Providing Strict Quality of Service in HSDPA for real time services

Joseph S. Gomes[1], Mira Yun[1], Hyeong-Ah Choi[1], Jae-Hoon Kim[2], JungKyo Sohn[3], Hyeong-In Choi[3]
[1]*Department of Computer Science, George Washington University, Washington, DC*
[2]*Access Network & Mobile Terminal R&D Center, SK Telecom, Seoul, Korea*
[3]*Department of Mathematics, Seoul National University, Seoul, Korea*
E-mail:{joegomes,mirayun,hchoi}@gwu.edu, jayhoon.kim@gmail.com, {jgsohn,hichoi}@snu.ac.kr

Abstract

Increasing demand for high data-rate real time services has led to the use of shared channels in the forward link for real-time services in 3G wireless networks, such as HSDPA. Strict and differentiated QoS requirements for real-time services, along with time varying channel capacity impose a great deal of challenge on the MAC-hs scheduler. In this paper, we present our policy-based scheduling algorithms to support differentiated QoS requirements along with the HSDPA network simulator in OPNET that we developed with newly introduced features, such as link adaptation, HARQ, fast scheduling at NodeB and quality based admission control.

1. Introduction

High Speed Downlink Packet Access (HSDPA) is introduced to increase WCDMA [1] downlink packet data throughput in the Release 5 of the 3GPP UTRAN specifications. HSDPA offers theoretical peak data rates of above 10 Mbps, which is achieved by implementing a fast and complex channel-control mechanism based upon short physical layer frames (2 ms), Adaptive Modulation and Coding (AMC), fast Hybrid-ARQ and fast scheduling [1]. However, generally such high data rates and correspondingly high throughput cannot be realized due to shortsighted decisions made by the scheduler.

Fairness issues in scheduling have been studied in depth in [2], [3]. The proportional fair (PF) algorithm, most popular packet scheduling algorithm considering fairness, was thoroughly investigated in [4], [5], and [6]. Three packet-scheduling algorithms namely Max C/I, Proportional Fairness (PF), and Round Robin (RR) are compared in [7] focusing on the achievable throughput of each user in HSDPA. Reference [8] proposes QoS based scheduling algorithms for real-time data users over shared wireless link. Reference [9] proposes an algorithm that provides QoS guarantee using barrier functions and [10] considers multi-user transmission in one slot.

A quality based admission control algorithm is derived in [11]. The interaction between admission control and fast scheduling have been unexplored under mixed service requirements in HSDPA networks, although some work has been done in this area for UMTS networks [12], [13] that use dedicated channels unlike HSDPA. None of the existing works provides policy driven QoS guarantee either. For example, a service operator may have a certain policy on how to prioritize QoS classes during times of overload while another policy may determine how to distribute surplus capacity when the total guaranteed bit rates (GBR) do not exceed the capacity. Another service provider may have a different allocation policy during times of overload and may choose to improve channel quality by increasing effective code rate when there is surplus capacity. In this paper, we consider such policy based QoS support [16] [17], present two different algorithms and show that our algorithms very effectively satisfy the users by guaranteeing such QoS requirements.

2. HSDPA System Overview

Figure 1 sketches the hierarchical structure of an HSDPA network. In a typical HSDPA network, data enters the HSDPA/UMTS core network directly from the Internet or from the PSTN through a Packet Data Switching Node (PDSN). In the core-network, data is routed through the GGSN and SGSN, and into the RNC of the radio access network (RAN). It is then forwarded to the Node-B (base station) and finally to the mobile station over the radio link.

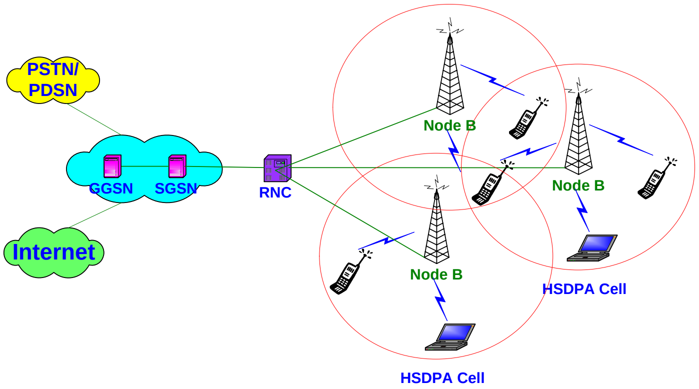


Figure 1: HSDPA Network Architecture

2.1 Newly added channels

Instead of dedicated channels as in release 99, HSDPA uses shared wireless downlink channels for high-speed data. A new downlink transport channel HS-DSCH (High Speed Downlink Shared Channel) was introduced, that carries data to the selected UE (or UEs) during each transmission time interval (TTI) of two ms. The transported bits from HS-DSCHs are mapped onto up to 15 physical downlink shared channels (HS-PDSCH) each using a separate orthogonal cdma code with a spreading factor of 16. The associated High Speed Shared Control Channel (HS-SCCH) is used to communicate control information between the UE and Node-B.

