3G HSDPA Performance In Mobile Internet Connections

XIAOXIN WANG

KTH Microelectronics
and Information Technology

Master of Science Thesis
Stockholm, Sweden 2004

IMIT/LCN 2004-02

3G HSDPA Performance In Mobile Internet Connections

Xiaoxin Wang

Master of Science Thesis
Performed at TeliaSonera AB, Mobile Network R&D

22nd March 2004

Examiner & Adviser:

Prof. Gerald Q. Maguire Jr.

Royal Institute of Technology

Industry Supervisor:

Doctor Anders Dahlén

TeliaSonera AB, Mobile Network R&D

TeliaSonera

Abstract

A key objective for 3G wireless networks, such as the Universal Mobile Telecommunication System (UMTS), is the explicit support for data communications for mobile users. Today, the dominant transport protocol in the Internet is the Transport Control Protocol (TCP). Since TCP was not tailored for wireless networks, there are some performance issues occurring when TCP traffic is transferred over a UMTS radio link.

In this paper, the characteristics of TCP and UMTS are specified and the problems of TCP over UMTS dedicated channels (DCHs) are analyzed. The problem is high delays implying low utilization of allocated resources for small file transfers or due to packet losses. High Speed Downlink Packet Access (HSDPA)'s potential to solve the problem is studied. Since in HSDPA, High-Speed Downlink shared channel (HS-DSCH) is introduced as an alternative of DCH on downlink packet access, a model of HS-DSCH is built and simulations are performed in order to compare its TCP performance with DCH. The focus is on studying retransmission delay and Block Error Rate (BLER) targets. Some scheduling methods are also compared. The results indicate that HS-DSCH gives better TCP performance than a DCH, and that advanced scheduling methods gives similar result as round robin if there are packet losses. Moreover, a somewhat surprising result is found regarding fast retransmission and channel utilization for increasing BLER targets.

Keywords: 3G, UMTS, TCP performance, HSDPA, HS-DSCH

Sammanfattning

I denna rapport specificeras karaktäristiken för TCP and UMTS, och problemet med TCP över dedikerade UMTS kanaler analyseras. Problemet är höga fördröjningar som medför låg utnyttjande grad av allokerade resurser vid överföring av små filer eller vid paket förluster. High Speed Downlink Packet Access (HSDPA)'s potential att lösa problemet studeras. En modell av HS-DSCH konstrueras och simuleringar utförs för att jämföra dess TCP prestanda med DCH. Fokus är på att studera återsändningsfördröjning och BLER riktvärden. Några scheduleringsmetoder jämförs också. Resultatet indikerar att HS-DSCH ger bättre TCP prestanda än DCH, och att avancerad scheduleringsmetoder ger liknande resultat som round robin vid paket förluster. Dessutom har ett ganska överraskande resultat funnits kring snabb återsändning och kanal utnyttjande grad då BLER riktvärdet ökar.

Nyckelord: 3G, UMTS, TCP prestanda, HSDPA, HS-DSCH

Table of Contents

List of Figures

List of Tables

1. Introduction

A key objective for 3rd generation wireless networks, such as the Universal Mobile Telecommunication System (UMTS)*, is the explicit support for data communications for mobile users. Today, the dominant transport protocol in the Internet is the Transport Control Protocol (TCP). Many important applications such as file transfer, Hyper Text Transfer Protocol (HTTP), web services, and E-mail use TCP. TCP is a generic protocol designed to promote *stability* in heterogeneous environments; hence, it is not specifically tailored to wireless networks. Therefore it is important to understand how TCP-based applications will behave if a wireless access link with certain special characteristics is involved in the transmission path. (See [5] and [6].)

In UMTS most services are provided over dedicated channels, which mean that each user has a radio link allocated solely to this user. The UMTS operator sets a block error ratio (BLER) target for the different services. The BLER target determines the average power level to be used by the radio link. For speech services, the BLER is usually 1%. This means that 1% of the radio blocks are received with errors after radio transmission. A 10% BLER is intended to be used for best effort services, such as Internet connections. It has been argued that, for these services, since the radio link retransmits the erroneous radio blocks, a higher BLER can be accepted despite increasing delay and delay variations. When the BLER is increased, the average power level for the radio link decreases. By lowering power, more users can simultaneously be served per base station, in another word, using a high BLER reduces the investment cost of building a 3G network.

The problem is that higher BLER leads to more packet retransmissions, which increases the delay. Long delays imply that TCP will send at a bit rate lower than the maximal bit rate of the radio link for long periods of time, for example, during the initial slow start and as a result of congestion control [1]. Hence, the utilization of the system resources will be poor when small files are transferred or when there are packet losses. Thus the gain by using 10% BLER will not be as high as expected, furthermore the customer may think that the performance is bad.

Today, a lot of research is being performed on the topic of TCP over wireless networks. Much research has focused on modifying TCP, of these the most promising solution to the TCP over UMTS problem might be TCP Westwood [4] together with a more aggressive initial slow start. The main problem with TCP modification is that it will take a long time for wide deployment and operators have little means to force Internet service providers to adopt new TCP versions. Moreover, TCP Westwood is not yet accepted as a

* Appendix B lists all the abbreviations need in this thesis.

standard by IETF[1]. Other proposals are split TCP [6] and radio bearer switching [2]. Split TCP is a rather controversial solution because it splits the TCP connection and breaks TCP's end-to-end semantics, whereas radio bearer switching does not increase the bit rate, but only increases the radio resource utilization.

In this thesis, a UMTS enhancement called High Speed Downlink Packet Access (HSDPA) is studied. HSDPA consists of a shared channel with a shorten retransmission delay. The performance of this radio channel is compared to the performance of dedicated channels. The BLER target and the retransmission delay impact as seen in TCP's performance are evaluated. Moreover, the effect of sharing channels is studied. In HSDPA scheduling is an important feature, therefore several scheduling methods are compared.

The performance evaluation is carried out through simulations in NS2. Hence, a model of UMTS and HSDPA has been implemented. Some simulations have been performed with packet loss and some with small files.

[1] IETF: The Internet Engineering Task Force. http://www.ietf.org/

2. TCP congestion control policy

The Transmission Control Protocol (TCP) is a transport protocol providing reliable end-to-end connections for Internet applications. TCP handles retransmission of packet loss and delay as well as prevents network congestion. (See [8] and [9].)

In order to provide error-free-guaranteed data delivery, a TCP sender numbers the segments in sequence before sending them. TCP corrects packet loss on the IP or Link layers using a sliding window and a retransmission strategy. TCP sends data segments according to a sender window and the receiver replies for each segment with a receipt based on a cumulative acknowledgement (ACK) policy. Since the receiver expects packets to come in sequence, it sends out so-called duplicated acknowledgements if a packet is received, but this packet has a higher sequence number than expected. Thus, a duplicate acknowledgement (DUPACK) indicates that a packet is missing, but one or more following packets got through, hence congestion was not the cause. The sender window limits the number of segments that can be sent, and additional segments cannot be sent before the ACK of the first segment in the send window is received. Lost packets are retransmitted when TCP's retransmission timeout occurs. Nowadays, TCP also incorporates fast retransmission, wherein duplicate acknowledgements (3 or more DUPACK) trigger immediate retransmission.

Congestion control is a very important role of the TCP protocol in TCP/IP protocol suite. TCP starts each connection with a low data rate and increases the rate as acknowledgements arrive, until packet loss is detected. The limit of the transmission rate is controlled by the congestion window size (cwnd). TCP can have at most cwnd bytes of data outstanding[2]. Initially, cwnd is set to a very small value. The congestion window will be increased by one SMSS (Sender Maximum Segment Size) for every received acknowledgement. This is so-called slow start, which increases the transmission rate exponentially (See Step 1 of Figure 1).

As soon as a packet loss is detected, network congestion is assumed. Therefore the congestion window is decreased, which leads to a reduced transmission rate. There are two ways to detect packet loss: (1) the retransmission timer expires, in this situation, the congestion is assumed to be very severe, and the cwnd is set to one SMSS; then the slow start algorithm will be used, which increases cwnd exponentially, until the cwnd equals

[2] The congestion window (cwnd) is a sender-side limit on the amount of data the sender can transmit into the network before receiving an acknowledgment (ACK), while the receiver's advertised window (rwnd) is a receiver-side limit on the amount of outstanding data (the receiver buffer size). The minimum of cwnd and rwnd is the send window, which governs data transmission.(RFC 2581 [8])

or exceeds ssthresh[3] (See, Steps 4, 5 and 6 of Figure 1), then the congestion avoidance algorithm is used, which increases cwnd linearly; (2) The other is the so-called fast retransmission method. It assumes that a packet is lost if some (usually 3) duplicate acknowledgements for a single segment are received. In this situation, the congestion is assumed less severe since packets following the missing packet have arrived. The congestion window is halved and the transmission rate is increased according to congestion avoidance algorithm. (See Steps 2 and 3 of Figure 1)

The send window limits TCP throughput. The maximal send window size is also limited by the receiver's advertised window size, since the sender window size is determined by the minimum of a congestion window and the receiver advertised window.

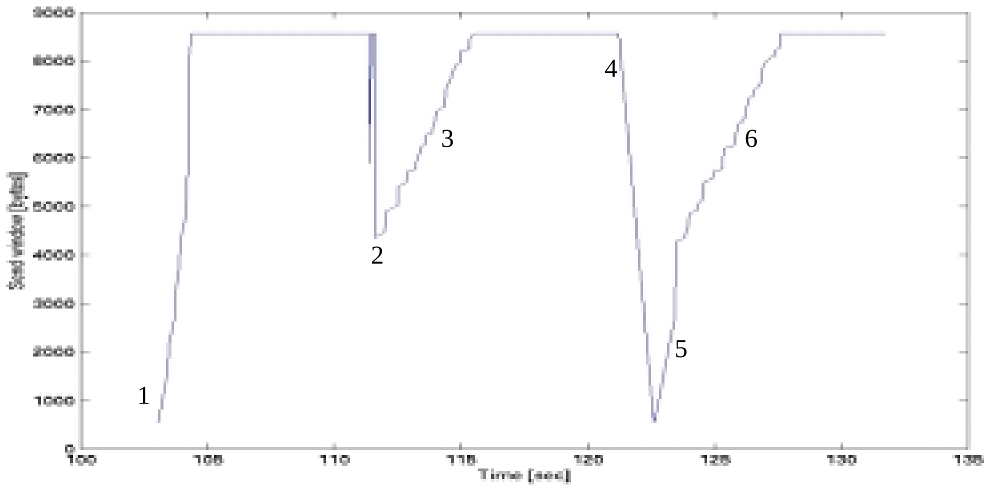

Figure 1. TCP Send Window, 6 steps: 1. initial slow start; 2. fast retransmission event; 3. congestion avoidance window recovery after a fast retransmission; 4. timeout event; 5. slow start recovery after a timeout; 6. congestion avoidance after the slow start recovery.

