users in HSDPA described earlier is used in a way to formulate the Quality constraints described above; using this model and the quality constraints the scheduling problem is turned into a feasibility problem.
Feasible solution is suited best in this case, as it satisfies the quality constraints balancing it with the optimal system throughput. This is put into practice by adjusting the fairness parameter that supplies the necessary priorities to the users where necessary to keep the quality standards in tact.



As shown in the Flow chart a joint scheduling algorithm and MCM assignment policy is outlined, $\Lambda = \{\phi, \psi\}$, where in 'ϕ' is the Scheduling policy and 'ψ' is the MCM assignment rule (the function that maps the system state to a pair of multi-code and MCS number at k).

Here the MCM represented by $\alpha(i, x_k)$ is calculated in a way that the throughput of user is maximized (in case of scheduling user).

$$\alpha(i, x_k) = \arg\max r_k(m,n)\left[1 - f^i\left(m,n,c_k^i\right)\right]$$

An equation for the parametric joint scheduling is given below which depends on the fairness parameter the buffer size and the per TTI capacities.

$$\phi(x_k) = \arg\max \gamma_i \mu_k^i V_k^i$$

This proposed algorithm enables a smooth play-out for the HSDPA along with supporting the quality constraints for the smooth play-out for maximum number of users if possible.

Proportional Fairness Algorithm

Introduction
This algorithm tends to explore the variations in the channel conditions of different users due to fading and other effects. It prioritizes the users that show superior performance in terms of channel quality, in contrast to the average throughput of that user.

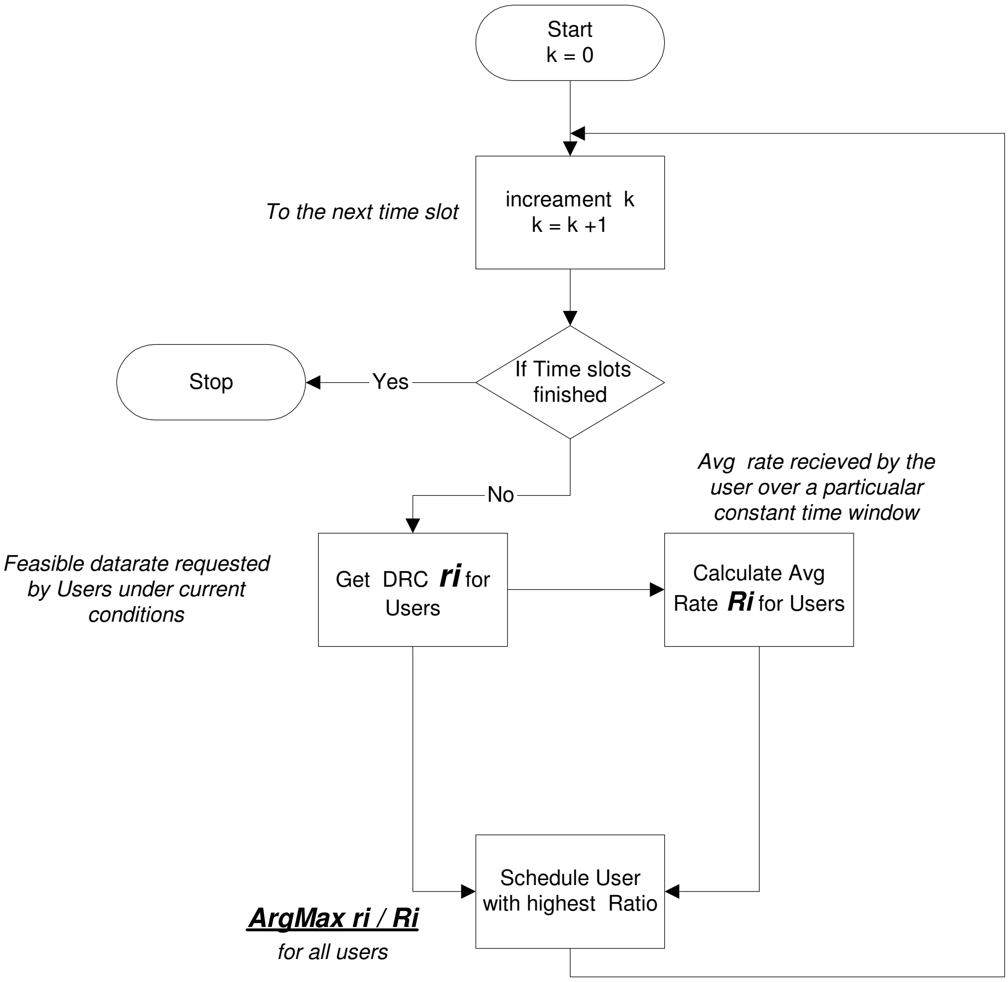
Defined as
Proportional fairness algorithm schedules the users, selecting those with the largest relative channel quality. Relative channel quality is the instantaneous data rate associated with the channel quality condition of the user divided by its current average throughput.

$$\text{(User Scheduled) } i = \arg\max {r_i}/{R_i} \quad \text{for all users}$$

Where ri is the instantaneous data rate of the user i, and Ri is defined as the average data rate effectively received by user i.

Description
The feasible rates ‘r’ for the various users vary over time due to the changing channel condition and quality.
In order to estimate the feasible rates, the base station relies on feedback information from the users on the instantaneous rates that can reliably be supported, so assuming that the base station has perfect knowledge of the feasible rate for every user at the start of the time slot.
The Scheduling of a user is based on the user’s current estimated feasible rate compared to its previous average performance

Phone: +46 (0)23778000
Fax: +46 (0)23778050
URL: http://www.du.se



As can be seen above in the flowchart at the start of each time slot 'k' the user is scheduled that has the highest ratio "ri/Ri " out of all the users currently participating in the transmission process. Update of the average rate is done in each slot, according to the following rule

$$R_i(k+1) = \left(1 - \frac{1}{k_c}\right) R_i(k) + \frac{1}{k_c} r_i$$

where 'kc' is the constant time window over which the average data rate of a user is calculated, 1/Kc is the soothing factor, here k is the current time slot.

QoS constraint described in the description of the problem, are addressed in this algorithm. The algorithm provides a mechanism that makes sure of the users quality of service needs are kept up to an acceptable level by implementing the value of parameter Kc, which is the maximum amount of time for which an individual user can be starved and receive no service. As the algorithm attempts to serve each user at the peak of its channel condition, the scheduler will see a drop in channel condition as temporary until the poor channel conditions persists for more than Kc seconds.

Maximum Carrier to interference Algorithm

The Base station relies on feedback information from the users on the instantaneous rates that can reliably be supported. Assuming that the feasible rates for the various users vary over time according to some stationary and ergodic discrete time stochastic process $\{R_1(t),...,R_N(t)\}$, with $R_i(t)$ representing the feasible rate for user in time slot i.

Maximum CIR algorithm schedules the users i with largest instantaneous supportable data rate at time slot t

$$i = \arg\max_{i=1,...,N} R_i(t)$$

This algorithm is excellent in providing the highest cell throughput; apparently this is due to its scheduling principle.
However, this algorithm has an obvious drawback in that it ignores those users with bad channel conditions, which may lead to starvation. Hence in spite of the fact that the network throughput is maximized, the throughput fairness receives a serious backlash.

Scheduling Performance

Performance Measures

Performance is measured and evaluated based on the buffer variations. The following suggested performance metrics over each simulation run may be provided as congestion parameters. The effects of congestion i.e. loss, delay and overflow, for every user is calculated as follows;

Overflow probability is the probability that, if the buffer is inspected at an arbitrary point in time, the buffer is found to be held at its maximum.

$$O_i = \frac{1}{T}\sum_i \tau_i$$

The buffer overflow probability is estimated from the measured buffer saturation time τ_i and the time T of the total measurement period.
Similarly the loss is expressed as Loss probability

$$L_i = \sum_i \frac{Lfluid_i / T}{M_i}$$

The mean waiting time or delay is expressed for each user through the equation

$$W_i = {}^{N_q}\!\big/_{M_i}$$

Where Nq denotes the average queue size and M denotes the mean offered bit rate.
In general the program runs with several iterations carried out for each type of simulation scenario. The traffic scenarios include variable delay penalty weight, variable traffic ratio, variable maximal queue length, and variable normalized reward parameter. The results are averaged over the program runs and plotted in graphs

Performance Comparison

Finding a comparison on the performance of the three algorithms mentioned above is cumbersome as the scheduling standards have not been frozen because of the HSDPA technology evolving as yet in the scheduling regard at least. So simulations performed in exactly same conditions could not be found, especially to the conditions matching the experiments performed for this thesis.
Here all of the three algorithms make use of the variations in the channel conditions. Comparing the algorithms by the Throughput achieved it was seen from the literature [14] [15] [16] that the maximum C/I scheme tends to achieve higher throughput gain then that of the proportional fairness algorithm where the variation of the channel condition has a larger standard deviation, while as the variation in the channel conditions reduces, the difference in the throughput of both the algorithms also cuts down. On the contrary the fairness between users shows a reverse effect. Hence Maximum C/I experience the worst performance in terms of user satisfaction when the channel conditions between users are subject to large variations because it only serves those users with the best channel conditions while ignoring the rest. While on the other hand Proportional fairness shows better results.
Also the PF scheduler is [17] fair (in terms of the distribution of the users' average throughputs) only in ideal cases where users experience similar channel conditions. However, Proportional fairness is found to be unfair and unable to exploit multi-user diversity in more realistic situations where users usually experience different channel conditions.

In the comparison among Maximum C/I and the Opportunistic scheme for streaming multimedia users, in the scenario whilst the users experiencing similar channel conditions [12] , It is a relatively good solution for maximum C/I scheme to only pick the instantaneously best channel without regarding their queue lengths. The total average throughput will be maximized in this case and because of the symmetry, no user will be particularly starved. In the long run, each user receives more or less the same portion of the maximized throughput and hence the overall performance is relatively good. This situation gives the result of almost the same system throughput for both the algorithms. But when the Users tend to practice quite varying conditions The Max C/I algorithm completely fails in this scenario, because users that are further away from the BS are not served at all, while the Opportunistic algorithm performs quite well in terms of the maximum number of users served with the desired QoS.

7. SIMULATION

In this Chapter the simulated system model as a whole is presented and explained, along with different experiments conducted depicting several scenarios, the results of these have been shown as Line graphs. The experiments or simulations run are used to show the performance of the three scheduling algorithms in the HSDPA system. The performance of the chosen algorithms is measured and assessed based on the congestion parameters i.e. loss, overflow and delay experienced by the users.

Simulation Setup

The HSDPA system was modeled and then simulated (i.e. from a specification model to a computational model) as a Discrete-Event Model that had been developed in C language. The Network system model includes the simulation of previously discussed Flat fading channel simulation, On/Off Model for traffic generation and the three scheduling algorithms detailed in the previous chapter.

Discrete-Event Model

Discrete-event simulation [18] is a way to build a model, so that the dynamic (time based) behavior of the system can be observed. In the system each event occurs at an instant in time and marks a change in the state of the system. During the experimental phase the Discrete-event model is executed (run over time) in order to generate results. The results can then be used to provide insight into a system and forms a base to make decisions on.
The general steps involved in the development of a DES model starts by

1. Determining the Goals of the system to be developed
2. Building of a conceptual model.
3. Converting it into a specification model.
4. Followed by converting the specification model into a computational model.
5. Verifying the system developed in the previous step and finally the validation (computational model being consistent with the system being analyzed) of the system.

In this thesis the discrete-event simulation is used at the network call layer to access the performance and behavior of the packet scheduling algorithms, as to how these algorithms perform under different conditions, the common characteristics of the algorithms, the complexities of these algorithms etc. Discrete event simulations can be implemented in any of the four following methods; event based, process based, activity based and the three phase approach. In addition to the representation of system state variables and the logic of what happens when system events occur, discrete event simulations include the following main components [19]:

- *Clock:* to keep track of the current simulation time, in whatever measurement units are suitable for the system being modeled. Because the events are instantaneous, the clock uses time ´hops` to keep track of the simulation events occurring.
- ***Event List:*** The simulation maintains a list for the simulation events. An event must have a start time, some kind of code that constitutes the performance of the event

itself, and possibly an end time. In some approaches, there are separate lists for current and future events. Events in their lists are sorted by event start time.

- ***Random number Generator:*** The simulation needs to generate random variables of various kinds, depending on the system model. This is accomplished by one or more pseudorandom number generators.
- ***Statistics:*** The simulation usually keeps track on the system's data, which calculates and analyzes the features of interest in the system.
- ***Ending Condition:*** it is practical to end the simulations execution, as the simulation would run for ever until an ending condition be specified. Typical choices are “at time t” or “after processing n number of events.