2.2. Adaptive Modulation and Coding (AMC)

In the uplink direction, HS-SCCH channels are used by the UEs to notify the Node-B of the *Channel Quality Indicator* (CQI) and a positive or negative acknowledgement pertaining to the received frame. CQI indicates the instantaneous channel quality experienced by the user, so that the Node-B can adjust its transmission parameters (modulation type, coding rate, number of codes) to cope with variations in channel conditions. The CQI reported by the UE corresponds to transmission parameters (modulation type, coding rate, number of codes) that would result in the maximum data rate possible while providing an acceptable *block error rate* (BLER) for the current link conditions. The reported CQI value is then used by Node-B to determine the appropriate parameters or *modulation and coding set* (MCS) for the next packet transmission to the UE. The actual parameters used are notified to the UE using the downlink HS-SCCH channel. Several AMC schemes are proposed including QPSK and 16QAM with coding rates of 1/3, 1/2 and 3/4. In order to reduce the latency involved in link adaptation the scheduling functionality has been moved from RNC to Node-B.

2.3. UE Categories

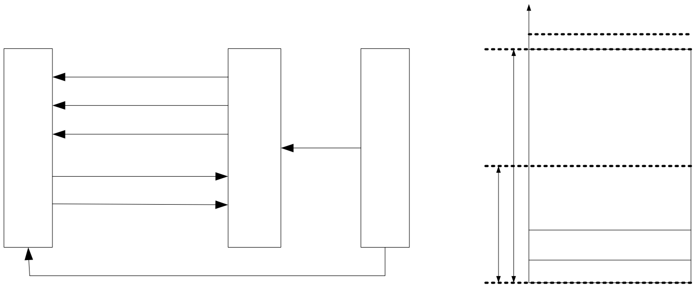
The UEs are divided into 12 different categories based on their capabilities. For example, UE category k for k = 1 or 2 can only support data rates up to 1.2 Mbps using 5 simultaneous physical channels (codes) and has minimum inter TTI interval *min_TTI(k)* of 3. If user u_i from category k is scheduled to transmit in TTI *t*, the earliest TTI in which u_i can be scheduled next is *t+min_TTI(k)*. Theoretically, category 7 can support up to 7.21 Mbps using 10 codes (under perfect link conditions) and has minimum inter TTI interval of 1.

2.4. Hybrid Automatic Repeat Request

Hybrid automatic repeat request (HARQ) is a technique where UE stores previous transmissions that are in error in soft memory to be combined with future re-transmissions for decoding. For each packet, the UE sends HARQ feedback (ACK or NACK) to inform the Node-B whether a retransmission is required or not. HARQ uses one of two different schemes, with identical retransmissions, often referred in the literature as Chase combining, or with non-identical retransmissions, otherwise called incremental redundancy.

2.5. Admission Control and Scheduling

Admission control is done at the RNC as in release ’99, while scheduling of admitted users is done by MAC-hs scheduler at the Node-B. Resource allocation, admission control, and scheduling are collaborative efforts between the RNC and the Node-B as shown in Figure 2. The Node-B receives guaranteed bit rate (GBR) and scheduling priority indicator (SPI) values from the RNC, which it can use to make scheduling decisions. On the other hand, RNC allocates channelization codes and power for HSDSCH and HS-SCCH transmissions based on load and performance measurements provided by Node-B. These measurements include but are not limited to the total carrier power, non-hsdpa power and the HS-DSCH required power. HS-DSCH required power is reported per SPI and is an estimate of the total power needed to serve all admitted user with that SPI at their GBR. The RNC also receives an E_c/N_0 measurement of the CPICH channel from the new user seeking admission. Based on all these reports, the RNC can estimate whether the user can be granted access without deteriorating the services to the existing users.



(a) Exchanging parameters and Measurement reports

(b) Power budget

Figure 2: Admission Control Parameters 1

3. OPNET HSDPA Simulator

We developed an HSDPA model by extending the UMTS model in OPNET. Our system extensively simulates all the key entities of the network such as UEs, Node-B, RNC, SGSN etc. with most of the functionalities at all the protocol layers. In the following section, we briefly describe the relevant portions of our simulator implementation. A quality based access control algorithm is implemented at the RNC, which will be described in section 3. Several packet-scheduling algorithms are implemented in NodeB including our LPS and SPS algorithms, which will discussed in Section 4.