[3]The slow start threshold (ssthresh), is used to determine whether the slow start or congestion avoidance algorithm is used to control data transmission.

3. UMTS

The Universal Mobile Telecommunication System (UMTS) is a third generation (3G) mobile telecommunication system. It is defined by 3GPP (the 3rd Generation Partnership Project) and provides high bit rate data communications to the mobile user.

3.1. UMTS Architecture

The UMTS system architecture is shown in Figure 2.

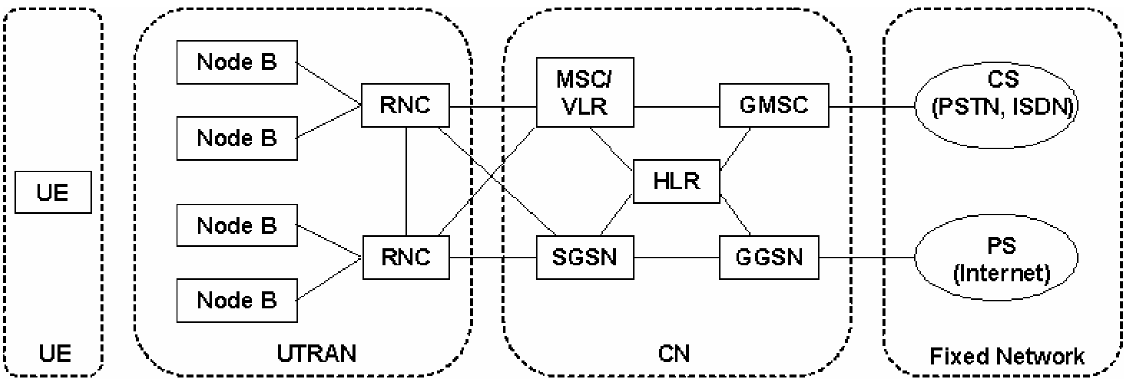

Figure 2. The Architecture of UMTS

The UMTS network consists of three parts: Core Network (CN), UMTS Terrestrial Radio Access Network (UTRAN), and User Equipment (UE). The CN handles routing in the system and stores user information. Since UMTS can be connected to both channel switched (CS) and packet switched (PS) external network, the CN needs to support both of these switching techniques. The connection between the external networks and CN goes through a Gateway MSC (GMSC) in the CS case, and a Gateway GPRS (GGSN) in the PS case. Furthermore, CS services are routed through a network of Mobile Services Switching Centers (MSCs), and the PS services are routed through a network of Serving GPRS Support Nodes (SGSNs).

The UTRAN consists of a Radio Network Sub-system (RNS), which in turn consists of a Radio Network Controller (RNC) and one or more Node B. Each RNC is connected to both an MSC and an SGSN. The main purpose of the RNC is to manage the radio resources using the Radio Resource Control (RRC) protocol. The Node B is usually called a base station and is the link between user equipment (UE) and the UMTS network.

3.2. Radio Link Control

The Radio Link Control (RLC) layer in the RNC has a very high impact on TCP performance. The RLC layer performs segmentation of IP packets into radio frames, and causes retransmissions of these frames to provide a reliable link.

To prevent spurious TCP timeouts, the RLC retransmissions should be performed much faster compared to the retransmission timeout of the TCP layer. The BLER target gives an indication of the probability for having several retransmissions. With a 10% BLER target, spurious timeouts will be rare (as verified in our simulations, see section 9.1). However, in practice, it is hard to avoid pathological cases completely [6].

In the RLC layer, configuration of in-order and out-of-order delivery of received IP packets up to higher layers can be chosen. When out-of-order delivery is used, the retransmissions of the RLC layer can also cause packet disordering, which is a critical issue for TCP. Out-of-order segments trigger duplicate acknowledgements, and if reordering of **three** segments occurs, an unnecessary retransmission is initiated, which not only wastes network resources, but also reduces TCP throughput by decreasing the cwnd window by half the window size. This problem is avoided by using only in-order delivery. However, in-order delivery increases the delay, since packets have to wait for delivery until all previous packets have been received without errors. Since spurious timeouts will be rare, using in-order delivering is recommended for TCP traffic.

4. The problem of TCP over UMTS Release 99

4.1. DCH

In UMTS release 99, a so-called dedicated channel (DCH) is used for the downlink and uplink packet access. A DCH is dedicated to a single user and is subject to fast power control. It can support different maximal bit rates. The higher the maximal bit rate is, the more resources must be reserved for this user. Thus, for a high bit rate burst of data, there will be a waste of resources if the channel is not released immediately after a single burst of data has been transmitted. On the other hand, the allocation and the release of the channels cause unwanted control signaling and unwanted delays [21]. Moreover, resources might not be available when a subsequent burst arrives.

4.2. The problem analyses

If in-order delivery is used in the RLC, a high round trip time (RTT) together with a relatively high radio link bit rate is the main factor for TCP's performance over UMTS. Unfortunately, the bit rate of the dedicated radio link is poorly utilized.

We can analyze this problem using TCP's Bandwidth-Delay Product (BDP):

$$\text{BDP (bits)} = \text{bandwidth (bps)} * \text{RTT (s)} \qquad (1)$$

- Bandwidth represents the bandwidth of the limiting part of the TCP connection
- RTT is the end-to-end round trip time

Theoretically, the congestion window should be as close as possible to this product, because a lower window implies poor utilization of the limiting link (generally the radio link), whereas, a larger congestion window implies more data than the limiting link can handle which leads to congestion (assuming that the advertised window is larger than the BDP). In practice, due to delay variations, the congestion window should be 1-2 times the BDP so as not to limit the throughput in the path.

A paper written by a researcher from Aalborg University and a researcher from Nokia [15] analyses the BDP for UMTS, see Table 1, and the number of end-to-end RTT cycles that are required to reach different BDPs, see Table 2. Table 2 considers the congestion window evolution during slow start, which is:

$$\mathbf{cwnd}\left(n \cdot RTT + \frac{2^{n-1}-1}{BW}\right) = 2^n \qquad (2)$$

- n is the cycle number. A cycle refers to the period since a segment is transmitted until its corresponding acknowledgement reaches the transmitting entity; and
- BW represents the radio link capacity (bandwidth) measured in IP datagrams per second. Then, 1/BW is the average time to transmit an IP datagram over the radio link.

Table 1. BDP EXAMPLES FOR SOME CHANNEL BIT RATES AND RTTs

Radio Channel Bit Rate (Rb)	Round Trip Time (RTT)		
	50 ms	*150ms*	***300ms***
32kbps	0.16kbytes	0.48kbytes	0.64kbytes
64kbps	0.32kbytes	0.96kbytes	1.92kbytes
128kbps	0.64kbytes	1.92kbytes	3.84kbytes
256kbps	1.28kbytes	3.84kbytes	7.68kbytes
384kbps	1.92kbytes	5.76kbytes	11.52kbytes

Table 2. NUMBER OF REQUIRED CYCLES TO REACH BDP

MSS	536Bytes		1460Bytes	
RTT	*150ms*	*300ms*	*150ms*	*300ms*
Rb = 32kbps	1	2	1	1
Rb = 64kbps	2	3	1	2
Rb = 128kbps	3	4	2	2
Rb = 256kbps	4	4	2	3
Rb = 384kbps	5	6	3	4

Table 1 shows the BDP for common bit rates of UMTS DCH when the RTT is 50ms, 150ms, and 300ms, respectively. While 50ms RTT might be realistic for fixed network connections, 200-300ms RTT is more realistic for UMTS connections. Table 2 gives the number of required end-to-end RTT cycles to reach the BDP when the RTT is 150ms and 300ms.

Table 2 shows that maximal radio link bit rate is reached earlier with a shorter RTT or with a larger SMSS. For a dedicated channel with a 384 kbps channel rate, it takes more than 3 cycles (3 cycles @ 150ms is 450ms!) to reach the maximum link utilization. Therefore, it is obvious that TCP’s slow start algorithm causes low utilization of system resources.

Small file transmissions can cause rather low system utilization. For instance, when a file with a size larger than 1500 bytes is transferred with 1460 bytes SMSS through a dedicated channel with 384kbps and 300ms RTT, the system utilization will be only 12% during the first RTT (since the BDP is 11.52kbytes and the TCP send window is 1.5kbytes, i.e. 1.5/11.52 =12%). If the file size is smaller than 1500 bytes, the system utilization can even be worse. Even when a file of 22.5kbytes in size is transferred, theoretically, the system will never reach the maximal utilization since as soon as the sender reaches the maximal send rate, the file transmission has finished. For this 22.5kbytes file, the channel utilization is only 50%.

When the BLER of the radio link is high, for example 10%, the RLC layer will use retransmissions to correct block errors, which will introduce even longer delay. Long delay implies that the TCP congestion window increases more slowly and that BDP increases as well. Hence, it will take a longer time for TCP to utilize the maximal bit rate of the radio link.

A formula for the end-to-end RTT over UMTS is given in [5], and simplified in [6] where it is written as:

$$\mathbf{RTT_{end\text{-}to\text{-}end} = k{*}RTT_{RLC} + 2{*}(Internet\ delay + CN\ delay)} \qquad \mathbf{(3)}$$

- where $k \approx 3$ for 10% BLER and $k \approx 1.5$ for 1% BLER.

The formula indicates that the radio bearer's BLER target affects the average RTT to a great extent. For an IP packet, a high BLER target results in many radio link retransmissions and each retransmission, RTT_{RLC}, takes about 100 ms [5]. The Internet delay is usually less than 50 ms and the core network delay is probably less than 20 ms[6]. Thus, for 1% BLER, RTT ≈ 1.5*100 + 2*(50+20) ≈ 300 ms, and for 10% BLER, RTT ≈ 3*300+2*(50+20) ≈ 450 ms. We can see that for 10% BLER, the $RTT_{end\text{-}to\text{-}end}$ is much longer than the one for 1% BLER. Therefore, it takes a longer time to reach the maximal radio link bit rate with 10% BLER. In the formula, 1% or 10% BLER in both downlink and uplink are assumed.

5. UMTS Release 5 and HSDPA

5.1. HSDPA

WCDMA Release 99 provides dedicated channels (DCHs) with data rates of 384 kbps for wide area coverage and up to 2 Mbps for hot-spot areas, which is thought to be sufficient for most existing packet-data applications (Although for small files, these rates are only ~10-20 % of the link rate). However, as the use of packet data services increases and new services are introduced, perhaps greater rates will be required.

The concept of High Speed Downlink Packet Access (HSDPA) was introduced in UMTS Release 5. HSDPA is designed to increase packet throughput on the downlink, as an evolution of DSCH (which is a shared channel introduced in Release 99 but rarely deployed). HSDPA consists of a new High Speed Downlink Shared Channel (HS-DSCH), which is shared among several users. The aim of HSDPA is to achieve significantly higher peak data rates than the 2 Mbps possible in Release 99 of UMTS, namely 8-10 Mbps. Moreover, HSDPA should be more efficient than Release 99 UMTS. Thus it should also give a capacity increase. This means that each cell can transport a larger number of bits per second. The operator is thus able to support more users per cell even if each user gets the same average data rate as in Release 99. Alternatively, with a fixed number of users, each gets a higher data rate with HSDPA than with Release 99.