Fluid Flow Model

Certain discrete-event simulation techniques have helped in the increase in the model scalability i.e., the size of network and the traffic densities that can be executed in real-time. Fluid-based modeling [20] is used to simplify traffic flows in a network simulation. With a fluid model, events are only generated when the rate of a flow changes.
In the fluid simulation model, network traffic is modeled in terms of a continuous fluid flow, rather than discrete packet instances. A cluster of closely-spaced packets may be modeled as a single fluid chunk with a constant fluid rate, with small time-scale variations in the packet stream being abstracted out of the model [21].
In fluid simulation, the higher level of abstraction suggests that less processing might be needed to simulate network traffic. Intuitively, this is not surprising as a large number of packets can be represented by a single fluid chunk. For simple network components, where traffic flows do not compete for resources, the fluid simulator outperforms the packet-level simulator. One drawback of a fluid model is that the accuracy of the interest measures is compromised due to the abstraction.

Markovian on-off source models are often used in network research to capture the bursty nature of the network traffic. The source transits between an ON and OFF state, remaining in each state for an exponentially distributed amount of time. When in the on state, fluid source sends out fluid at a constant rate. No fluid is sent during the OFF period. On/Off sources are commonly used as traffic models in the fluid simulation.

The simulation of Traffic for the network has been implemented as Fluid Flow Markov On/Off model. In the traffic simulation the buffer is modeled as the inflow and outflow of data, as such that the buffer is seen as a fluid reservoir with a hole in the bottom and the arriving of information as fluid running into the reservoir. Hence has the name Fluid Flow. Each time the event of inflow occurs (rate in change of information from No information to some rate of information) for a particular source the state of that source is said to be in the ON state, and while there is not any inflow of information the source is in the Off state. These times for inflow of information are exponentially distributed.
Each buffer is of finite size B Mbits with inflow rate of information coming into the buffer, an outflow rate of information flowing out of the buffer and a netflow being the difference between the inflow and outflow [22]. It is assumed that, in a fluid simulation the inflow fluid remains (roughly) constant over long time periods with information coming into the buffer at

a peak data rate depending on the type of class of traffic. The system is modeled with sources representing traffic classes such as streaming class and interactive class.
The arrival of information and departure (i.e. burst arrival & burst departure) are modeled as discrete event. The simulation of the DES has its foundation in the event-list, which is a linked list (data structure) of event records [23].Each event has a continuous time entry when the event should occur. The event list is sorted by the event occurrence time, in increasing time order. The head of the event list contains the next event that should happen. The time when the event should occur, or time between events, is determined by the Inverse method using a random number generator. A random number with distribution F(x) is determined by F−1(U) where U is a uniform random number in the interval [0,1].
At each Instantaneous Time Interval a decision is made accordingly to the simulated scheduling algorithm at to the outflow on the users (scheduling a user). That is based on the channel conditions of the users, these channel conditions are provided by the simulated flat fading channel model implemented as a first order Markov model for fading channels.
The performance of these scheduling algorithms, as discussed previously are measured and evaluated based on the buffer variations with the aid of congestion parameters i.e. loss, overflow and delay for each user.

The network system computational model is executed and runs through in a sequential manner passing through the following main steps shown in the diagram and explained further on.

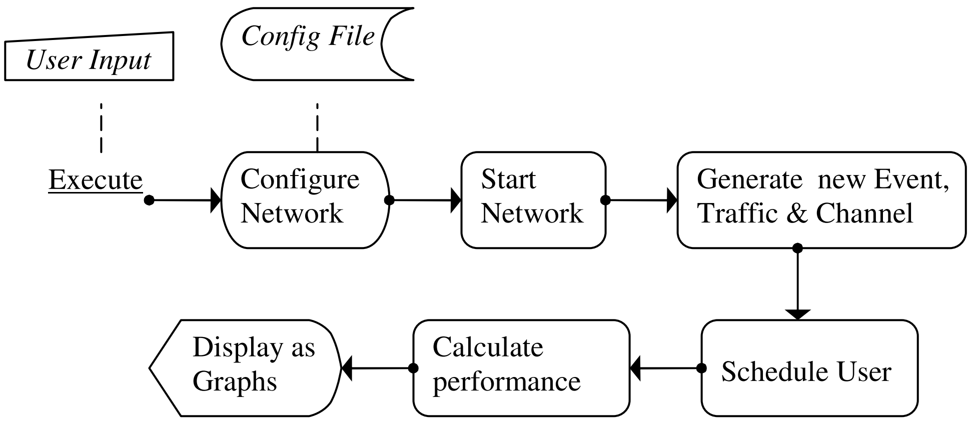


Figure #: Sequential flow of the Simulation

- The simulation program executes as the user sets the simulation parameters by giving input arguments for the simulation i.e. name of configuration file, type of scheduling downlink/uplink, the scheduling algorithm, the type of simulation etc.

- Following it the next step, configures the network taking values from the configuration file written by the user with realistic parameters and storing them in the simulation for further use. These configuration parameters involve the following:

    - Cell number.
    - Number of sources.

  - Number of Channel states per source.
  - Burst traffic parameters (Peak rate, On-Off and Off-On rate).
  - Markov channel state probabilities.
  - Link capacity.
  - Buffer size per source.

- Once the network is configured the simulation proceeds with initializing and starting up the network. Here the network creates the event list and inserts the first events representing the first traffic bursts for each of the sources. Also the first TTI event is inserted into the event list.

- The next step generates the subsequent events, building the event list until the numbers of specified events are reached (the simulation time is reached 10 sec in this case). In these events is inserted the generation of traffic burst or departure of the traffic burst depending upon the On/Off model distribution and also the insertion of new TTI event happens where each source is given a channel state depend on the Markov model. After which the specified scheduler is called for the decision making of the user to be scheduled.

- Further in the simulation the calculation of the performance parameters is done based on the variations of the buffer levels of the users.

- The last step involves the analysis and performance measurement of the different scheduling algorithms used by plotting the performance parameters in the form of graphs.

Simulation Configuration

The simulation is run under parameters configured for Realistic values, so that the results obtained could form a basis for the planning and designing of radio recourse networks. The parameter values can be found from the literature [24], [25], [26] and [27].
In order to run the simulation optimally a few adjustments had to be made to the values of the configuration parameters in the simulation. These adjustments would make the system work in the same manner as with the above listed realistic values, just that the system would be perceived as a minor version of the original one.
Peak data Rates for Terminal/User depends on the category type of the terminal, one of the 12 categories of terminals available. Hence, depending on the MCS and channel state. The number of users in a cell is dependent on the type of service and traffic class required by the users in that cell e.g. 40 numbers of simultaneous users of 128kbps streaming in 5 MHz.

Traffic type	**Parameter Name**	**Parameter Value**
Streaming/Mixed	Maximum downlink channel capacity	10.2 Mbps
Interactive	Maximum downlink channel capacity	1.0 Mbps

Streaming/ Interactive	Transmission time interval	2 ms
Streaming/ Interactive	Buffer Size	0.9 Mbits - 28.8 Mbits *(0-10 Mbits for simulation)*
Streaming/ Interactive	Channel state transition probability matrix	Any kind i.e. Uniform, fixed, calculated etc
Streaming	Peak Rate, Off→On rate, On→Off rate	7.2 Mbps, 0.8, 0.3
Interactive	Peak Rate, Off→On rate, On→Off rate	0.7 Mbps, 0.2, 0.7
Streaming/ Interactive	Number of Markov channel states	10 *(5 for simulation)*
Streaming/ Interactive	Number of users/sources	Varies depending upon the type of users and the type of traffic generated
Streaming/ Interactive	Number of users/sources	1000000 (10 sec)

Table

The algorithms need to be fine tuned according to realistic values [28] [13] for the simulation as well. The Opportunistic algorithm for streaming users and Proportional algorithm have parameters that control the scheduling of the users while the Max C/I is implemented straight forward as it schedules users with best channel conditions leaving less room for adjustment of itself at least in the case of this thesis.
Following are the values used in the algorithms for the different parameters.

Opportunistic algorithm	
Max Instantaneous FER range f^i	$(0 < f^i < 0.2)$
MAX instantaneous constraint: $f^i_{\max}$	0.1
Buffer Threshold: θ	0.3
Unbiased Estimate: δ	0.1 $(0 < \delta < 1)$
Proportional algorithm	
Smoothing parameter: a	$1/t_c$
Time constant: t_c	1000 slots

In the Opportunistic algorithm the value for f^i are read into the simulation program from a separate file with the name "oppsetting.dat". the program uses 5 values from the range $(0 < f^i < 0.2)$, further manipulated depending on the type of channel of the user possess.

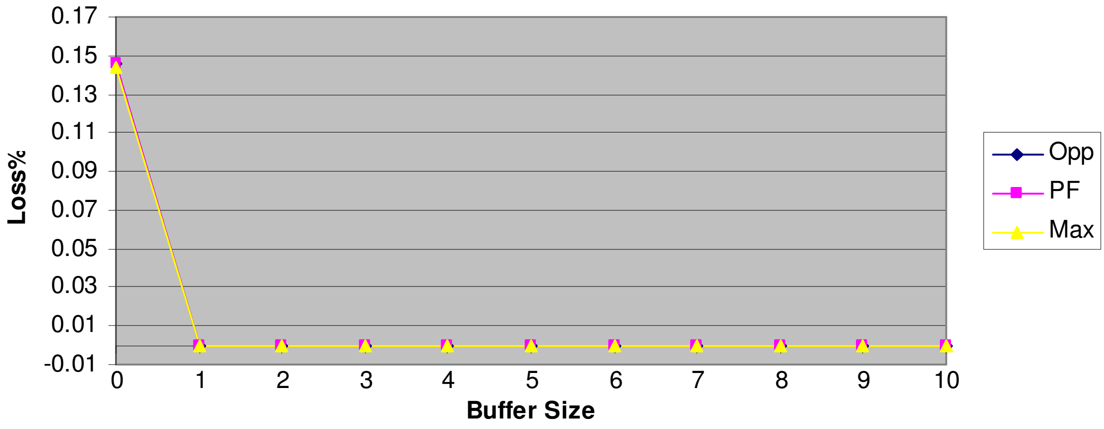
Experiments & Results

The simulated scheduling techniques are run for the purpose of evaluation and comparison, each of the simulation is run for a number of iteration. The simulation is considered to e executed under different conditions/scenarios.

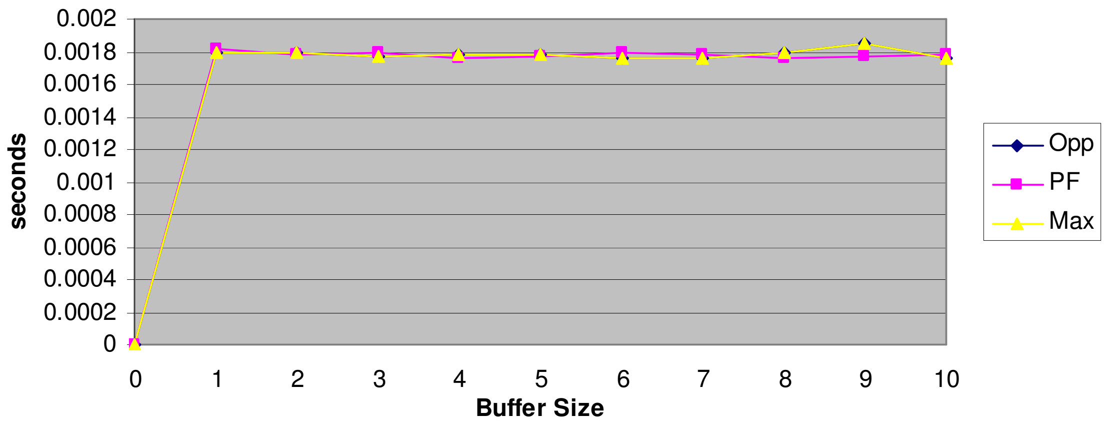
Scenario *1*:

In The first scenario the simulation is run under realistic parameters listed in the table above for user's of traffic type belonging to Streaming Class. The simulation results are plotted for the congestion parameters against the buffer size (expressed in Mbits).
The simulations under these conditions are run for the following number of users per cell
For 1 USER / CELL