However, can HSDPA solve TCP's performance problem over UMTS? In this chapter, some investigations of what other researchers have done in this area are described and following this, a discussion of how HSDPA helps to solve the problem is presented.

5.2. The results and conclusions of recent research

In recent research on HSDPA, positive results were achieved. HSDPA has been shown to improve the performance of TCP over UMTS and solve the problem of low utilization of system capacity when TCP traffic is sent over UMTS radio network. (See [17] and [18])

In reference [17], both the performance of TCP based applications and the effects of TCP on system capacity using HS-DSCH are evaluated. It is claimed that the introduction of HS-DSCH increases both the application performance for TCP based applications and the system capacity compared with using dedicated channels.

According to reference [18], through appropriate management of radio resources (scheduling), the new high-speed packet data bearer is shown to perform well with TCP for WWW services.

5.3. Why HSDPA solves the problem

The RTT over HS-DSCH will be shorter than over dedicated channels. The reduction in delay is achieved by reducing the amount of interleaving and by not using macro diversity for HS-DSCH. The interleaving delay of 2 ms (3 slots) is significantly lower than the minimal interleaving delay of 10 ms for dedicated channels. The reduced interleaving delay reduces processing delays in various nodes, resulting in a significant reduction in the total delay.

By not using macro diversity, retransmissions can be performed completely independently by Node B. This results in faster retransmissions. The ARQ process is also enhanced with soft combining using hybrid ARQ, which will reduce the number of retransmissions needed [17].

Because of faster retransmissions, lower delay and the possibility of better scheduling, we believe that with HS-DSCH it should be possible to mitigate many of the problems that TCP introduces on dedicated channels. The possibility of better sharing of resources (i.e. the effect of multiplexing many burst channels) alone should improve channel utilization since when one user’s bit rate is low, other users may utilize the extra available data rate.

6. Goals

The aim of this thesis is to understand the performance of TCP traffic over HS-DSCH via simulations and to quantify the performance relative to dedicated links. Hence, a model of HS-DSCH and DCH need to be created for these simulations.

How the new features of HS-DSCH affect the TCP performance can be one of the most important aspects when comparing to the TCP performance of DCH. The lower retransmission delay will be a feature that we will focus on.

The flexibility of sharing a channel will also be studied, thus different scheduling methods will be considered.

7. HS-DSCH

HSDPA introduces a new transport channel called the High-Speed Downlink shared channel (HS-DSCH). This is a shared channel, as the name indicates it transports data from the network to the terminal. A certain number of the channel codes and a certain amount of transmission power in a cell are allocated to the shared channel and these resources are dynamically shared among users, primarily in the time domain (i.e., time shared). Shared-channel transmission makes more efficient use of available codes than always assigning dedicated channels.

HSDPA also supports the following features [19]:

- Fast link adaptation – The transmission power is kept constant for the HS-DSCH, but the bit rate is adapted to the varying radio conditions (by varying code rate and modulation scheme), whereas DCH is designed to maintain constant bit rate by varying the power level to the radio conditions (through a power control loop). The method in HS-DSCH is commonly referred to as link adaptation; this method is more efficient than power control for services that can tolerate short-term variations in the data rate.
- Fast hybrid-ARQ – erroneously received data can rapidly be retransmitted to the terminal. The retransmission delay will be ~10 ms compared to around ~100 ms on a DCH. Prior to decoding, the terminal combines information from the original transmission with that of later retransmissions. This is called soft combining and it increases the probability of successful decoding.
- Fast channel-dependent scheduling – The scheduler determines, at any given moment, to which terminal the shared channel transmission should be directed. The basic idea is to exploit short-term variations in radio conditions by transmitting to terminals with favorable instantaneous channel conditions. It is claimed to greatly increase capacity and to make better use of resources [19].

7.1. Architecture

The UTRAN architecture in release 99 is illustrated in the figure below. The RNCs are connected to the core network. Each RNC controls one or more Node Bs, which in turn communicate with the user equipment (UE).

The features of fast link adaptation, fast hybrid-ARQ, and fast channel-dependent scheduling, require some new functionality in the UTRAN architecture. These features require rapid adaptation of the transmission parameters to time-varying channel condition and, therefore, the corresponding functions should be placed as close to the air interface as possible. Hence, a new entity, MAC-hs, which supports the new functionalities of HSDPA, is added to Node B.

The new features of HSDPA in Node B should not be seen as a replacement of the RNC, but rather as a complement to the RNC. The old RNC entities are kept.

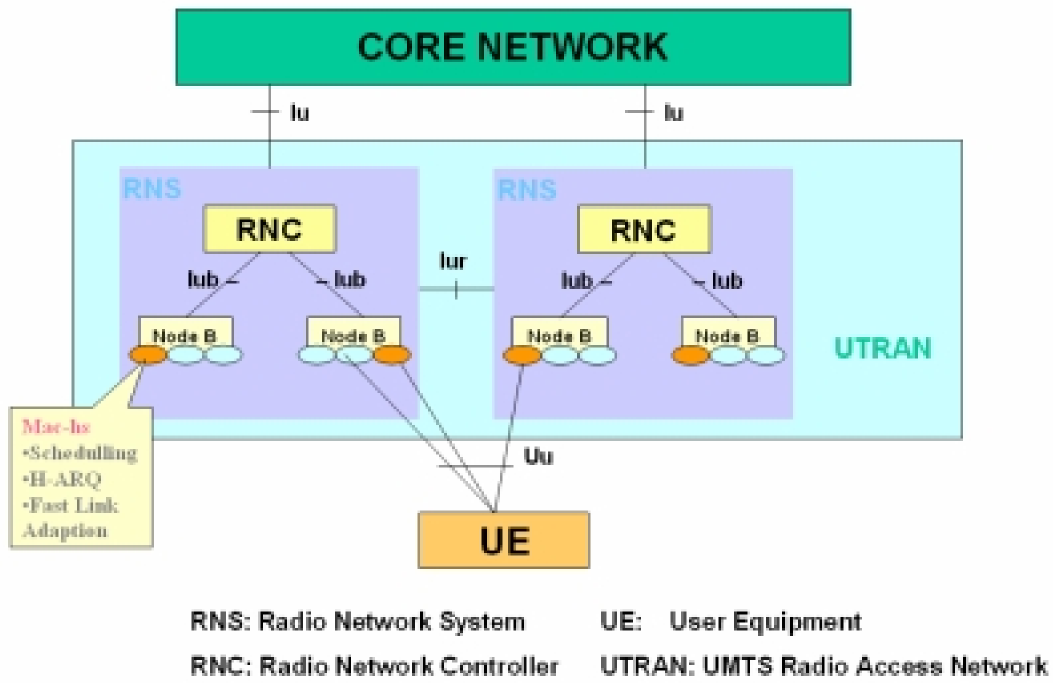

Figure 3. The UTRAN Architecture

7.2. Channel structure

The HS-DSCH can share a carrier with other DCH channels, as illustrated in the figure below. Therefore, there is no need to allocate additional spectrum in order to introduce HSDPA services. This is an efficient use of spectrum and allows flexible allocation of resources between services using HS-DSCH, such as packet-data applications, and services using DCH, such as voice.

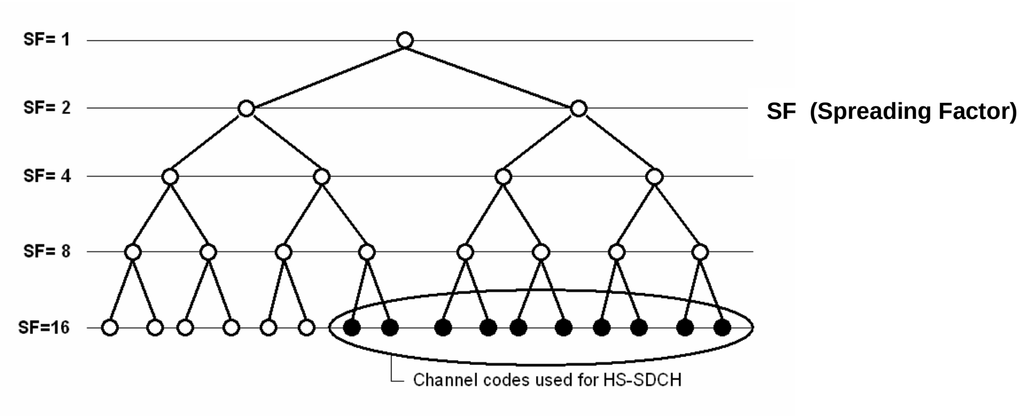

Figure 4. Multiple codes of spreading factor 16

HS-DSCH corresponds to a common code resource shared among several users, primarily in the time domain. Therefore, the allocation of the code resources in HS-DSCH is primarily based on one transmission time interval (TTI), which means the shared code resource is dynamically allocated by the radio base station (RBS) every TTI. Compared with the DCH TTI of 10ms, HS-DSCH uses a shorter TTI of 2 ms. Using a shorter TTI is supposed to reduce overall delay and to improve the tracking of channel variations, which are exploited by link adaptation and channel-dependent scheduling.

In HSDPA, a new downlink control channel, the high-speed shared control channel (HS-SCCH), is implemented to carry control information (such as identity of the terminal, hybrid-ARQ-related information and the parameters of the HS-DSCH transport format selected by the link-adaptation mechanism) from the MAC-hs in the RBS to the scheduled terminal.

The MAC layer uses ACKs (or NACKs negative acknowledgements) to request retransmissions. These ACK messages are sent on a new uplink control channel, called the high-speed dedicated physical control channel (HS-DPCCH). It enables retransmissions in approximately 8-12 ms.

7.3. Scheduling

The scheduler controls the allocation of the channel among the users. It is one of the functions of the HS-DSCH MAC layer and is implemented in the Radio Base station. As a key element, a scheduler determines the overall system behavior in the long run.

The functionality of a scheduler is to determine, for each TTI,

- Which terminal the common code resource should be assigned to;
- The link adaptation mechanism; and
- At what data rate the radio base station (RBS) transmits downlink data to the terminals.

When the scheduler prioritizes transmissions to terminals with favorable instantaneous channel conditions, a significant increase in capacity can be obtained. Thus, statistically the network experiences mostly good conditions. But with increasing user diversity and hence increasing system diversity, user satisfaction **deceases** since the fairness of radio resource assignment cannot be satisfied (compared to allocating resource sequentially). Therefore, a practical scheduling strategy exploits the short-term variations while at the same time maintaining some degree of long-term fairness between users, in another word, giving up some system throughput to increase fairness. Therefore, a trade-off should be reached.

There are three type of schedulers that can be used. They are C/I, Round Robin, and Proportional Fair scheduling algorithms [18]. Each is described briefly below.

Carrier-to-interface ratio (C/I)) scheduler directs transmission to the user with the momentarily best channel conditions, allowing for the highest possible data rate at each instant and thus maximizing the overall throughput. However, the price paid for high overall throughput is potentially large variations in service quality among the user population.

The round robin scheduler cycles through the list of active users and thus is fair in an average sense. However, as the round robin scheduler does not base its decision on the varying channel quality, the throughput performance suffers.

The proportional fair scheduler schedules the user with the currently highest ratio between instantaneous C/I and average transmission rate. This algorithm takes advantage of the short-term channel variations while at the same time preventing zero “long term” throughput (starvation) for any user.

7.4. Link adaptation and higher-order modulation

By introducing variable-rate turbo encoding, such as 16QAM modulation, as well as extensive multi-code operation, the HS-DSCH supports peak bit rates from 120 kbps up to and beyond 10 Mbps. The basic adaptive modulation and coding (AMC) process has a dynamic range on the order of 20dB, further expanded by the available number of multi-codes. This is typically sufficient to track the channel quality experienced at the UE(terminal). Table 3 shows the link between a possible transport format and resource combination and the corresponding user peak bit rate.

Table 3 . Possible Transport Formats and the Corresponding User Peak Bit Rates

Modulation/ Coding	1 code	2 codes	5codes	10 codes	15codes
QPSK ¼	120kbps	240 kbps	600 kbps	1.2 Mbps	1.8 Mbps
QPSK ½	240 kbps	480 kbps	1.2 Mbps	2.4 Mbps	3.6 Mbps
QPSK ¾	360 kbps	720 kbps	1.8 Mbps	3.6 Mbps	5.3 Mbps
16QAM ½	480 kbps	960 kbps	2.4 Mbps	4.8 Mbps	7.2 Mbps

16QAM ¾	720 kbps	1.44Mbps	3.6 Mbps	7.2 Mbps	10.7Mbps

7.5. Hybrid-ARQ with soft-combining

The hybrid-ARQ mechanism allows the terminal to rapidly request retransmission of erroneously received transport blocks, essentially fine-tuning the effective code-rate and compensating for errors made by the link-adaptation mechanism. The terminal attempts to decode each transport block it receives, reporting to the radio base station its success or failure within 5ms after the reception of the transport block. The hybrid-ARQ mechanism in the radio base station can thus rapidly respond to retransmission requests. During retransmission, the terminal employs soft-combining – that is, it combines (soft) information from previous transmission attempts with the current transmission (retransmission) to increase the probability of successful decoding.

8. Simulation Model

8.1. Introduction

A simulation model of the HS-DSCH was created to evaluate the possible performance improvements that HSDPA gives compared to DCH. The model and the parameters are designed carefully with 3GPP specifications and other research papers as sources of information. The model allows tests to be conducted without actually implementing the proposed solutions in a real system. A model is by definition a simplification and an abstraction of the reality, thereby the results from the experiments involving the model should only be seen as a guideline. We did some verification of DCH, which is mentioned in section 9.1. Since there are no equipments of HSDPA available, the verification of HSDPA of this model is out of this master thesis scope.

8.2. Modelling HS-DSCH

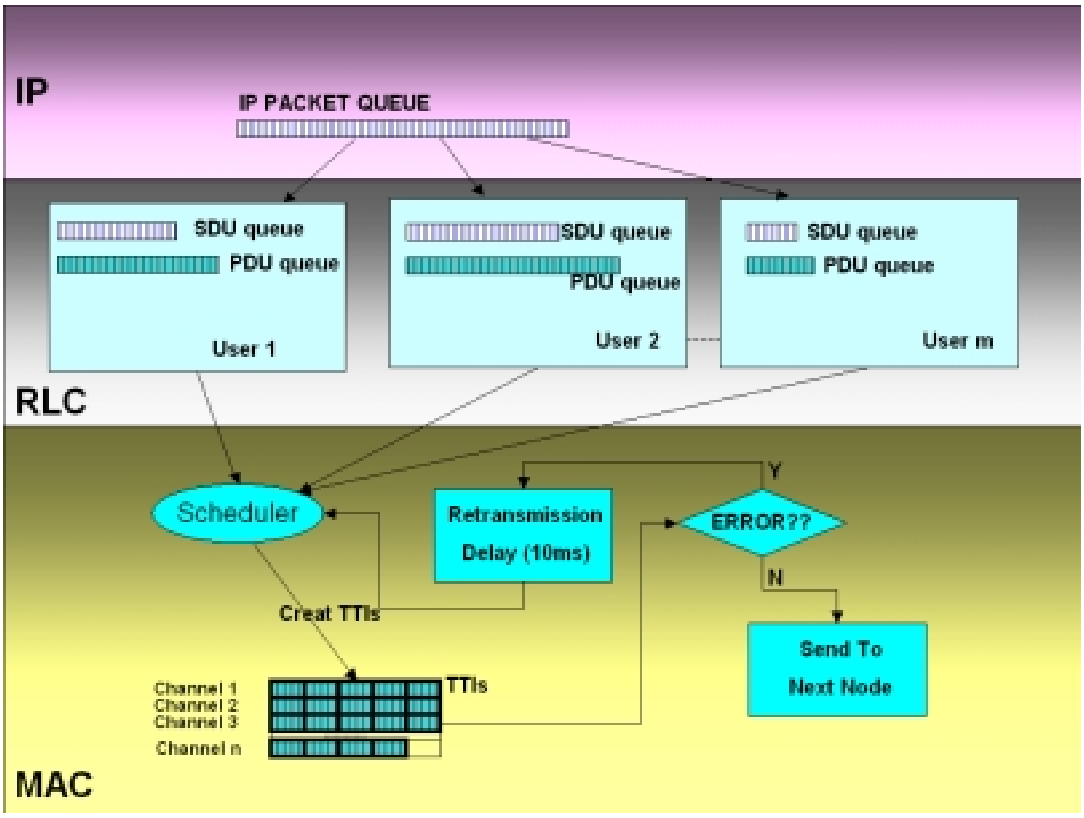

Figure 5. Modelling of HSDPA

In the HS-DSCH primarily the RLC and MAC layers were modeled. The HS-DSCH model consists of one MAC object and a RLC object for each user. We focus on modeling delays and bit rate of the channel, rather than protocol details, since these parameters are the most important factors for TCP performance.

As we mention in the section 7.1, HSDPA relies on some new functionality in each Node-B. Therefore, in our model, most of the functions are implemented in the MAC model while the RLC model is rather simple. The RLC model simply acts as a buffer for different terminals. When the radio base station receives IP packets from an upper layer, it will send the packets to the different RLC objects according to the destination IP address. The RLC layer segments IP packets into radio blocks, which are called Service Data Units (SDUs) and Protocol Data Units (PDUs) respectively. There are no retransmissions in this layer.

The MAC layer is in charge of creating TTIs, which are composed of PDUs from the upper layer (the RLC layer). HSDPA uses a rather small TTI compared to DCH in order to reduce the overall delay and to improve the tracking of channel variations. Error checking and retransmission will be taken place in the MAC with the hybrid-ARQ mechanism of HSDPA.

With a smaller TTI, the retransmission delay will be much shorter than the retransmission delay of the DCH. In our model, the hybrid-ARQ mechanism implementation lacks soft combining. This is called basic H-ARQ-type-I, which is the simplest retransmission method in HSDPA [22].

A TTI is determined to be erroneous or not by drawing a uniformly distributed random number between 1 and 100. If the random number is less than a constant "target number", this TTI is considered as erroneous, and is scheduled for retransmission 10ms later. The errors of the following TTIs are independently generated[4]. This error determination scheme is similar to a perfect BLER target power control of the DCH. Therefore, we will call this "target number" the BLER target, even though the errors are at the TTI level instead of on a RLC radio block level (which is the standard definition of the BLER target of the DCH)[5].

No control channels were modeled and no MAC acknowledgements or negative acknowledgement (NACKs) are sent back to the Node B. We do **not** model the possibility that NACKs are erroneously interpreted in the Node B (which would lead to unnecessary RLC retransmissions). In any case, such events should be rare.

Scheduling is a key function of the MAC layer. It determines to which terminal the radio resource is assigned and at what bit rate the channel transmits data. Different schedulers are implemented in our HS-DSCH model in order to compare their affects on users' throughput. The channel bit rate can be constant or varying according to a very simple "radio model", described in the next section.

A packet can be sent to the next node in-sequence or out-of-sequence. In our model, both of these methods are implemented and can be chosen at run-time for a given simulation. However, obviously, in-sequence should be chosen because out-of-sequence causes many unnecessary TCP fast retransmissions due to packet reordering as we mentioned in the section 3.2.

8.3. Assumptions and Descriptions

Parameters and details must also be specified in addition to the description of how the HS-DSCH is modeled. The assumptions made in this context are important and they must be handled with care in order to realize an accurate model. We believe we have succeeded fairly well in this by carefully choosing parameters according to information in network manufacturers' research papers and 3GPP specifications. Some of the most important attributes of the model and the surrounding area are described in this section.

[4] This way of modelling the radio block error generation is very common in other research papers.
[5] To have a BLER target on the RLC layer on HS-DSCH make little sense, since the MAC layer should perform essentially all retransmissions.

In the model, a shorter TTI (compared to DCH) of 2ms is used. This is one of the key characteristics of HSDPA, which can greatly reduce the overall delay. The hybrid-ARQ mechanism allows the terminal to rapidly request retransmission of erroneously received transport blocks. The terminal attempts to decode each transport block it receives, reporting to the radio base station its success or failure within 5ms after the reception of the transport block. Therefore, we assume that a TTI can be retransmitted 10ms after its decoding failure. This retransmit delay is much shorter than the DCH delay, which is assumed to be 89 ms.

In the MAC layer, we assume that an unlimited number of retransmissions can occur, thus there will be no RLC layer retransmissions. In real life, usually there are limited numbers of MAC layer retransmissions and RLC does retransmission if the MAC retransmissions fail[6].

H-ARQ has two possible schemes: Selective Repeat and Stop and Wait. The model does not explicitly adopt any of these techniques, but simply does an unlimited number of retransmissions where each retransmission has a fixed delay. The window (the number of radio blocks sent but not yet acknowledged) is unlimited.

HS-DSCH is a shared channel. In our model, we only consider time-sharing since it is the primary means of sharing common resources among users. Therefore in each TTI, only one user can be assigned the radio channel.

In this model, three schedulers were implemented. They are:

- Round Robin: The scheduler cycles through the list of active users and gives a fair sharing of the resource.
- Max C/I: The scheduler directs transmission to the user with the momentarily best channel conditions, allowing for the highest possible data rate at each instant and thus maximizing the overall throughput.
- Proportional Fair (PF): The scheduler schedules the user with the currently highest ratio between instantaneous C/I and average transmission rate.