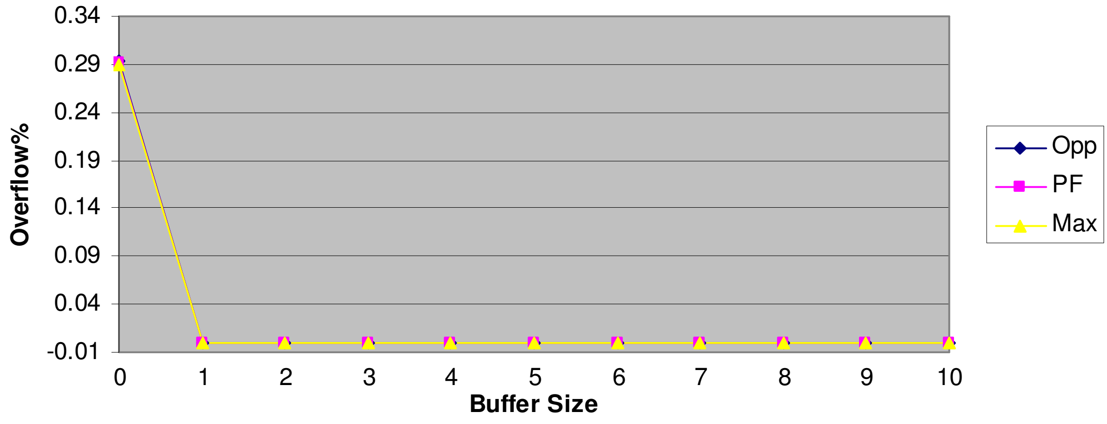
Graph: 7.1.1 ONE USER , LOSS against BUFFER-SIZE



Graph: 7.1.2 ONE USER, DELAY against BUFFER-SIZE

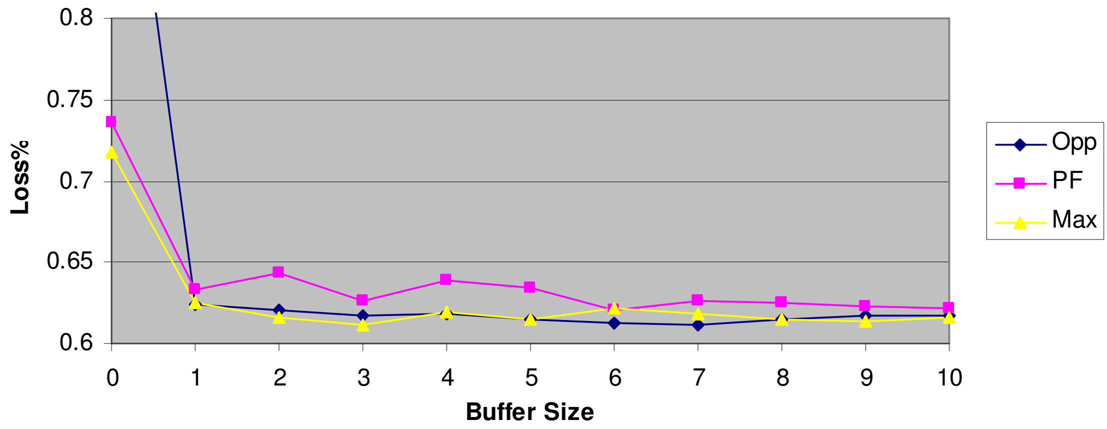


Graph: 7.1.3 ONE USER, OVERFLOW against BUFFER-SIZE

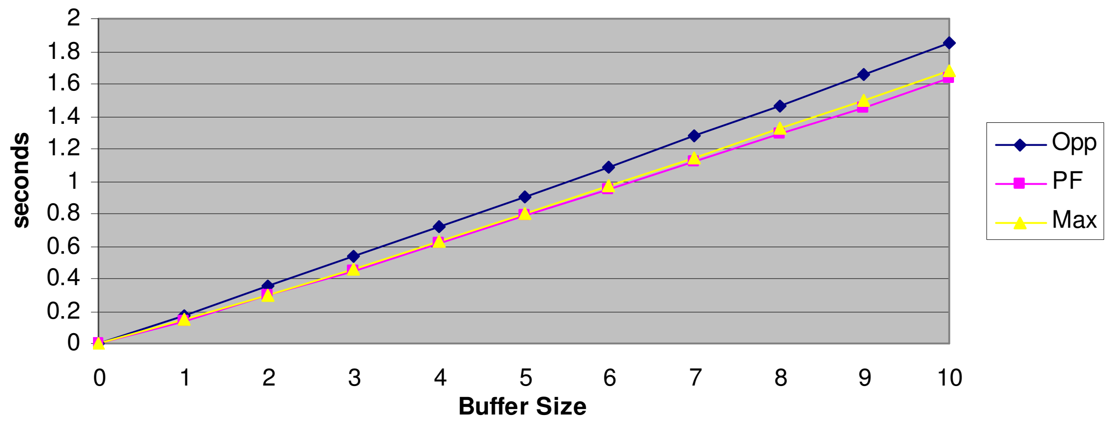


<u>For 5 USERS/CELL</u>

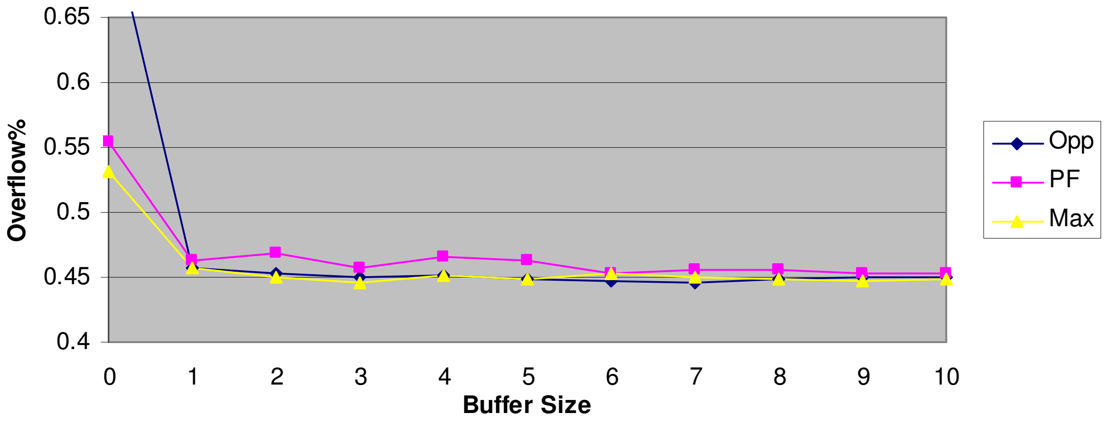
Graph: 7.1.4 FIVE USERS, LOSS against BUFFER-SIZE



Graph: 7.2.5 FIVE USERS, DELAY against BUFFER-SIZE

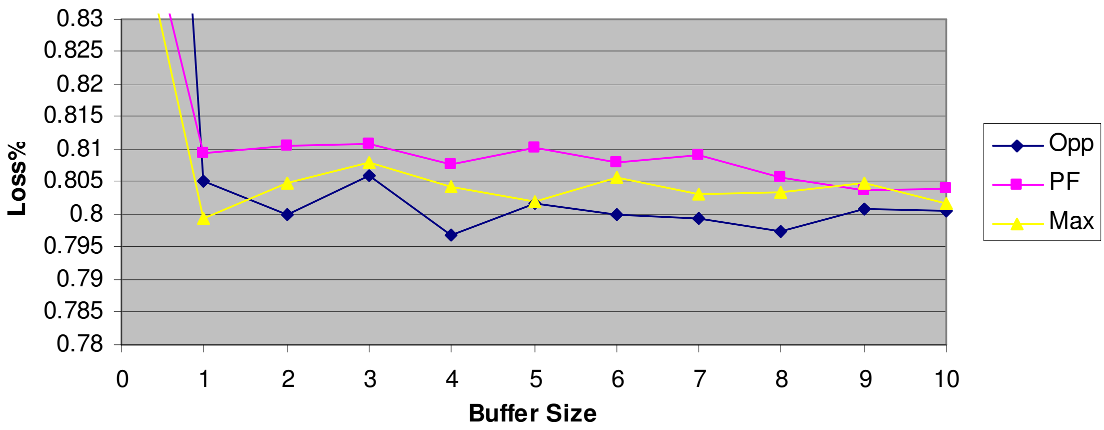


Graph 7.2.6 FIVE USERS, OVERFLOW against BUFFER-SIZE

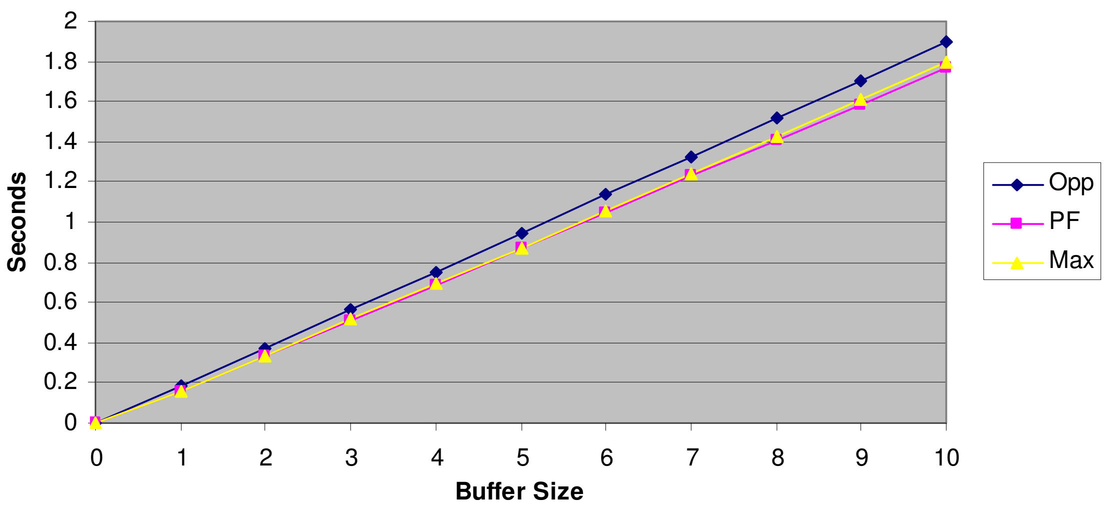


For 10 USERS/CELL

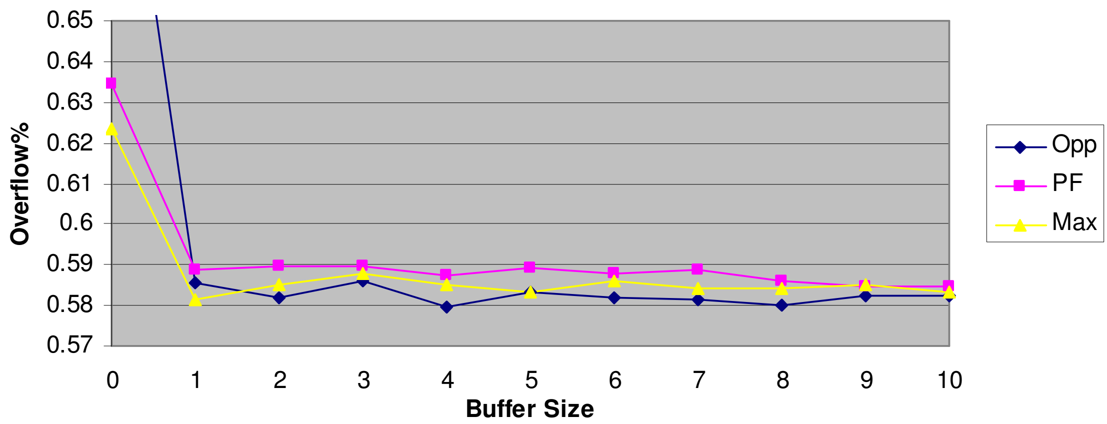
Graph 7.2.7 TEN USERS, LOSS against BUFFER-SIZE



Graph 7.2.8 TEN USERS, DELAY against BUFFER-SIZE



Graph 7.2.9 TEN USERS, OVERFLOW against BUFFER-SIZE

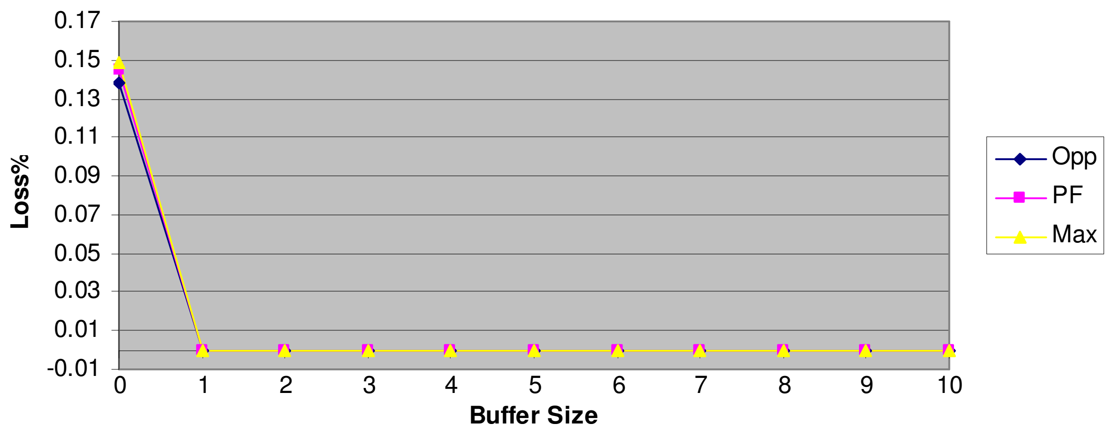


Scenario 2:

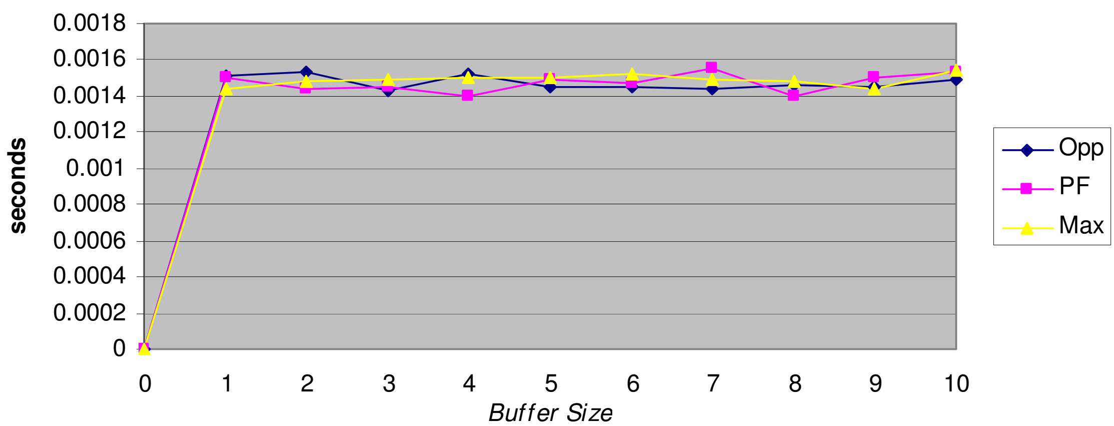
The second scenario is similar to the first scenario, with the only difference i.e. the users belong to the interactive traffic class. This set of experiments is also run under the realistic parameters with the same number of users.

<u>For 1 USER/CELL</u>

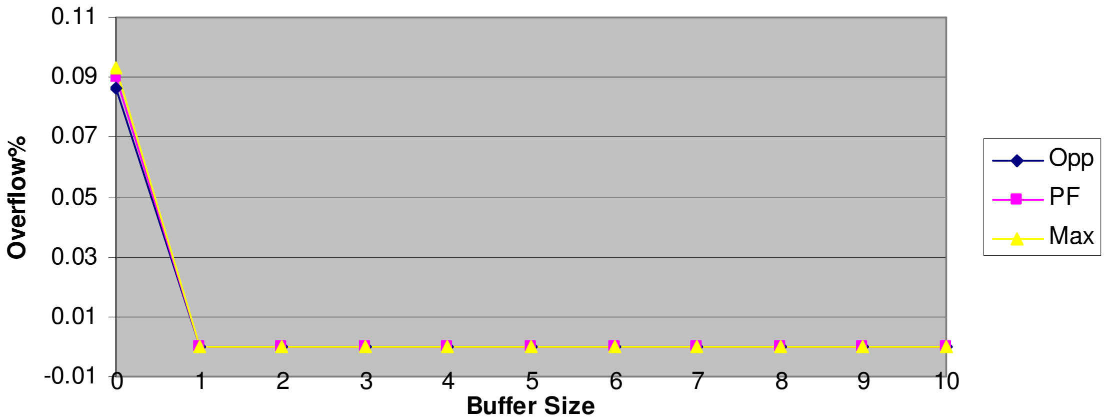
Graph 7.2.1 ONE USER, LOSS against BUFFER-SIZE



Graph 7.2.2 ONE USER, DELAY against BUFFER-SIZE

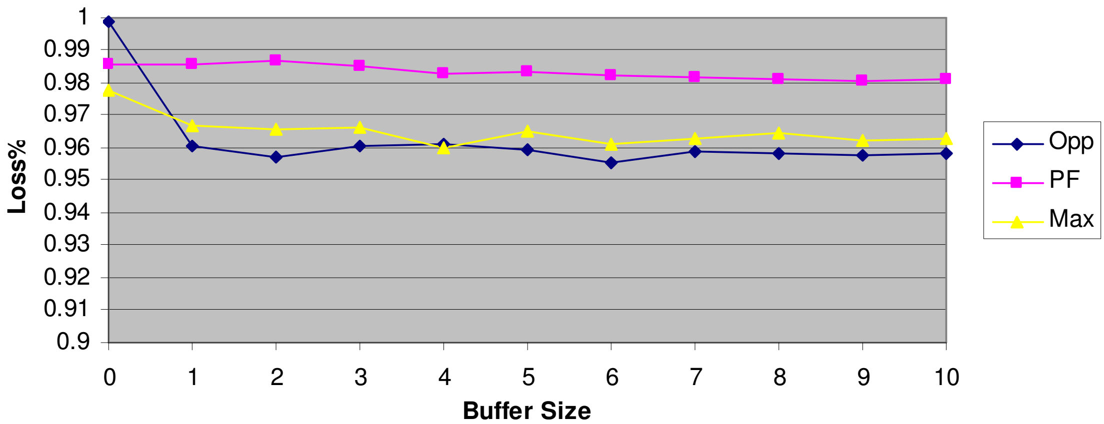


Graph 7.2.3 ONE USER, OVERFLOW against BUFFER-SIZE

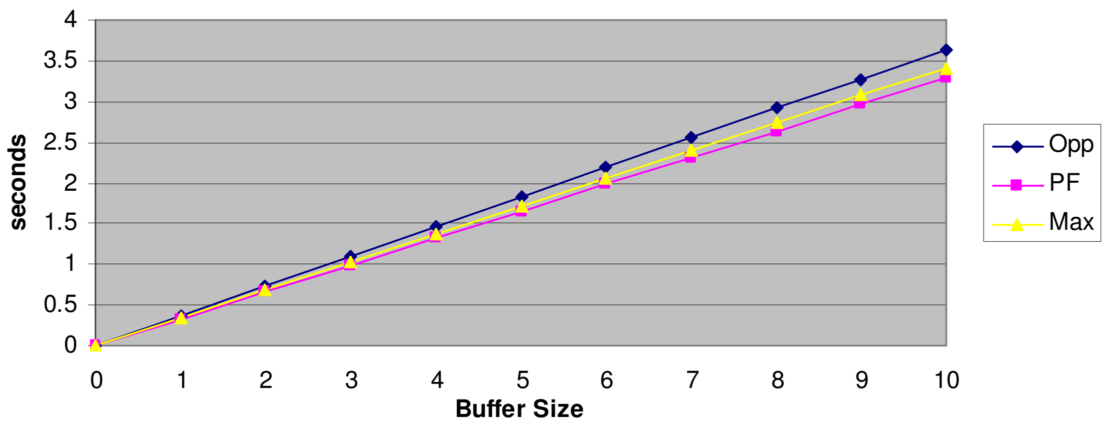


For 5 USERS/CELL

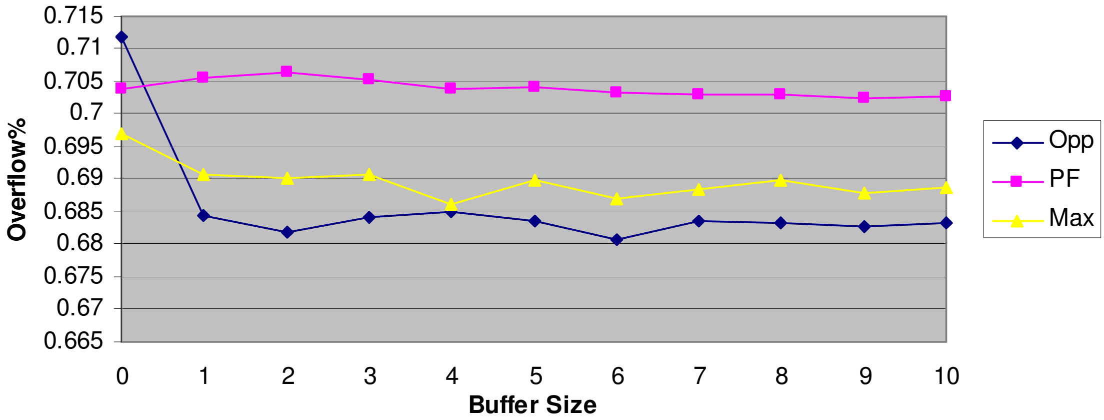
Graph 7.2.4 FIVE USERS, LOSS against BUFFER-SIZE



Graph 7.2.5 FIVE USERS, DELAY against BUFFER-SIZE

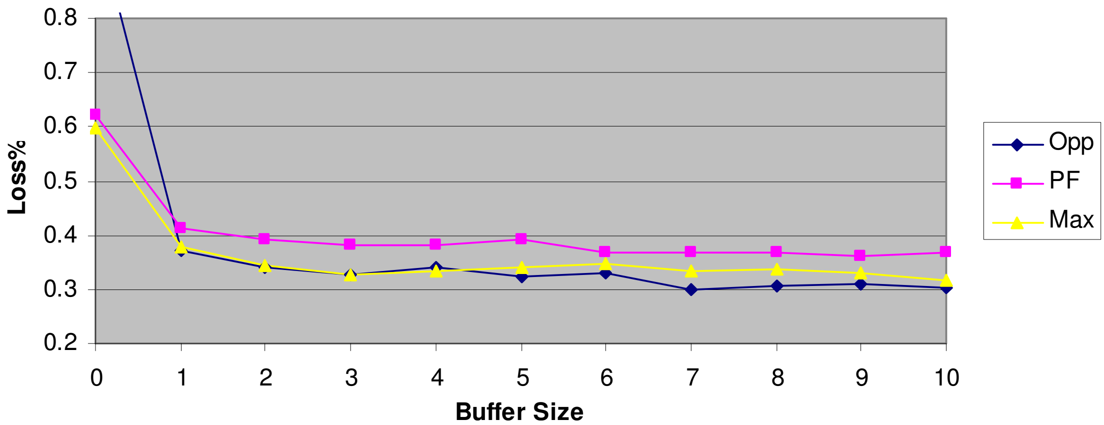


Graph 7.2.6 FIVE USERS, OVERFLOW against BUFFER-SIZE

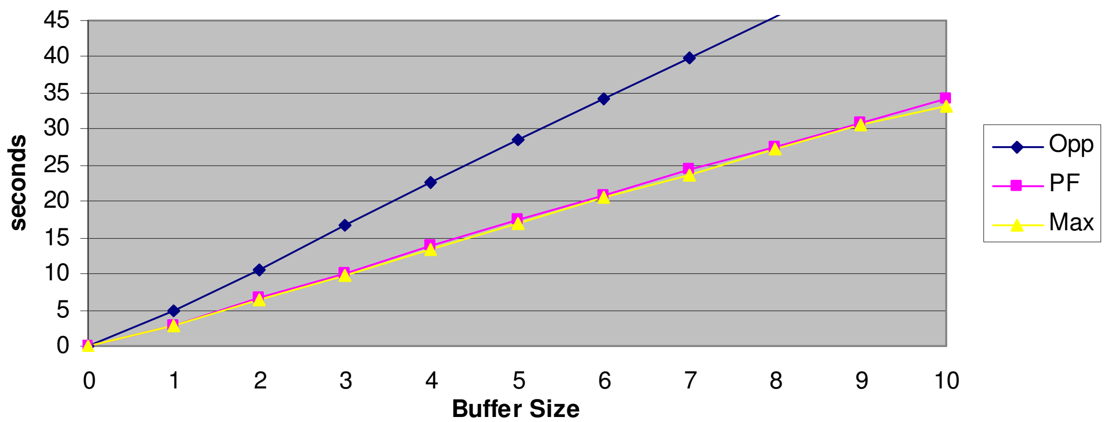


For 10 USERS/CELL

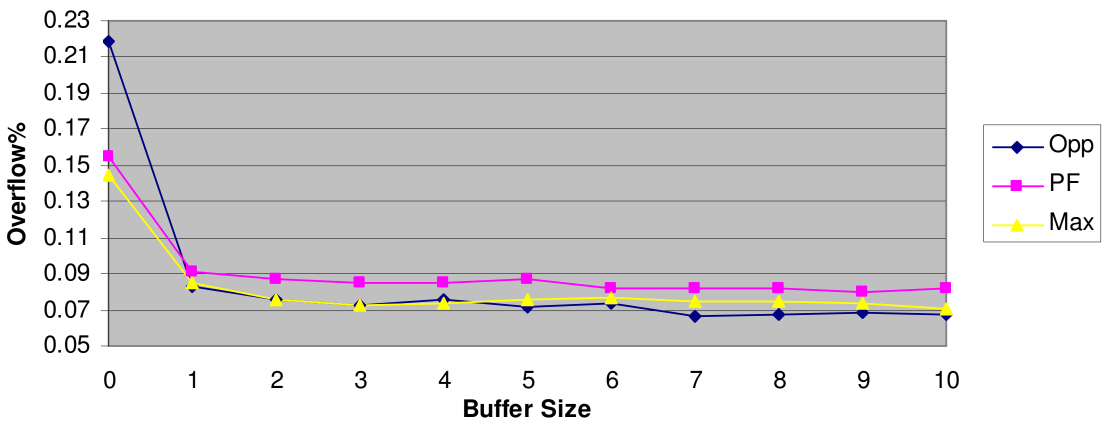
Graph 7.2.7 TEN USERS, LOSS against BUFFER-SIZE