The latter two schedulers need C/I values to make decisions. In another word, the C/I value should be known before the decision is made. In our model, we generate a C/I value for each user each TTI. When a terminal is moving within a cell, it will have different C/I values with the probabilities in the diagram below (Figure 6). These probabilities are based on the information in [21]. The C/I value is mapped to a certain bit rate using table 4.

[6] One reason for this assumption is that no RLC retransmission needed to be implemented. According to [22], a limit of 6 retransmissions gives the same result as unlimited retransmissions.

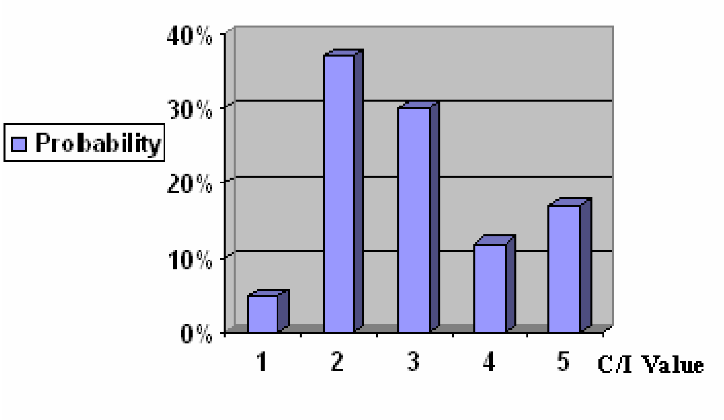

Figure 6. C/I Value Probability

Table 4. Map of C/I values and bit rates of a terminal

C/I Values	Bit Rate Speeds
1	120 Kbps
2	240 Kbps
3	360 Kbps
4	480 Kbps
5	640 Kbps

We use the same model to simulate the DCH channel, but with other parameter settings; see the table below.

Table 5 . Parameters of HS-DSCH & DCH

Parameters	HS-DSCH	DCH
TTI	2 ms	10 ms
Retransmission Delay	10 ms	89 ms
PDU size	30 Bytes	40 Bytes
Channels	Variable	12
PDUs in each TTI	Variable	1
Users	Variable	1

8.4. Simulation Topology

The topology of the simulation is very important, since it affects the simulation results a lot. We choose a simple fixed network topology because a clearer picture of how the HSDPA affects TCP performance can be shown without the interference of the factors of the fixed network.

The topology comprises three parts as shown in Figure 7. The first part is the rightmost, which represents a server residing on the Internet. The second part is the node in the middle, which represents a base station (Node B). (Actually, the function of Node B is implemented in the HS-DSCH link.) The third part is composed of those small nodes on the leftmost and one node (X node) between those small nodes and Node B. This part represents some UEs (user equipment, or terminal). The X node could be viewed as a part of the UE's physical layer that determines if the data is for this UE or not.

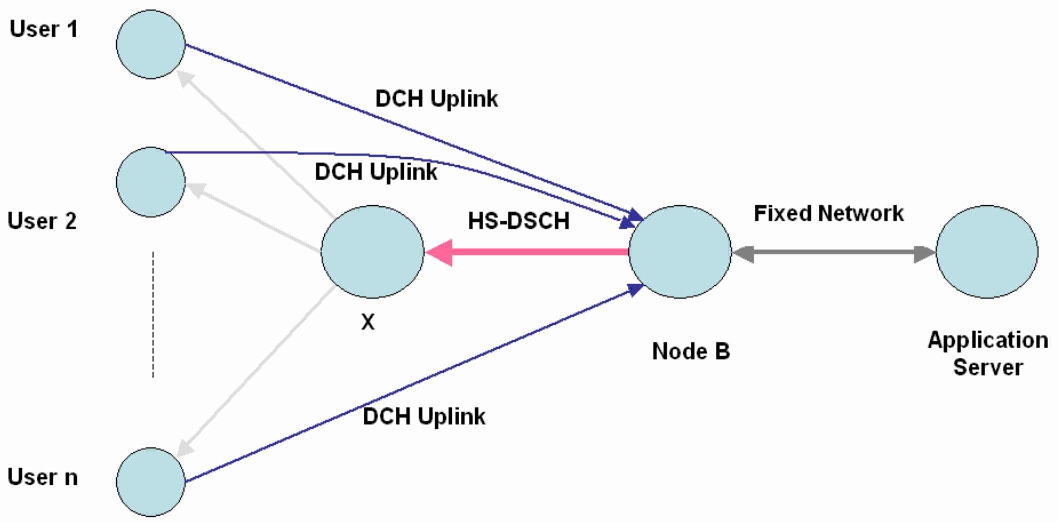

Figure 7. Simulation Network Topology

Table 6. Settings on the network topology

Link	Bit Rate	Delay	Packet Loss
Fixed network	10 Mbps	110 ms	0,1 and 3 %
DCH Uplink	64 kbps	20 ms	0 %
HS-DSCH	Variable	Variable	0 %

From Node B to the user equipment, is the High Speed Downlink Shared Channel (HS-DSCH). Since this is a shared channel for all the end users, we use one node (X node) to represent all of the end users. Each terminal has a dedicated uplink. In the simulation, this link carries download requests from the terminal and TCP acknowledgements. The uplink has no block errors and hence, no retransmissions.

9. Simulations and Results

All the simulations are performed using our extension to NS2 [23]. Three groups of simulations were done. The first group concerns the BLER targets affect on RRTs for HS-DSCH and DCH; since the RTT affects the TCP performance a lot (as we analysed in the section 5.2). The second group concerns the TCP performance on HS-DSCH compared to DCH. Comparisons are made between two kinds of links using an ftp application with different file sizes and different numbers of users. The third group considers how different schedulers affect HS-DSCH performance from the view of customer satisfaction.

Section 9.1 examines the simulations of the first group. Section 9.2, 9.3 and 9.4 examine the simulations of the second group. Section 9.5 examines the simulations of the third group.

9.1. Simulation results of RTT

The TCP layer in NS2 can directly measure the roundtrip time if the maximal TCP window size is set to one packet and the TCP clock granularity is high. This is done for a DCH and a HS-DSCH with 1% and 10% BLER targets. See the table below for detail parameter settings:

Table 7. The parameters When measured the RTTs

	DCH	HS-DSCH
Bit rate [kbps]	384	384
Retransmission delay [ms]	99	9.5
TTI [ms]	10	2
PDU size [byte]	40	32
Number of PDUs in a TTI	1	1
Number of channels	12	3
Packet size [bytes]	1500	1500
TCP max cwnd [packets]	1	1
TCP clock granularity [ms]	1	1
ACK delay at receiver [ms]	0	0
File size [kB]	150	150

The average RTTs obtained in the simulations are:

Table 8. The results of the Measurement of RTTs.

RTT [ms]	1% BLER	10% BLER	Difference [ms]
DCH	264	346	82 (+31%)
HS-DSCH	248	256	8 (+3%)

From these results it is clear that the fast retransmission of H-ARQ has a large impact on the end-to-end round trip time as the BLER target increases. The difference between using 10% and 1% BLER target for the DCH is 82 ms, but only 8 ms for the HS-DSCH. Apparently, the difference between the two links is a factor 10, just like the retransmission delay.

Note that the uplink DCH in these simulations is error free. In reality it could lead to additional delay if its BLER target is 10%, but this additional delay would be the same when comparing downlink DCH with HS-DSCH.

RTTs of 250 ms have been observed for a DCH with 1% BLER in a live network. Hence, the measured delays in the simulation model should be realistic for low BLER targets. For high BLERs, recall the RTT model in section 4.2, which indicates a150 ms increase in average delay if downlink and uplink BLER target increases from 1% to 10%. Here we get an 82 ms extra delay (i.e. 31%) when increasing the BLER in the downlink only.

Live network measurements for HS-DSCH cannot be done, since no HSDPA system has been implemented yet.

9.2. HS-DSCH performance compared with DCH

As we mentioned in section 6.3, HSDPA can be a solution to TCP's problems over UMTS. Therefore, some simulations of TCP performance on DCH and HS-DSCH were performed in order to verify that HSDPA can help improve TCP's performance.

The same parameters were used in all the simulations from this section (9.2) to the end of this chapter (i.e., Section 9.5) unless there is some special indication.

Table 9. TCP Parameter Settings

Packet size [bytes]	1500
TCP version	NewReno
TCP initial cwnd [packets]	2
TCP max cwnd [packets]	42
TCP ssthresh [packets]	40
ACK delay at receiver [ms]	100

Table 10. Radio channel Parameter Settings

Parameters	DCH	HS-DSCH
Bit rate [kbps]	384	384 ; 360
Retransmission delay [ms]	89	9.5
TTI [ms]	10	2
PDU size [byte]	40	32 ; 30
Number of PDUs in a TTI	1	1
Number of channels	12	3

Additionally, in-order delivery was used for the reason mentioned in section 4.2 [2]. The BLER and the file size are also two very important parameters. They are not specified here because they will be different in the different simulations.

Simulations with file sizes of 16M bytes, 100k bytes, 50k bytes, and 20k bytes were performed on both DCH (384 kbps) and HS-DSCH (384 kbps and 360 kbps). The reason we use two different speeds of HS-DSCH is that 360Kbps is reasonable in practice while 384kbps is more comparable to DCH.

The formulas we use to calculate the results are:
(1) Application Bit Rate (kbps) = Application data / Time
(2) Retransmit Bit Rate (kbps) = [File size $_{\text{(with IP/TCP Header)}}$ / Time] / (1- BLER)
(3) Channel Utilization (%) = Retransmit Bit Rate / Channel Bit Rate

- Time (s) is the average download time of the FTP file.

9.2.1 The results for 16M byte files

Simulations of 16M byte file downloads were performed in order to see the TCP performance over UMTS when large files are transmitted. Here we only tested the situation when there is no packet loss on the Internet part of the path. In the topology shown in section 8.4, this means that there will not be any packet loss on the fixed network.

The following two figures and table show the simulation results when the file sizes are 16Mbytes.

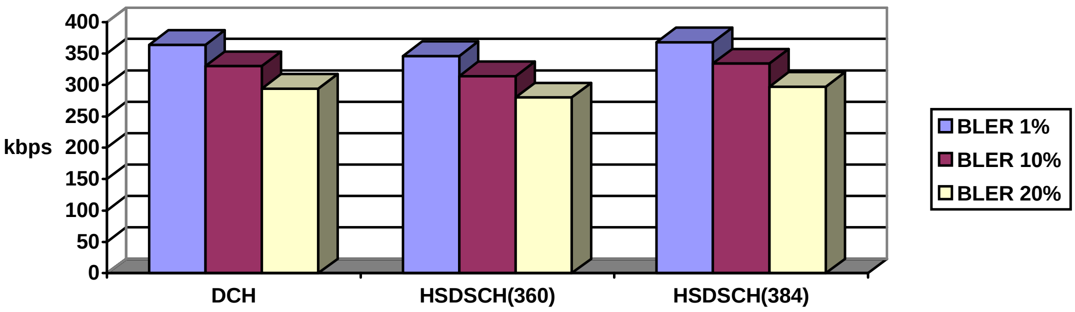

Figure 8. Application Bit Rates with a file size of 16M bytes

When the BLER increases, the application bit rates decrease; this is because there are more radio link retransmissions as the BLER increases. Note that no TCP retransmissions have been observed.