Graph 7.2.8 TEN USERS, DELAY against BUFFER-SIZE



Graph 7.2.9 TEN USERS, OVERFLOW against BUFFER-SIZE

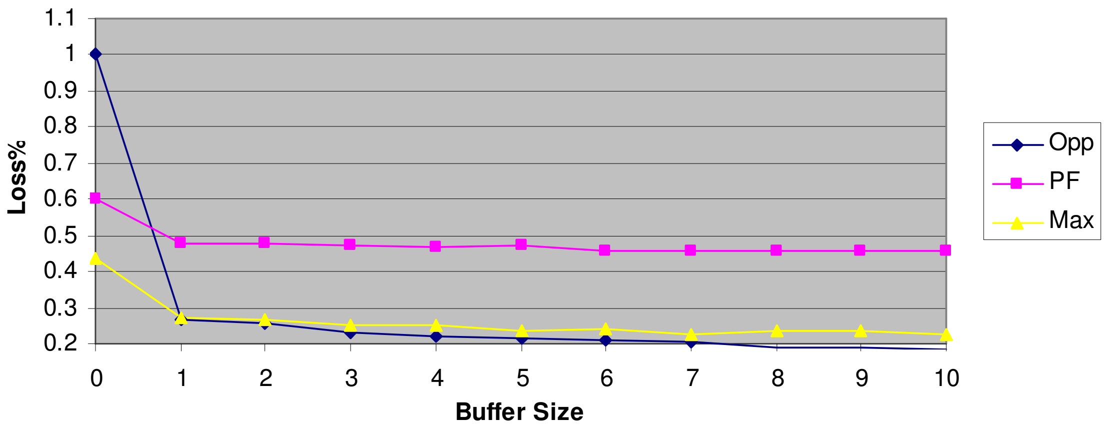


Scenario 3:

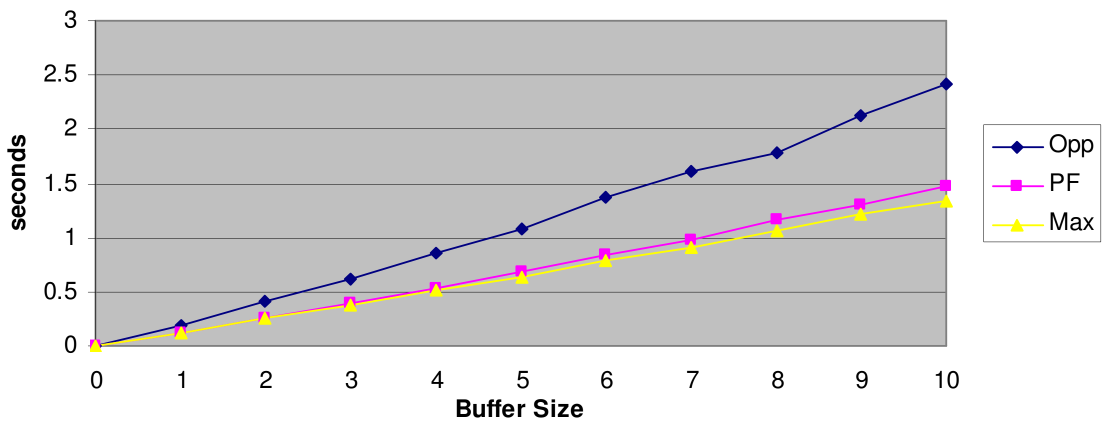
In the third scenario the simulations are run for 5, 10 and 100 users of the mix traffic class (both Interactive/Streaming). The remaining parameters remain as previous.

<u>For 5 USERS/CELL</u>

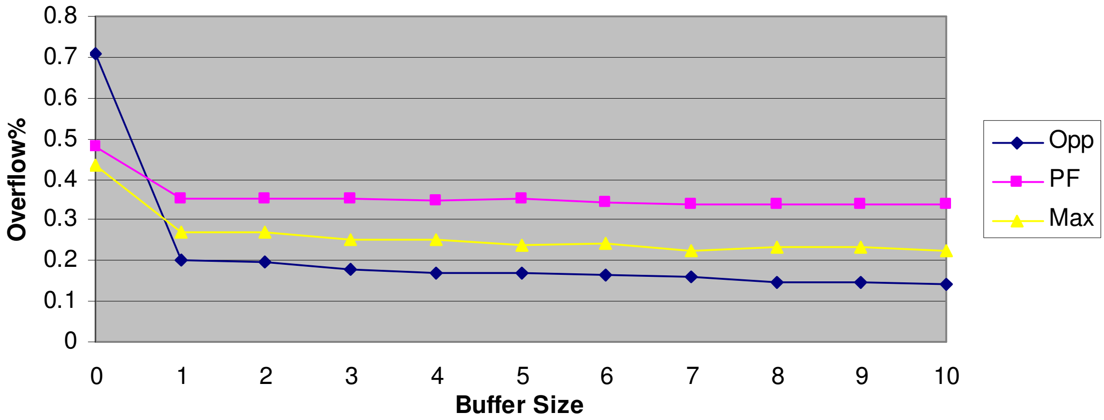
Graph 7.3.1 FIVE USERS, LOSS against BUFFER-SIZE



Graph 7.3.2 FIVE USERS, DELAY against BUFFER-SIZE

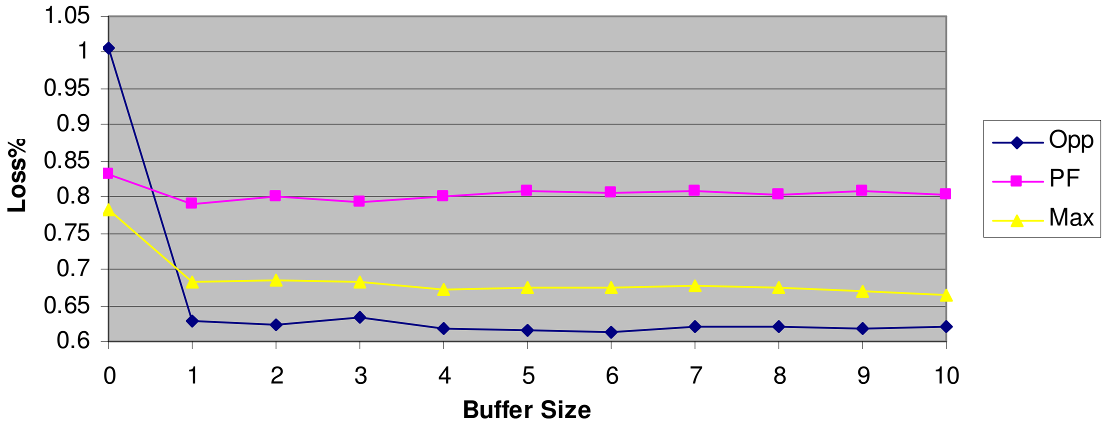


Graph 7.3.3 FIVE USERS, OVERFLOW against BUFFER-SIZE

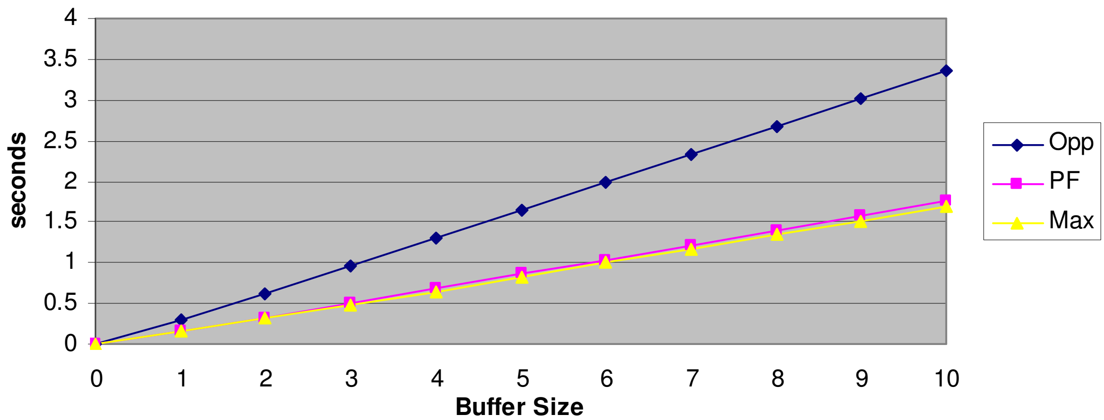


For 10 USERS/CELL

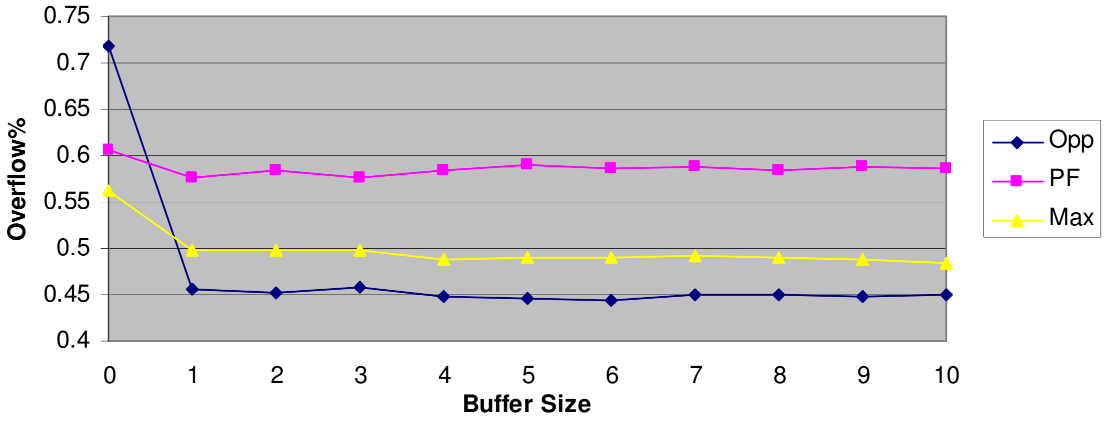
Graph 7.3.4 TEN USERS, LOSS against BUFFER-SIZE



Graph 7.3.5 TEN USERS, DELAY against BUFFER-SIZE

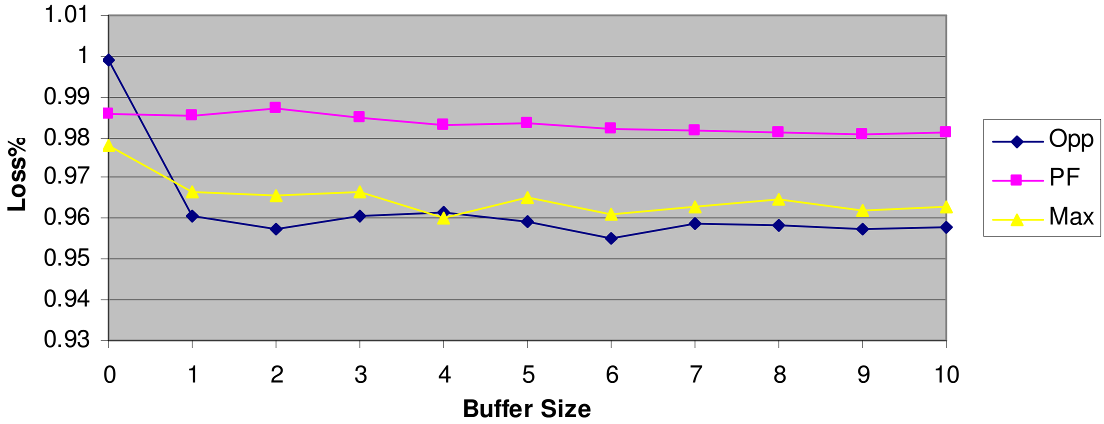


Graph 7.3.6 TEN USERS, OVERFLOW against BUFFER-SIZE

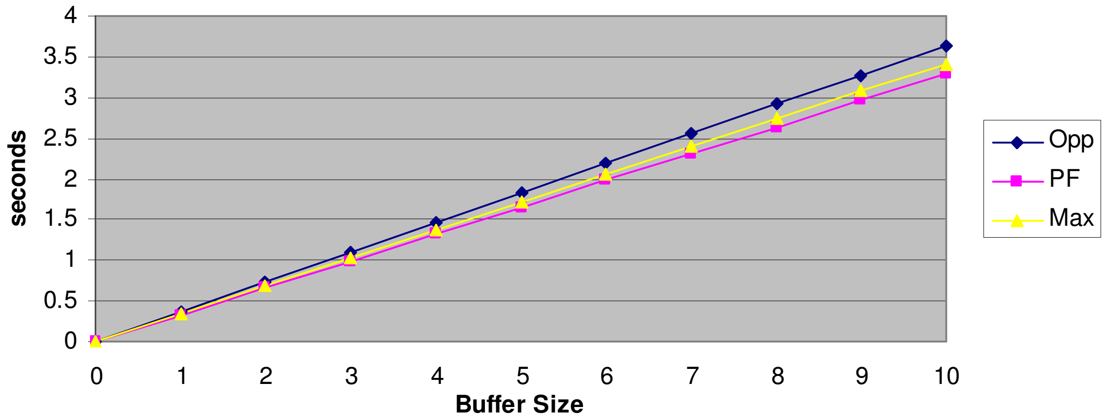


<u>For 100 USERS/CELL</u>

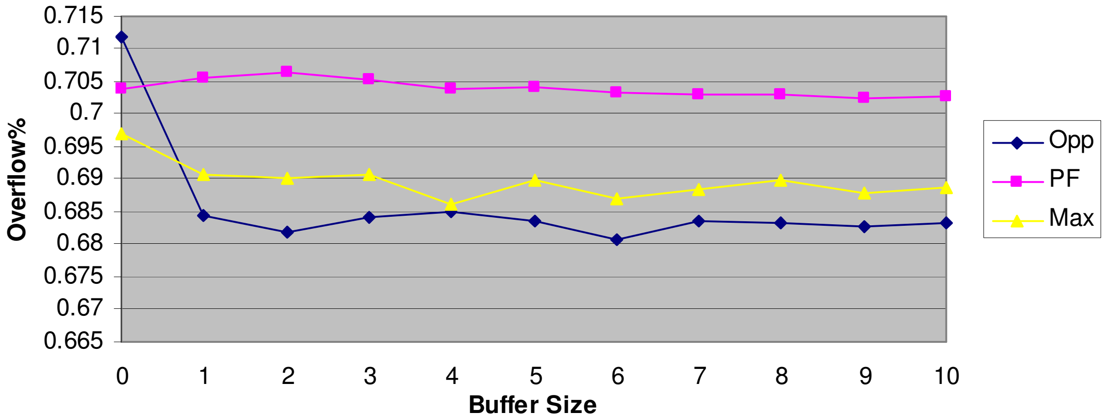
Graph 7.3.7 100 USERS, LOSS against BUFFER-SIZE



Graph 7.3.8 100 USERS, DELAY against BUFFER-SIZE



Graph 7.3.9 100 USERS, OVERFLOW against BUFFER-SIZE

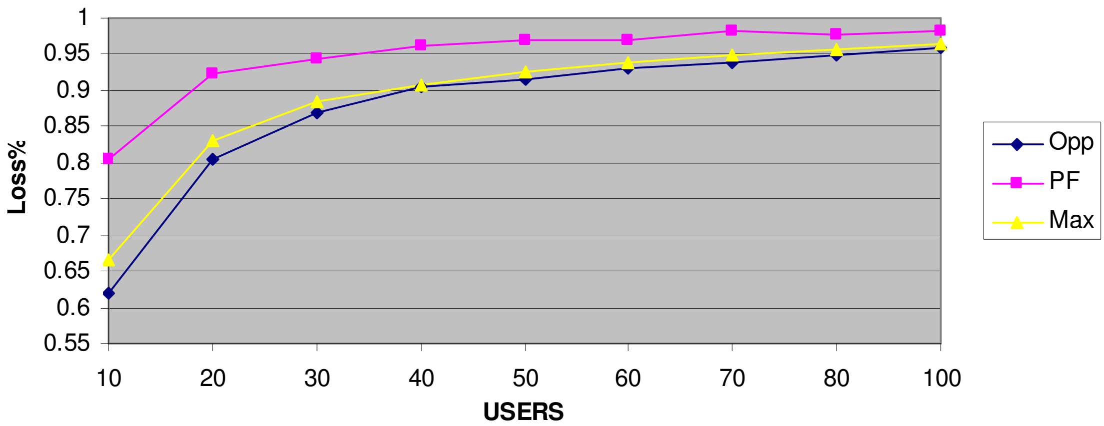


Scenario 4:

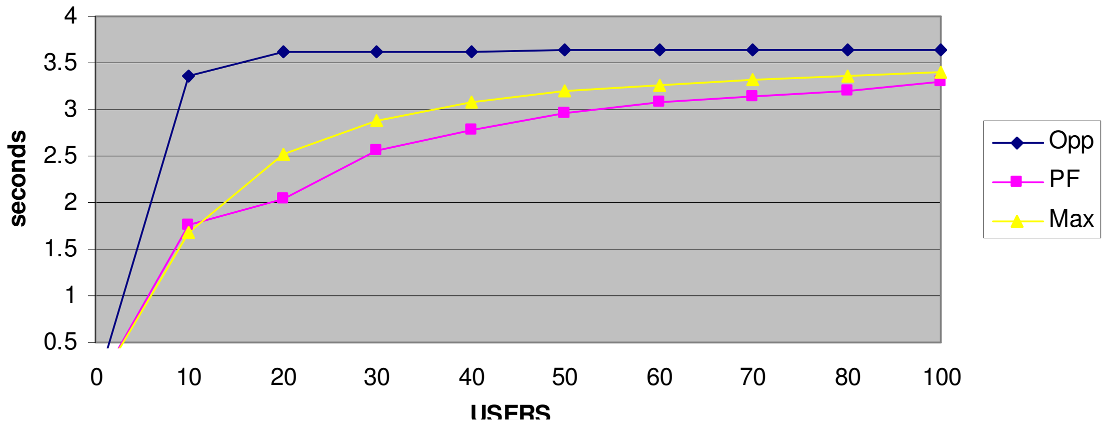
The fourth scenario is for Mix traffic users, the number of users in this experiment varies from 10 - 100 users per cell incrementing in each simulation with 10 users. Each simulation is run with 10 Mbits buffer-size per user. The graph plotted is between the number of users against the performance parameters.

<u>For 10-100 USERS (10 MBITS BUFFER EACH)</u>

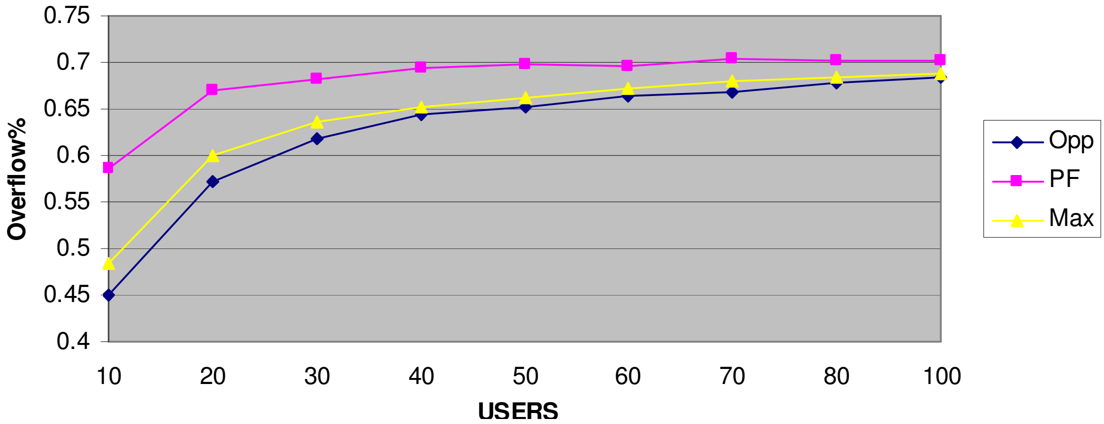
Graph 7.4.1 TEN MBITS BUFFER, LOSS against RANGE of USERS



Graph 7.4.2 TEN MBITS BUFFER, DELAY against RANGE of USERS



Graph 7.4.3 TEN MBITS BUFFER, OVERFLOW against RANGE of USERS

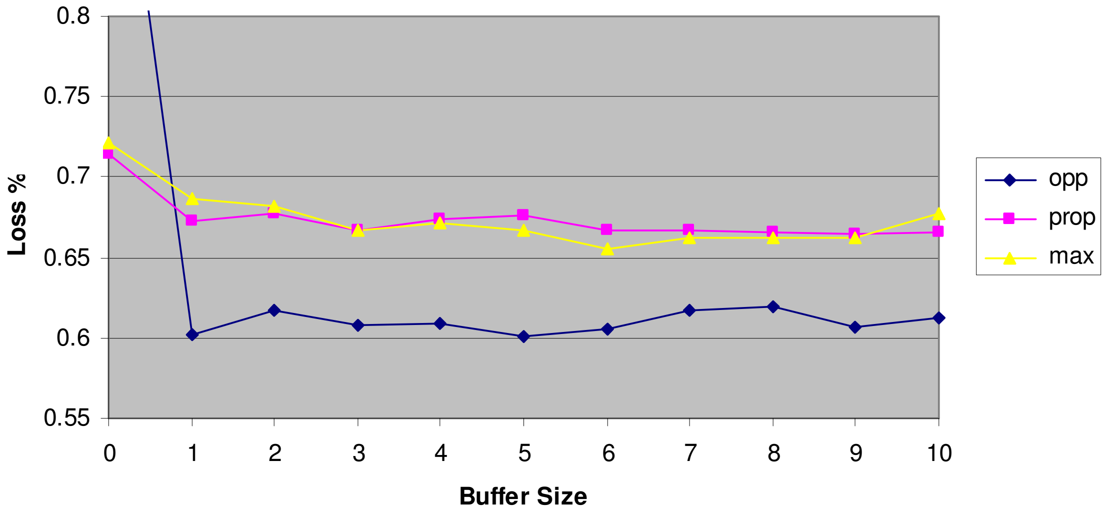


Scenario 5:

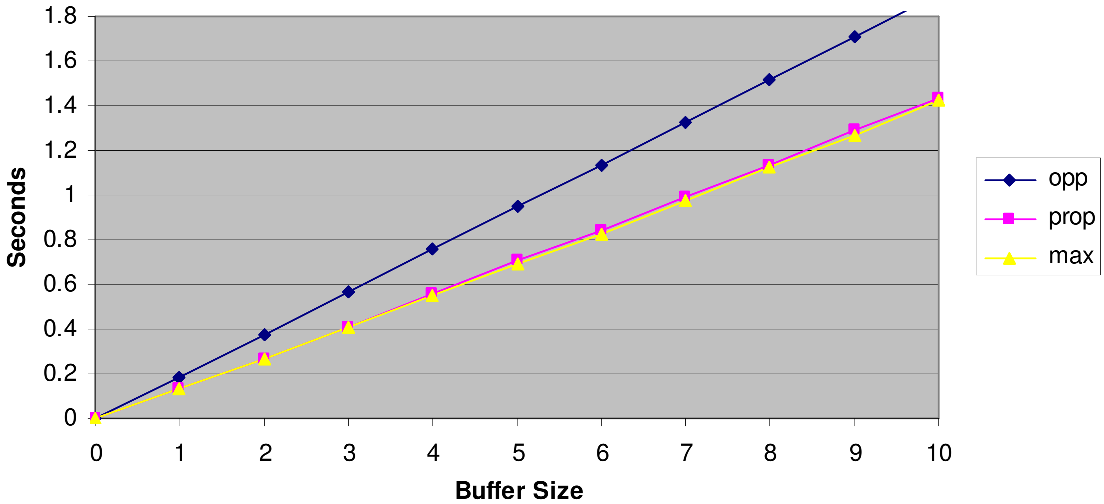
In the fifth and last scenario the simulation is run for Streaming Traffic users. Here unlike the previous experiments the channel conditions for the users have been kept almost similar to each other. The simulation is run for 5, 10 and 100 users per cell. The graphs plotted are the range of buffer-size 0-10 against the performance parameters.
The last simulation is i.e. for 100 users is run for only the MAX C/I and PF algorithm due to computational complexity further discussed in the result analysis section.