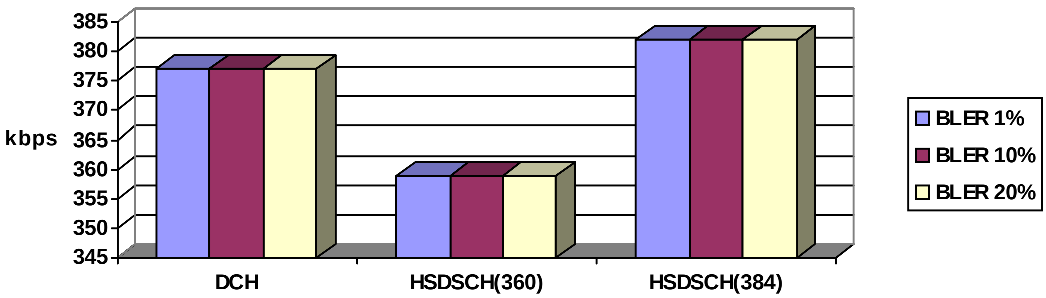

Figure 9. Retransmit Bit Rates with a file size of 16 Mbytes

Table 11. Channel Utilization for transfers of 16 Mbytes (%)

BLER (%)	**DCH (384 Kbps)**	**HS-DSCH**	
		360 Kbps	**384Kbps**
1	98.3	99.7	99.4
10	98.2	99.7	99.4
20	98.3	99.8	99.4

When the BLER increases, the retransmit bit rates does not change at all. Hence, from the channel utilization point of view, there is nearly no waste of resource. This is because: when a large file is transmitted, the wasted channel utilization because of slow start is rather small compared to the whole transmission time.

From these two figures and the table, we can see that when a large file is transmitted over UMTS, either using DCH or HS-DSCH, the performance shows little difference. They nearly reach the maximum bit rates.

Therefore, when there is no packet loss on the Internet part of the path, there is no significant TCP performance problem when large files are transmitted over UMTS.

9.2.2 The results for 100k byte files

For this section, some simulations of medium sized files were performed in order to investigate TCP's performance for medium sized files over UMTS and to see if HS-DSCH can improve TCP's performance for medium sized files compared to DCH. Downloading files of 100k bytes, 50k bytes, and 20k bytes were simulated using our model. Here we only tested the situation when there is no packet loss on the Internet part of the path. Each download was executed 10 times and the results listed below are the average values of these 10 results.

The following figures and table show the simulation results when the file sizes were 100kbytes.

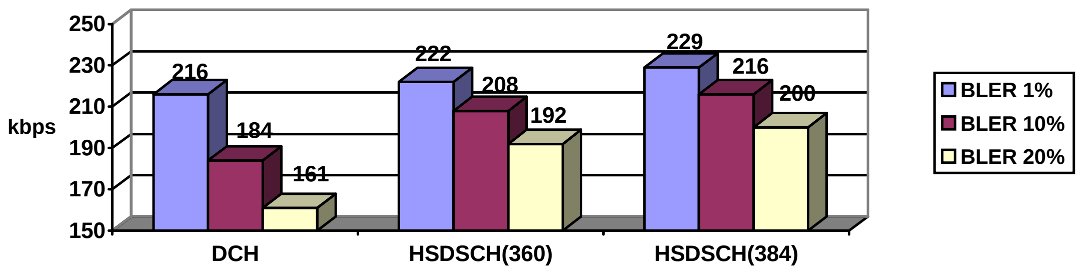

Figure 10. Application Bit Rates with a file size of 100k bytes

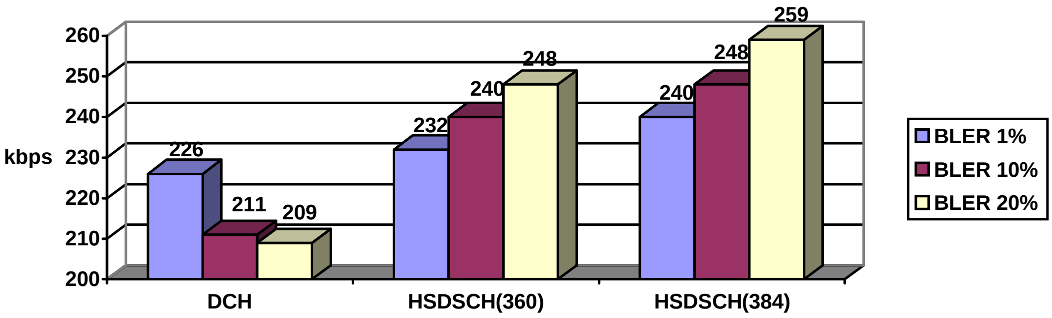

Figure 11. Retransmit Bit Rates with a file size of 100k bytes

With file sizes of 100k bytes, for both DCH and HS-DSCH, the application bit rate decreases when BLER increases. However, the application bit rate of HS-DSCH decreases more slowly than for DCH. Therefore, from an end user's point of view, there are some improvements by using HS-DSCH. Moreover, while the retransmit bit rate of DCH decreases with increasing BLER, the retransmit bit rate increases for HS-DSCH.

Table 12. Channel Utilization for transfers of 100k bytes (%)

BLER (%)	**DCH (384 Kbps)**	**HS-DSCH**	
		360 Kbps	**384Kbps**
1	59	64	62
10	55	67	65
20	54	69	68

The channel utilization for a 100k byte file is only 50% -70%. Compared with the result for 16M bytes, we see that medium size files have more of a TCP performance problem than large files (as we expected).

The result is that the channel utilization of HS-DSCH increases with increasing BLER is even greater than expected. This will be examined in greater detail in the simulation of transfers of 50kbyte and 20kbyte files.

9.2.3 The results for 50k byte files

The simulations of downloading with a file size of 50k bytes, which is half the size of the 100k-byte files, are performed here. The aim of this simulation is to show how the application bit rates and retransmit bit rates change when the BLER increases. Understanding the utilization of the channel is another goal of these simulations since it will be compared with those of larger and smaller files.

The following figures and table show the simulation results when the file sizes are 50kbytes.

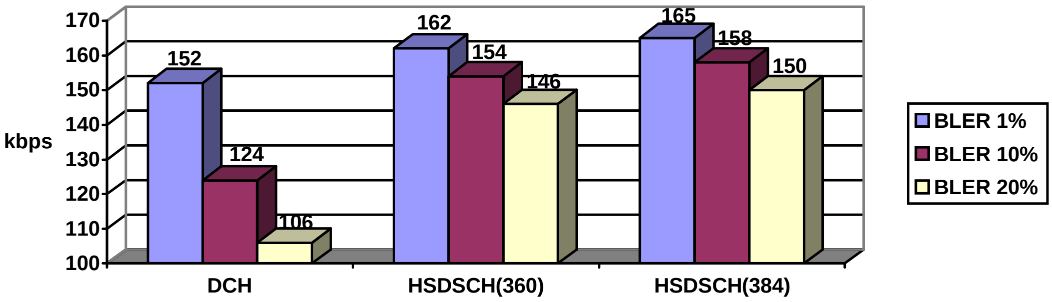

Figure 12. Application Bit Rates with a file size of 50k bytes

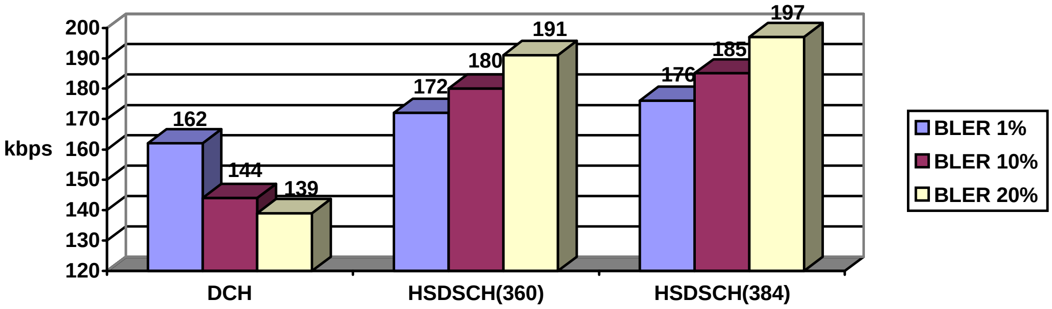

Figure 13. Retransmit Bit Rates with a file size of 50k bytes

Similar to the simulation results of the 100kbyte file, when BLER increases, for both DCH and HS-DSCH, the application bit rate decreases. Furthermore, the application bit rate of HS-DSCH deceases more slowly than those of DCH. Therefore, we have the same conclusion that, from an end user's point of view, there are some improvements by using HS-DSCH.

Table 13. Channel Utilization for transfers of 50k bytes (%)

BLER (%)	**DCH (384 Kbps)**	**HS-DSCH**	
		360 Kbps	**384Kbps**
1	42	48	46
10	38	50	48
20	36	53	51

The channel utilization for transfers of 50k byte file is only 35% - 55%. Compared with the result for 100kbyte files, we can find that the smaller the file size is, the more severe the TCP problem over UMTS is.

9.2.4 The results for 20k byte files

The simulation of a 20kbyte file transfer is very interesting since in our analyses in section 4.2, we mentioned that: if the file size is less than 22.5kbytes, theoretically, the channel utilization can **never** reach the maximum bit rate. From the simulation, the severity of this problem can be quantified.

The following figures and table show the simulation results when the file sizes are 20kbytes.