<u>For 5 USERS/CELL in Similar Channel Conditions</u>

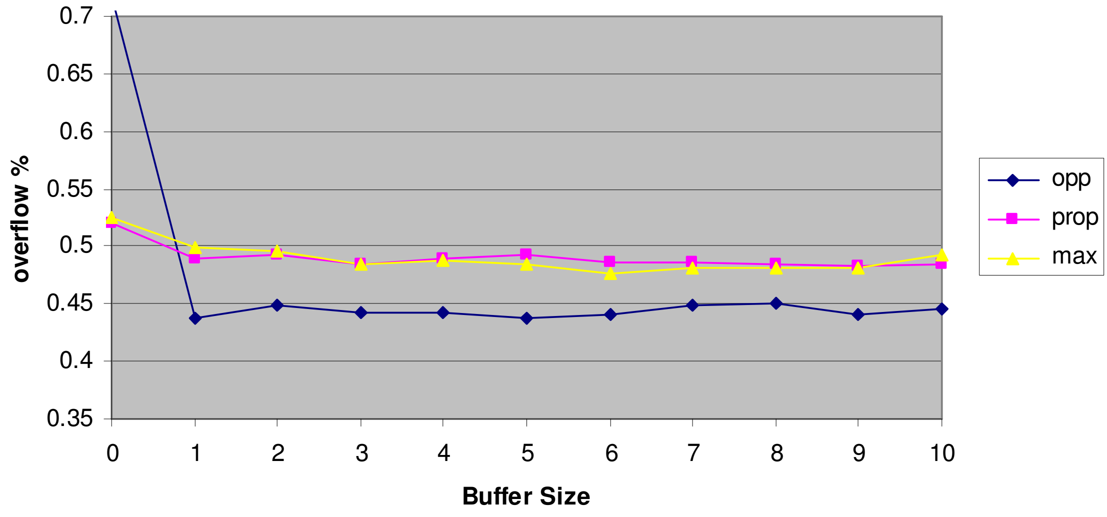
Graph 7.5.1 FIVE USERS, LOSS against BUFFER-SIZE



Graph 7.5.2 FIVE USERS, DELAY against BUFFER-SIZE

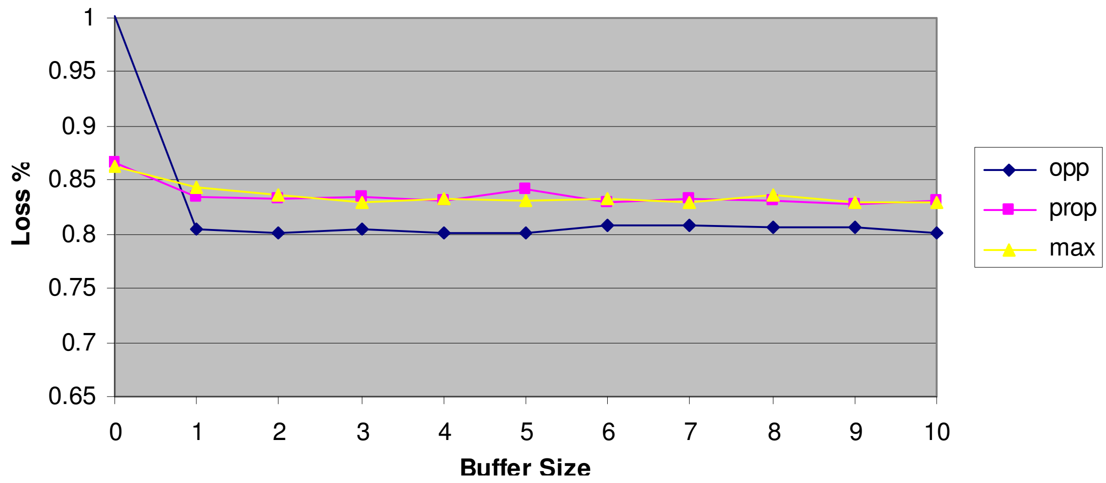


Graph 7.5.3 FIVE USERS, OVERFLOW against BUFFER-SIZE

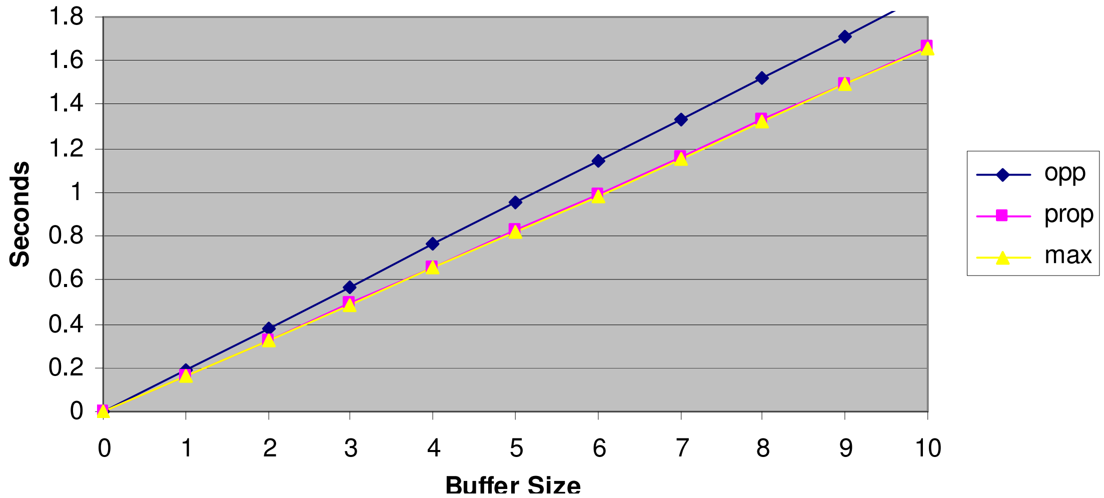


For 10 USERS/CELL in Similar Channel Conditions

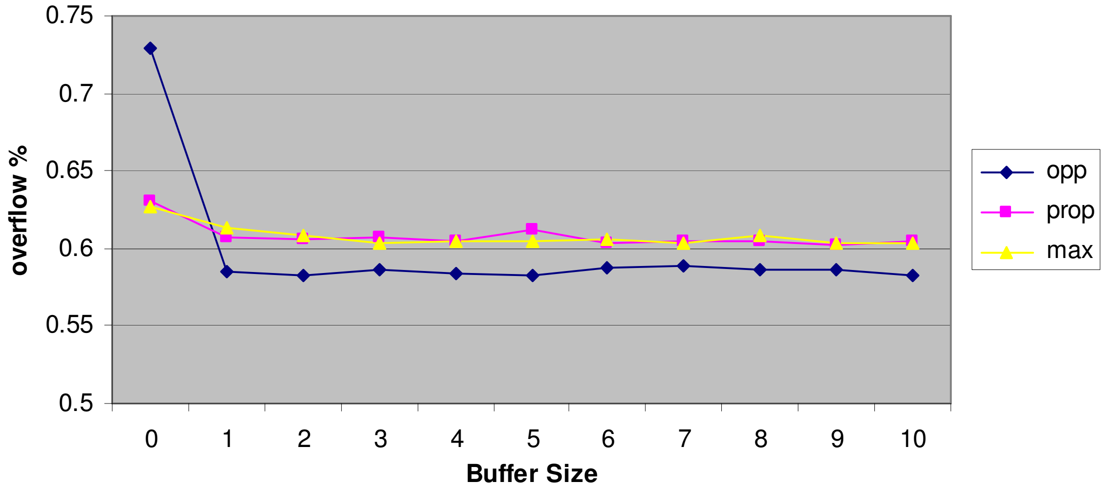
Graph 7.5.4 TEN USERS, LOSS against BUFFER-SIZE



Graph 7.5.5 TEN USERS, DELAY against BUFFER-SIZE

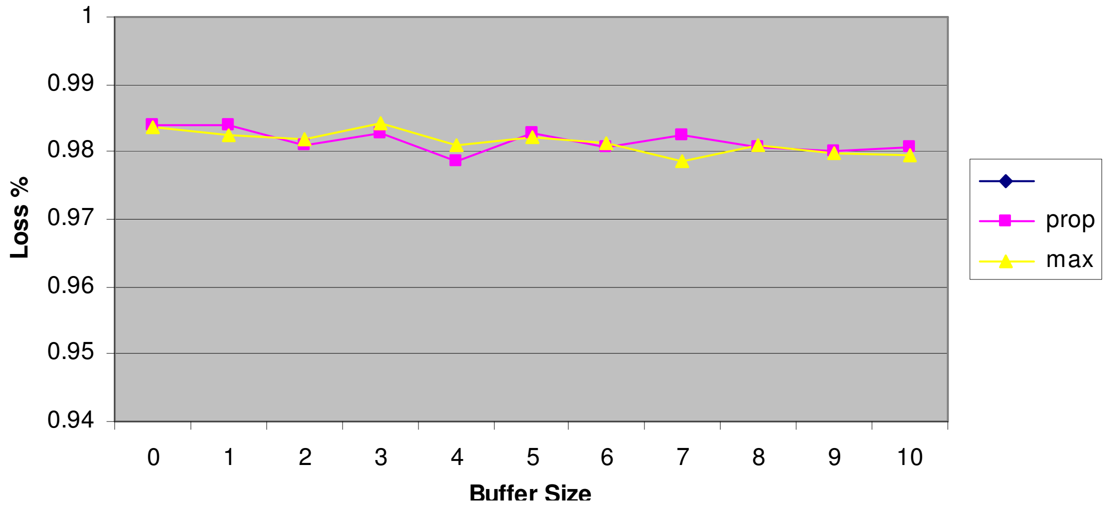


Graph 7.5.6 TEN USERS, OVERFLOW against BUFFER-SIZE

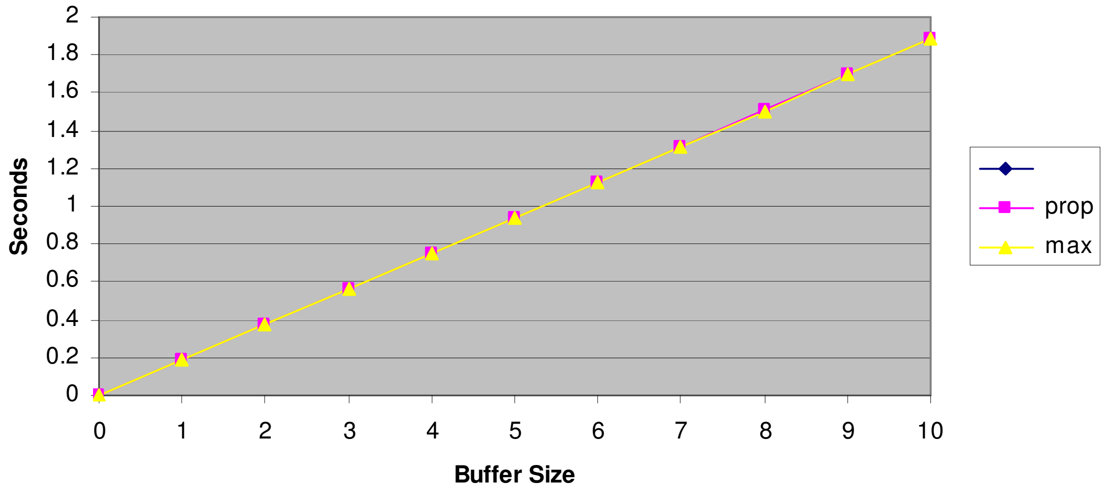


<u>For 100 USERS/CELL in Similar Channel Conditions</u>

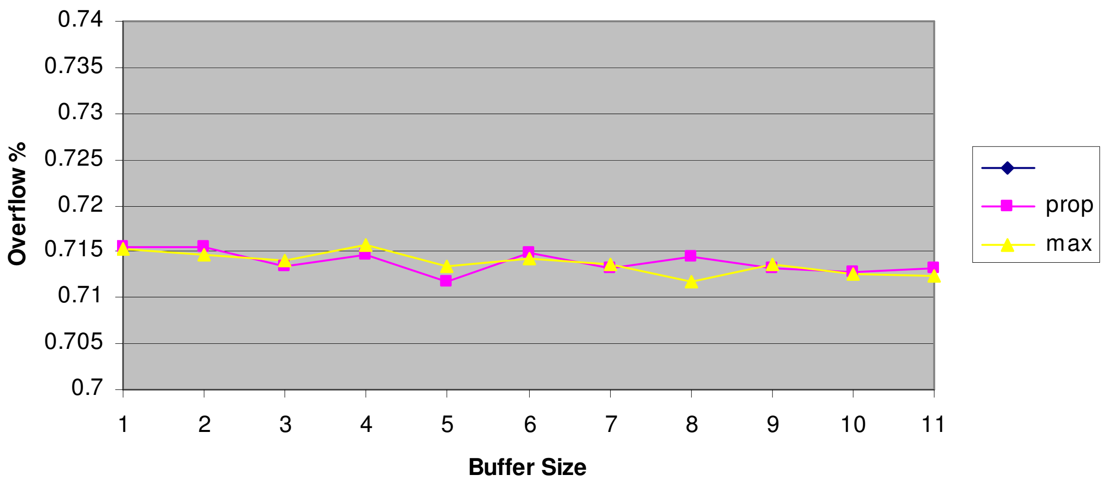
Graph 7.5.7 100 USERS, LOSS against BUFFER-SIZE



Graph 7.5.8 100 USERS, DELAY against BUFFER-SIZE



Graph 7.5.9 100 USERS, OVERFLOW against BUFFER-SIZE



Result Analysis

The scheduling techniques when run under varying channel conditions for different traffic types i.e. Interactive, Streaming, and a mix of both classes, it can be observed from the results, under the mix class the opportunistic and Max C/I techniques perform better then the PF. For scenario of Interactive traffic users the difference in performance of the three techniques does not come with much variation, however OPP still is found slightly to be better then Max C/I and PF. This is due to the low data rate required by the users which is easily provided by all of the three techniques when exploiting the channel variations. Here one interesting observation can be made i.e. in spite of the Opportunistic technique giving overall better result for loss and overflow the delay in comparison to the streaming and mix traffic class is proportionally a bit higher, probably because of the fact that the system has to check for the buffer level rather then for only the channel conditions for the opportunistic technique. Over all the delay difference is not much and doesn't raise any alarms. Experiments were carried out under five different scenarios each one is discussed below

In the first scenario one, five and ten users are scheduled for the steaming class for a range of buffer sizes up to 10 Mbits. In the first experiment for one user it is quite straight forward the system performs at its maximum value with only one user present to be scheduled. The performance for all the three techniques is found to be alike. For 5 users and 10 users a difference is experienced in the result in loss and overflow. In case of 5 users the PF technique shows the higher loss while the other two techniques show mixed results.

The second scenario is for interactive data traffic users also undertaken for one, five and ten users. This case is quite similar to the first scenario and here also opportunistic and Max C/I techniques are found to be better then the Proportional Fairness. With the only difference was that the gulf between the delay increased between the Opportunistic and the other two techniques.

The third scenario was attempted for five, ten and hundred users of the Mix of the two traffic classes i.e. interactive and streaming classes. In this case the opportunistic technique performed well with minor delays the worst was the PF, while Max C/I was found to be better then PF with some margin. This behavior was seen for all the three experiments. As the number of users increased it was seen that the values in the overall system performance also decreased producing comparatively more loss, buffer overflow and delays, which is understandable as the load on the system effected its performance

In the fourth scenario experiment performed was with the buffer-Size kept constant at 10 Mbits for a range of users up to hundred users. In this scenario all of the users belong to the Mix traffic class. This scenario shows the maximum number of users that can be satisfied with least loss of data contents. Clearly it can be observed that the opportunistic accommodates the most users with the least loss while the next best is Max C/I which gives a good competition. The number of users increase from forty the difference decreases. The PF algorithm gives the least performance in the loss and buffer overflow, but has less delay then its competitors.

The fifth scenario is quite different in a sense that it has been performed for users with similar channel conditions, rather then the previous scenarios where the channel conditions had high varying probabilities. In this case PF technique gives good competition to MAX C/I, rather outperforms it at times for different buffer sizes especially for the 100 users experiment, here each of the users belongs to the streaming traffic class. It can also be seen that the delay for the PF and MAX C/I are almost similar for this type of scenario.

The difference between loss and overflow probabilities is that the loss probability is the ratio of the mean loss rate and the mean offered rate. The overflow probability occurs when the queue length becomes greater than the buffer length. It has also been observed that the loss probability has been less than the overflow probability through out the simulations performed, this represents good channel conditions.

When we look at the performance of each algorithm separately, It has been observed that OPP technique has more Delay then the PF and MAX?
The time taken to serve (clear) the current buffer is called the waiting time
OPP experiences more delay because it has fairness characteristic to deliver the minimum throughput, and schedule the time rather fairly to all the users.
OPP technique while scheduling takes into account the Buffer levels along with the channel conditions, as delay or waiting time means the time needed to clear the contents of the buffer, so while scheduling the users with higher buffers are given more attention then the ones with lower buffer levels. Hence at the time when many of the users are busy transmitting it takes more time to clear the contents of the buffers completely. This increases the overall waiting time for the system, when compared to the other two scheduling techniques.

It has been observed that PF has more loss and overflow then MAX. This can be explained from the literature mentioned earlier in the thesis i.e. PF scheduler is fair (in terms of the distribution of the users' average throughputs) only in ideal cases where users experience similar channel conditions, and unable to exploit multi-user diversity in more realistic situations where users usually experience different channel conditions. The fifth scenario which is conducted under similar channel conditions depicts this situation. It can be seen that the PF technique shows better results in terms of loss and overflow and gets more competitive in comparison to the MAX technique getting better-quality at times.

Opportunistic technique is expected to guarantee minimum throughput and be fair in scheduling the users, it takes care of the buffer levels, resulting in minimizing loss and overflow of the buffer. In terms of the overall user satisfaction and fairness in scheduling the OPP technique shows the most fairness while MAXCI follows up,
From the fourth scenario it can be seen that PF on accommodates least number of users under the realistic conditions of varying channel conditions. PF is expected to perform better when channel conditions experienced by the users are similar and hence increasing its Fairness towards scheduling.

The algorithms used in this thesis have different complexities i.e. it was clearly seen that MAX CI was the least complex, and so was fast in process huge amounts of data (USERS). Further more MAXCI used the least amount of memory relating closely to the PF algorithm. This is because the nature of both these algorithms tends to be similar in a manner as both of these only use the channel condition as the input to the decision of scheduling. Where as OPP was the worst in case of complexity, either in processing large or small amounts of data, mainly because of its complexity in comparison to the other two algorithms and the large amount of memory it has to maintain. OPP algorithm hence was the hardest to implement.

CONCLUSION

In this thesis HSDPA functionalities have been studied and Different scheduler mechanisms have been analyzed and successfully simulated and evaluated. The model of the HSDPA system over which the schedulers have been simulated is implemented as a Channel model for channel simulation and a Data model for the generation of traffic, both of these combining to form the HSDPA system.
The channel model is put into practice as Finite State Markov chain model, which limited number of states (Channels) and a uniform state transition probability matrix. For the traffic simulation, On-Off traffic model was used for burst level traffic generation for users.
The Scheduling algorithms under study in this thesis are called fast scheduling algorithms; the algorithms utilize the channel conditions of users and need to make decisions every TTI (2ms) to better exploit fast variation of channel conditions.
These algorithms in the HSDPA system were simulated and experiments were performed under different traffic conditions on limited number of users against varying buffer sizes. The results of these experiments and the algorithm's performances were evaluated on the bases of congestion parameters i.e. loss, overflow and delay.
The conclusions drawn can be summarized as follows,

All the three algorithms analyzed are fast scheduling, each of the algorithms take into account the varying channel conditions while scheduling the users. In addition the OPP algorithm also takes care of the buffer levels of the users in the process of scheduling the timeline to the users.

The complexity of the MAXCI algorithm was the least and it was the fastest to execute. PF technique had the function of calculating the average CIR which increased its complexity and was rather slightly slower then MAXCI. OPP technique managed to be the most difficult technique of all the three, the fact that it had to take into account the buffer-levels as well, hence was the slowest of all.

In terms of performance OPP algorithm was seen to accommodate maximum number of users under varying channel conditions giving the least loss. MAXCI was the second best, while PF gave the least performance.
OPP technique largely experienced more delay in contrast to the other two techniques. This was due to the fact that it was fairer in terms of scheduling the users (as it took into account also the rising buffer levels of the users).

PF algorithm when implemented under steady channel conditions (similar channel conditions) experienced better results, and exploited multi-user diversity giving better performance at times better then MAXCI.
Under varying channel conditions OPP algorithm was the fairest as its technique involved quality constraints, followed by MAXCI and PF techniques respectively.

Bibliography

[1] Janne Pöllönen, Quality of Service Based Admission Control for WCDMA Mobile Systems, HELSINKI UNIVERSITY OF TECHNOLOGY Department of Engineering Physics and Mathematics year unknown

[2] Harri Holma, Antti Toskala. WCDMA for UMTS : Radio Access for Third Generation Mobile Communications (3rd Edition) John Wiley & Sons, Incorporated. 2005

[3] Radio-electronics:http://www.radio-electronics.com/info/cellulartelecomms/

[4] Ingo Forkel a,b, Hartmut Klenner a,b, Andreas Kempe, High Speed Downlink Packet Access (HSDPA)—Enhanced Data Rates for UMTS Evolution

[5] Qiu R.C; Wenwu Zhu; Ya-Qin Zhang ,Third-Generation and beyond (3.5G) wireless networks and its applications Journal: Circuits and Systems, 2002 ; ISCAS 2002; IEEE International Symposium , year: 2002 , vol: 1 , pages: I-41-I-44

[6] Arnab Das , Nandu Gopalakrishnan , Teck Hu , Farooq Khan , Ashok Rudrapatna, Ashwin Sampath , Hsuan-Jung Su , Said Tatesh , Wenfeng Zhang,Evolution of UMTS toward high-speed downlink packet access

[7] Khan Muhammad Sohaib, Nguyen Kim Cuong, HSDPA SYSTEM SIMULATION, Blekinge Institute of Technology (2005)

[8] Sem Borst, User-Level Performance of Channel-Aware Scheduling Algorithms in Wireless Data Networks.

[9] Ghassane Aniba, Student Member IEEE, Sonia Aka, Senior Member IEEE FASTP ACKET SCHEDULING ASSURING FAIRNESSA ND QUALITY OF SERVICE IN HSDPA, University of Quebec

[10] Fading: http://en.wikipedia.org/wiki/Fading

---

[11] Hussein Al-Zubaidy, Jerome Talim and Ioannis Lambadaris. Optimal SchedulingPolicy Determination for High Speed Downlink Packet Access. University, IEEE International Conference on Communications.

[12] Taesup Moon, Efficient modeling of flat fading channels, Stanford university

[13] Arsalan Farrokh, Member, IEEE, and Vikram Krishnamurthy, Fellow, IEEE Opportunistic Scheduling for Streaming Multimedia Users in High-Speed Downlink Packet Access (HSDPA)

[14] T. Bonald, A Score-Based Opportunistic Schedulerfor Fading Radio Channels France Telecom R&D

[15] Hua Fu and Dong In Kim, Senior Member, IEEE, Analysis of Throughput and Fairness with Downlink Scheduling in WCDMA Networks

[16] Pablo José, Ameigeiras Gutiérrez, Packet Scheduling And Quality of Service in HSDPA, Ph. D. Thesis,Department of Communication Technology, Institute of Electronic Systems, Aalborg University

[17] Bader Al-Manthari and Hossam Hassanein, Queen's University Nidal Nasser, University of Guelph, (2007), Packet Scheduling in 3.5G High-Speed Downlink Packet Access Networks: Breadth and Depth, , IEEENetwork.

[18] George S. Fishman: Discrete-Event Simulation: Modeling, Programming, and Analysis Berlin: Springer-Verlag 2001

[19] Discrete event simulation http://en.wikipedia.org/wiki/Discrete_event_simulation Wikipedia

[20] D. Anick, D. Mitra, and M.M. Sondhi, "Stochastic theory of a datahandling system with multiple sources," The Bell System Technical Journal, vol. 61, no. 8, pp. 1871 – 1894, Oct. 1982.

[21] Benyuan Liu, Daniel R. Figueiredo, Yang Guo, Jim Kurose, Don Towsley
A Study of Networks Simulation Efficiency: Fluid Simulation vs. Packet-level Simulation, Department of Computer Science University of Massachusetts

[22] Layo Olumide Babagbemi. (2006) Performance Analysis of Packet Schedulers, Department of Computer Science, Dalarna University, Sweden.

[23] Ernst Nordstrom, Ming Fan, Resource management in multi-servicenetworks: technical project description, Department of Computer Science, Dalarna University, Sweden.

[24] Simon Binar , HSDPA and HSUPA Functional Testing,Tektronix Inc, Protocol Monitoring Division, Berlin, Germany

[25] HSDPA in W-CDMA. http://www.umtsworld.com. (2006), The UMTS World.

[26] Wasif Iqbal , Rate Scheduling for high speed uplink packet access (HSUPA) in universal mobile telecommunication system (UMTS), (2007),Department of Computer Science, Högskolan Dalarna

27 http://cp.literature.agilent.com/litweb/pdf/5989-0390EN.pdf
Signal Studio for HSDPA over W-CDMA,E4438C ESG Vector Signal Generator

28 A. Jalali, R. Padovani, R. Pankaj, Data Throughput of CDMA-HDR a High Efficiency-High Data Rate Personal Communication Wireress System, Qualcomm, Inc.