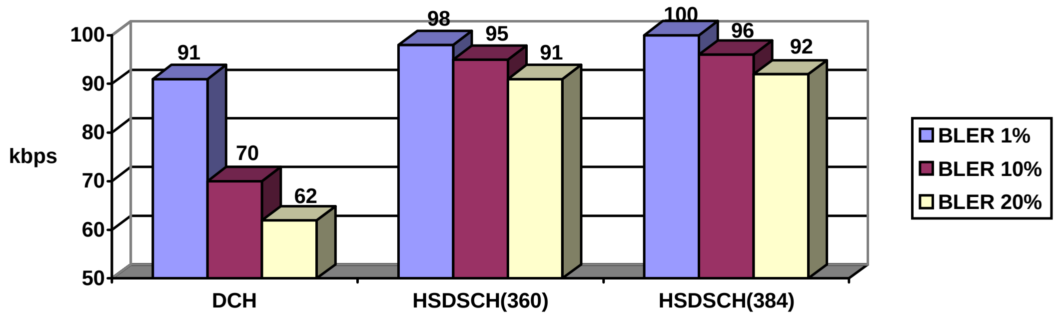

Figure 14. Application Bit Rates with a file size of 20k bytes

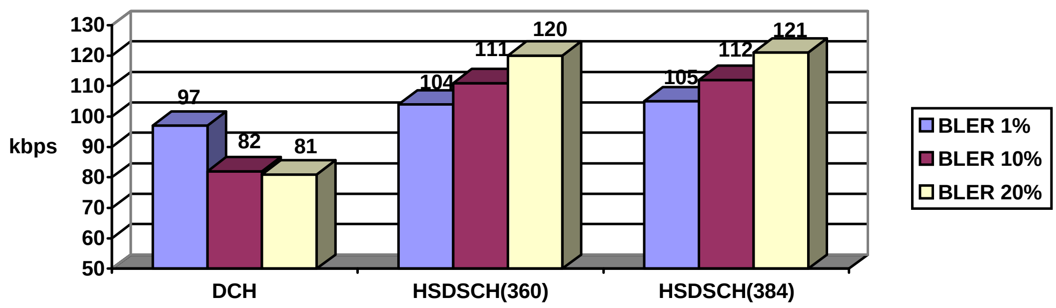

Figure 15. Retransmit Bit Rates with a file size of 20k bytes

Similar to the simulation results for the transfer of 50kbytes and 100kbytes files, for both DCH and HS-DSCH, the application bit rates decreases when BLER increases. Additionally, the application bit rate of HS-DSCH deceases more slowly than for DCH. Again, we reach the conclusion that, from an end user's point of view, there are some improvements for HS-DSCH.

Table 14. Channel Utilization for transfers of 20k bytes (%)

BLER (%)	**DCH (384 Kbps)**	**HS-DSCH**	
		360 Kbps	**384Kbps**
1	25	29	28
10	21	31	30
20	21	33	32

The channel utilization for 20kbyte file transfer is only 20% - 32%. Compared with the result for 50kbyte file, we confirm that the smaller the file size is, the more severe the TCP problem over UMTS is.

9.2.5 Discussion of the results

The small file simulations are very interesting since the TCP performance problem over DCH is more severe than for large files.

In general, we find that there are some improvements when using HS-DSCH compared to DCH, especially when the BLER is higher. When the file size is 20k bytes, the improvement is significant.

We find that the channel utilization for HS-DSCH **increases** as BLER increases. This is better than our expectation. We think these improvements are due to the shorter retransmission delay since this is the main difference between DCH and HS-DSCH in this set of simulations. Therefore, further simulations with different delay were performed to verify our analysis.

Here we have assumed that the RAN delay is the same for HS-DSCH and DCH. However, the small TTI for HS-DSCH will probably result in a significant decrease of the RAN delay. Moreover, if soft combining had been implemented, the difference in performance when increasing BLER would be even smaller. Hence, HS-DSCH will probably perform better in practice than the result indicates here.

9.3. Simulation of 50Kbytes transfer with Different Delays

In order to understand whether it is the retransmission delay that affects the increasing/decreasing of channel utilization as BLER target is increased, we examined transfer of 50k byte files and changed the retransmission delay of the HS-DSCH link. 89ms and 50ms were chosen to compare with the simulations with normal HS-DSCH retransmit delay (9.5ms) and the simulations of DCH (where the retransmission delay of DCH is 89ms.)

Table 15. Channel utilization with Different Delays

Channels with different Retransmission Delay / BLER(%)	DCH (384 Kbps)	HS-DSCH (384Kbps)		
	89ms	89ms	50ms	9.5ms
1	42%	43%	44%	46%
10	38%	39%	44%	48%
20	36%	37%	44%	51%

From Table 15, we find that, as the retransmit delay increases on the HS-DSCH link, the difference between HS-DSCH and DCH becomes smaller. When using 89ms retransmission delay, the channel utilization for HS-DSCH decreases as BLER increases, which is the same as DCH; in this case HS-DSCH shows no improvement over DCH. When using a 50 ms retransmit delay, the channel utilization for HS-DSCH remains stable as BLER increases, thus HS-DSCH shows some improvement over DCH. Using a delay of 9.5ms, we have as much improvement as we had in the previous simulations in section 9.2 and the channel utilization increases as the BLER increases.

Therefore, the results verify our earlier hypothesis that, it is the shorter retransmission delay that causes HS-DSCH's channel utilization increases as BLER increases. In the DCH situation, the channel utilization decreases as BLER increases. This is can be considered to be an improvement of HS-DSCH over DCH.

9.4. DCH & HS-DSCH performance comparison with 3 Users

In order to test if there are any performance differences between a shared channel and a dedicated channel, we simulate 3 users using the channel at the same time. For DCH , the 3 users will use different channels; each user has his own 384kbps channel. In HS-DSCH case, the 3 users share the same channel with a speed of 1152kbps (3 times 384kbps).

In this simulation, there is 1% packet loss on the Internet part of the path. Each user downloads a 16M bytes file at the same time. The purpose for setting 1% packet loss on Internet is that: when there is packet loss on Internet, either TCP fast retransmission or TCP timeout will occur, thus, the cwnd window will either be halved or start from 1 SMSS respectively. In the case of DCH, the channel that experiences packet loss will have low utilization. In the case of HS-DSCH, the other users can use the extra available bit rate.

Table 16. The average values of 3 users when BLER is 1%

Channels	Application Bit rate (kbps)	Re-Tx bit Rate (kbps)	Channel Utilization
DCH	270.33	280.51	0.73
HS-DSCH	306.76	318.32	0.79*
Difference	13%	13%	8%
* Using the finishing download time of the last user to calculate utilization			

Table 17. The average values of 3 users when BLER is BLER 10%

Channels	Application Bit rate (kbps)	Re-Tx bit Rate (kbps)	Channel Utilization
DCH	214.96	245.37	0.64
HS-DSCH	292.94	334.38	0.85*
Difference	36%	36%	33%
* Using the finishing download time of the last user to calculate utilization			

From these results, it is clear that HS-DSCH has better channel utilization than a dedicated channel when there are packet losses, especially when BLER is high. When BLER is 1%, the difference between the channel utilizations is 8% while when BLER is 10%, the difference is 33%.

Furthermore, we can see that the end users of HS-DSCH have greater satisfaction than those of DCH. Because, the average application bit rate of the former is 13% greater than the latter when BLER is 1% and when BLER is 10%, the different is even larger (36%).

From these two points, we can conclude that sharing can help improve channel utilization. On the other hand, we should see that the faster retransmission of HS-DSCH is also reflected in the results. So the improvement for 1% BLER is mostly due to sharing, whereas at 10% BLER the fast retransmission has a large impact.

Another conclusion is, when there are packet losses on fixed network, HS-DSCH mitigates the TCP problem and therefore has some improvement over DCH.

9.5. Simulations of different schedulers

Since the schedulers need to use C/I value to make decision, before we simulate the schedulers, we should first simulate the possibilities of C/I values for the users. In practice, C/I value is dynamic. Therefore at each moment, the C/I value of a user is different. We generate a C/I value for a user each TTI. There are three user types (See Figure 16). User A is far away from the radio base station (RBS) and it is more likely to have lower C/I value. User A is near to the RBS and it is more likely to have higher C/I value. User B is in the middle of A and C and the probabilities of his C/I value are very near to Figure 6. The C/I value probabilities of the three types of users are listed in Table 18. These values in the Table 18 are generated according to the values in Figure 6, which was discussed in section 8.3 [21].

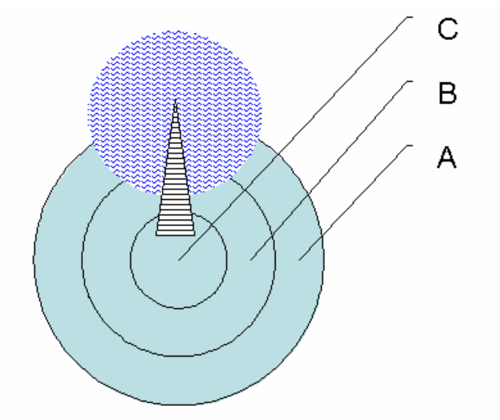

Figure 16. Users at the different areas of a cell

Table 18. C/I value Probabilities according to figure 15.

C/I Value	User A	User B	User C
1	9%	4%	2%
2	60%	32%	16%
3	24%	43%	23%
4	5%	9%	20%
5	2%	12%	37%

When we simulate 3 users, one is of Type A (User A), one is of Type B (User B), and one is Type C (User C). When we simulate 6 users, two are of Type A, two are of Type B, and two are of Type C. (With the indicated C/I value and respective bit rates)

The results of these simulations using different scheduling with 3 users and 6 users are list below. MVA is the mean value of 3 or 6 users' application bit rates. From this value, in some sense, we can find the user throughput when using different scheduling algorithms. STD is the standard deviation of 3 or 6 users' application bit rates, which indicates the difference of application bit rates among the 3 or 6 users. In some sense, STD shows the fairness of the resource assignment, which should be considered as an important criterion to qualify the different scheduling algorithms.

Flowing tables are the average value of the simulation results. The details about the simulation results are shown in Appendix A.

Table 19. MVA and STD of the Application Bit Rates for 3 users when using different schedulers

MVA: Mean Value of Application Bit Rate			
Scheduling	**Packet Loss on Internet**		
	0%	**1%**	**3%**
RR	365.27 kbps	282.91 kbps	155.56 kbps
MAX C/I	604.46 kbps	309.61 kbps	164.67 kbps
PF	486.32 kbps	288.81 kbps	164.55 kbps
STD: Standard Deviation			
Scheduling	**Packet Loss on Internet**		
	0%	**1%**	**3%**
RR	61.41	23.41	4.21
MAX C/I	293.26	57.63	13.80
PF	104.83	14.71	8.45

Table 20. MVA and STD of the Application Bit Rates for 6 users when using different schedulers

MVA: Mean Value of Application Bit Rate		
Scheduling	**Packet Loss on Internet**	
	1%	**3%**
RR	279.41 kbps	170.28 kbps
MAX C/I	315.70 kbps	169.63 kbps
PF	301.65 kbps	170.37 kbps
STD: Standard Deviation		
Scheduling	**Packet Loss on Internet**	
	1%	**3%**
RR	22.45	7.51
MAX C/I	31.64	8.10
PF	23.65	6.11

From the results, MAX C/I scheduler has the best system throughput while RR scheduling has the best fairness; PF is something in the middle. We can see an interesting result, that when packet losses on the fixed network increase, the difference between different scheduler is less and less. We think this is because: when packet loss increases, either TCP fast retransmissions or TCP timeout occurs, thus the cwnd decreases. This forces the end user who has the best C/I value to give up the channel to those with lower C/I value. Hence, the TCP window will limit the throughput rather than the scheduling of resources. When the packet loss is high, the speeds of all the users remain low. Therefore, the difference is rather small when the packet loss is 3%.

10. Conclusions and Future Work

10.1. Conclusions

The problems of TCP performance over a UMTS dedicated channel (DCH) have been both discussed and quantified in this thesis. HS-DSCH, which is introduced by HSDPA, is proposed as an alternative to DCH for data services. Both HS-DSCH and DCH were implemented in our model in order to compare their impact on TCP performance. Three groups of simulations were performed in order to quantify the potential enhancement that HS-DSCH can give. The analysis and discussion of the simulation results were presented, and from these, we now summarize our conclusions.

The RTT is a very important parameter that affects TCP performance a lot. The results of the simulations from the first group show that HS-DSCH has a shorter RTT compared to DCH especially when BLER is high. This means that by using HS-DSCH, TCP send window can increase faster than over a DCH, which improves TCP performance.

The results of the simulations from the second group show that the problems of TCP performance over a UMTS dedicated channel (DCH) are severe, especially when small files are transferred and the BLER is high. The faster retransmission of HS-DSCH reduces the problem in these situations (when small files are transferred and the BLER is higher). Furthermore, the fast retransmission can even imply that the channel utilization increases as BLER increases when medium or small files are transferred. This conclusion is drawn in the context that there is no packet loss on the fixed network and only one user uses the HS-DSCH (not considering sharing). Since soft combining is not implemented in the model, the fast retransmission is the main factor for the improvement. If soft combining had been implemented, the difference in performance of HS-DSCH would be smaller when increasing BLER. Hence, HS-DSCH will probably perform even better in practice.

Also from the results of the simulations of the second group, and **several** users, it is clear that HS-DSCH has better channel utilization than a dedicated channel when there are packet losses, especially when BLER is high. Furthermore, we can see that the end users of HS-DSCH are likely to be more satisfaction than those of DCH because HS-DSCH has better average application bit rates. Due to the affect of faster retransmission by HS-DSCH, the conclusion about the improvement due sharing can only partly be drawn. However, some gain due channel sharing could be verified.

Since we only tested some situations, it is too early to make far-reaching conclusions for HS-DSCH in general. But, from our simulations the advantages of HS-DSCH can be seen and these should reduce TCP's performance problem over UMTS.

Scheduling is one of the most important features in HS-DSCH. Simulations from the third group examined how it affects the TCP performance. We found that when packet losses on the fixed network increase, the difference between different schedulers is less and less. In such cases, the TCP window places more limits on the throughput than the scheduling of resources. Since limited simulation samples were made, more simulations in different situations and with different user numbers need to be done to draw more solid conclusions.

10.2. Future work

There are still things to study even though the area has been studied in detail and simulations have been conducted. The limited time for this work as well as the limited form that a thesis represents forced us to leave out some areas that are both relevant and interesting.

Much work in the area of TCP over wireless media remains to be done. Some of the following actions may be worth performing:

- Add soft combining in the model - Just a simple version of HS-DSCH was implemented due to limited time. In future work, additional features can be added to the model, such as soft combing. With a more complete model, a more detailed picture of HSDPA can be seen.

- Using HTTP as an application instead of FTP – HTTP is one of the most popular application on 3G mobile system. The average file size of HTTP is rather small. Using HTTP as an application when performing simulations can give a better analysis about the sharing feature of HSDPA.

- Do more research about scheduling – This is a very interesting topic, especially more evaluations of different algorithms ought to be made. . More research can be done on developing new scheduling algorithms tailored for TCP.

- Some tests in a real network – Although there are no real equipment supplies HSDPA, when equipment with HSDPA becomes available, some tests should be done in a real network to verify our theoretical study.

References

[1] Anders Dahlén and Per Ernström, "TCP over UMTS – An Overview". Technical Report 3/0363-FCPA1090248, Telia Research AB, Sept 2002.

[2] Anders Dahlén, Per Ernström and John Tillberg, "TCP over UMTS". Technical Report 9/0363-FCPA1090185, Telia Research AB, Jan 2002.

[3] Erik Lundsten, "Improving 3G Performance for the Mobile Internet", master thesis, Wireless@KTH, Royal Institute of Technology, Nov. 2002.

[4] S. Mascolo, C. Casetti, M. Gerla, M. Y. Sanadidi, and R. Wang. TCP Westwood: Bandwidth Estimation for Enhanced Transport over Wireless Links. In MOBICOM, pages 287--297, 2001.

[5] Janne Peisa and Michael Meyer, "Analytical Model for TCP file Transfers over UMTS", in proceedings of 3G Wireless 2001, San Francisco, May-June, pp. 42-47, 2001.

[6] Michael Meyer, Joachim Sachs, and Marus Holzke, "Performance Evaluation of a TCP Proxy in WCDMA Networks", IEEE Wireless Communications, Vol 10, No.5, pp. 70- 79 October 2003

[7] Anders Dahlén and Per Ernström, "Simulation of TCP traffic over UMTS", Technical Report in process, Telia Research AB, Jan 2002.

[8] M. Allman, V. Paxson, and W. Stevens, TCP Congestion Control, RFC 2581, April 1999.

[9] Douglas E. Comer, Internetworking with TCP/IP - Volume 1: Principles, Protocols, and Architectures (4^{th} edition), Prentice Hall, 2000.

[10] H. Inamura, et al., "TCP over Second (2.5G) and Third (3G) Generation Wireless Networks", IETF draft, draft-ietf-pilc-2.5g3g-12, March 2002.

[11] CAIDA, "Traffic Workload Overview," http://www.caida.org/outreach/recources/learn/trafficworkload/tcpudp.xml, September 2002.

[12] J. Border, et al., Performance Enhancing Proxies Intended to Mitigate Link Related Degradations, RFC 3135, June 2001

[13] V. Paxson and M. Allman, Computing TCP's Retransmission Timer, RFC 2988, Nov 2000.

[14] V. Jacobson, R. Braden, and D. Borman, TCP Extensions for High Performance, RFC 1323, May 1992.

[15] Pablo José Ameigeiras and Jeroen Wigard, "Impact of TCP Flow Control on the Radio Resource Management of WCDMA Networks", VTC 2002 (Spring), pp. 977 - 981 vol.2, May 2002..

[16] Joachim Sachs and Michael Meyer, "Mobile Internet – Performance Issues Beyond the Radio Interface", 10th Aachen Symposium on Signal Theory, pp. 163-168, Aachen, Germany, September 20th-21st, 2001.

[17] Janne Peisa and Eva Englund, "TCP performance over HS-DSCH", in proceedings of IEEE VTC 2002 (Spring), pp. 987-99, May 2002.

[18] A. Furushär, et al., "Performance of WCDMA High Speed Packet Data", in proceedings of IEEE VTC 2002 (Spring), pp.1116 – 1120 vol.3, May 2002.

[19] S.Parkvall, et al., "WCDMA evolved – High-speed packet-data services", Ericsson Review No.2, 2003, pp. 56-65.

[20] Troels Kolding, Frank Frederiksen, and Preben Mogensen, "Performance Aspects of WCDMA Systems with High Speed Downlink Packet Access (HSDPA)", Proceedings, VTC, September, 2002, Vol. 1, pp. 477-481

[21] B. Gull, et al., "Providing High Bit Rates In UMTS –An overview of High Speed Downlink Packet Access", Technical Report 11/0363-FCPA 1090171, Telia Research AB, July 2001.

[22] Johan Gustavsson and Johan Lewin, "Impact of MAC and RLC Protocols on UMTS release 5", master thesis, Department of Information Technology, Lund Institute of Technology, Lund University, Feb. 2004.

[23] The Network Simulation (version 2): http://www.isi.edu/nsnam/ns/doc/index.html, December 2003

Appendix A – Simulation Results for scheduling

Table A.1 Application Bit Rate of using different scheduling with 3 users (kbps)

Scheduling		**Packet Loss on Internet**		
		0%	**1%**	**3%**
RR	**User0**	313.71	257.06	155.10
	User1	348.88	302.67	151.59
	User2	433.21	288.99	159.98
	Average	**365.27**	**282.91**	**155.56**
	STD	**61.41**	**23.41**	**4.21**
MAX CQI	**User0**	366.81	254.06	149.11
	User1	514.36	305.66	175.41
	User2	932.20	369.12	169.50
	Average	**604.46**	**309.61**	**164.67**
	STD	**293.26**	**57.63**	**13.80**
PF	**User0**	395.89	274.20	156.48
	User1	461.86	288.62	163.82
	User2	601.22	303.62	173.34
	Average	**486.32**	**288.81**	**164.55**
	STD	**104.83**	**14.71**	**8.45**
(STD) Standard Deviation				

Table A.2 Application Bit Rate of using different scheduling with 6 users (kbps)

Scheduling		**Packet Loss on Internet**	
		1%	**3%**
RR	**User0**	252.63	166.63
	User1	261.79	182.43
	User2	296.87	174.17
	User3	263.46	170.48
	User4	298.21	160.32
	User5	303.49	167.64
	Average	279.41	170.28
MAX CQI	**User0**	283.62	169.71
	User1	335.09	169.90
	User2	360.65	183.83
	User3	281.75	160.19
	User4	301.93	163.55
	User5	331.13	170.58
	Average	315.70	169.63
PF	**User0**	300.37	174.22
	User1	268.08	169.26
	User2	339.74	165.04
	User3	289.08	169.76
	User4	309.21	180.25
	User5	303.45	163.69
	Average	301.65	170.37

Appendix B – Abbreviations

ACK	Acknowledgment
AMC	Adaptive Modulation and Coding
AP	Access Point
ARQ	Automatic Request for Retransmission
BDP	Bandwidth Delay Product
BLER	Block Error Rate
CN	Core Network
CS	Channel switched
C/I	Carrier-to-interface ratio
DCH	Dedicated Channel
DUPACK	Duplicate Acknowledgment
FTP	File Transfer Protocol
GGSN	Gateway GPRS Support Node
GMSC	Gateway Mobile Services Switching Center
HARQ	Hybrid ARQ
HSDPA	High Speed Data Packet Access
HS-DPCCH	High Speed Dedicated Physical Control Channel
HS-DSCH	High Speed Downlink Shared Channel
HTTP	Hyper Text Transfer Protocol
IETF	Internet Engineering Task Force
IP	Internet Protocol
I-TCP	Indirect TCP
MTU	Maximum Transfer Unit
PDU	Protocol Data Units
PF	Proportional Fair
PS	Packet switched
RAN	Radio Access Network
RBS	Radio Base Station
RLC	Radio Link Control
RNC	Radio Network Controller
RNS	Radio Network Sub-system
RR	Round Robin
RRC	Radio Resource Control
RTT	Round Trip Time
SDU	Service Data Unit

SGSN	Serving GPRS Support Node
SMSS	Sender Maximum Segment Size
TCP	Transmission Control Protocol
TTI	Transmission Time Interval
UE	User equipment
UMTS	Universal Mobile Telecommunication System
UTRAN	UMTS Terrestrial Radio Access Network
WAP	Wireless Application Protocol Forum

IMIT/LCN 2004-02

www.kth.se