3.1. AMC and HARQ

An UE measures the SINR for each received packet and reports the corresponding CQI back to the NodeB. For this, actual value interface (AVI) [14] is used to map a given *modulation and coding set* (MCS) to a received per-packet average SINR for obtaining a specific BLER (we use 0.1). For each UE category, we use a separate AVI table that maps an MCS to a corresponding threshold SINR. In Table 1 we show sample MCS to SINR mappings for ue categories 7 and 8. For example, a CQI of 16 for a UE of category 7 means that the NodeB can transmit 3565 bits to this UE in the next TTI using 16-QAM modulation and an effective code rate of 0.37. A UE reports the maximum possible CQI for its current channel condition once every 10 TTI. Error decision for a received packet is also made based on the threshold SINR for the MCS associated to the transmitted packet. If the received SINR is less than the threshold SINR the packet is in error, otherwise a uniformly distributed random number $y \in [0; 1]$ is generated. If $y \geq 0.1$ the packet is successfully received. We also model a delay of 12 ms for the Node-B to get the associated ACK/NACK response back, from the time of transmission.

We assume Chase combining for the H-ARQ process and use the following model [15]:

$$\left(\frac{E_s}{N_0}\right)_{C,n} = \varepsilon^{n-1}\cdot\sum_{k=1}^{n}\left(\frac{E_s}{N_0}\right)_k$$

, where $(E_s/N_0)_{C,n}$ represents the combined E_s/N_0 after n transmissions and $(E_s/N_0)_k$ corresponds to the k-th transmission. ε is the Chase combining efficiency, which is set to 0.93. In Chase Combining, an identical version of the original frame has to be retransmitted. Therefore, when multi-user physical layer frames (code multiplexing) are used, even the users that received their packet correctly during the original transmission will receive the retransmission.

CQI value	Transport Block Size	Number of HS-PDSCH codes	Modulation	Data rate (Mbps)	ECR (per code)	SINR (dB)
0	Out of Range					
1	137	1	QPSK	0.07	0.15	-6.5
2	173	1	QPSK	0.09	0.19	-5.5
3	233	1	QPSK	0.12	0.25	-4
…	…	…	…	…	…	…
16	3565	5	16-QAM	1.78	0.37	14.68
17	4189	5	16-QAM	2.09	0.44	16.59
…	…	…	…	…	…	…

Table 1: MCS to SINR mapping table for ue categories 7, 8

3.1. Priority Based Admission Control at RNC

In HSDPA, the scheduling has been moved from RNC to the NodeB. However, the task of admission control remains in the RNC. The RNC process model in Opnet is depicted in Figure 2. When an HSDPA packet destined for a particular UE arrives at RNC in the FROM_CN state, it is forwarded to the serving NodeB. At the beginning of each session, if the necessary resources are not already allocated to the UE, the RNC performs admission control algorithms in the ADM_CNTL state for both uplink and downlink to verify if enough resources are available to serve the UE. The uplink admission control algorithm remains the same as in UMTS due to the use of UMTS (DCH channels) in uplink direction. For downlink, we use a priority based admission control algorithm that was proposed in [11] with some modifications. In this approach, a new user is admitted only if it can be served with its target bit rate without degrading the throughput of all the users with the same or higher SPI (priority) from their target bit rates. We added attributes such as spi, downlink gbr, ue category etc. in the mobile station node model to individually set the scheduling priority, GBR and category for each UE as shown in Figure 3.

In the priority-based algorithm, first the average HS-DSCH power allocated to user k is estimated as,

$$P_k = \frac{1}{N}\sum_{m=1}^{M_k} P_k^{'}(m) \qquad (1)$$

, where $P^{'}_k(m)$ is the allocated transmission power in the *m-th* TTI where the user was scheduled, and M_k is the number of TTIs user k was scheduled out of the N TTIs. When code multiplexing is used, we proportionally adjust $P^{'}_k(m)$ according to the multicodes assigned to user k during that TTI. The average bit rate provided to user k over N TTIs is,

$$R_k = \frac{1}{N \cdot T_{tti}}\sum_{m=1}^{M_k} B_k(m) A_k(m) \qquad (2)$$

Here $B_k(m)$ is the transport block size (TBS) of the m-th transmission to the user. Notice if the user does not have enough data to send then much of the transport block could be empty and the user will be experiencing a much lower avg. bit rate than the provided bit rate. $A_k(m) = 1$, if an ack was received from the user and 0 otherwise. Given equations 1 and 2 the required power to provide all the users, that have SPI value x, with their GBR can be approximated as,

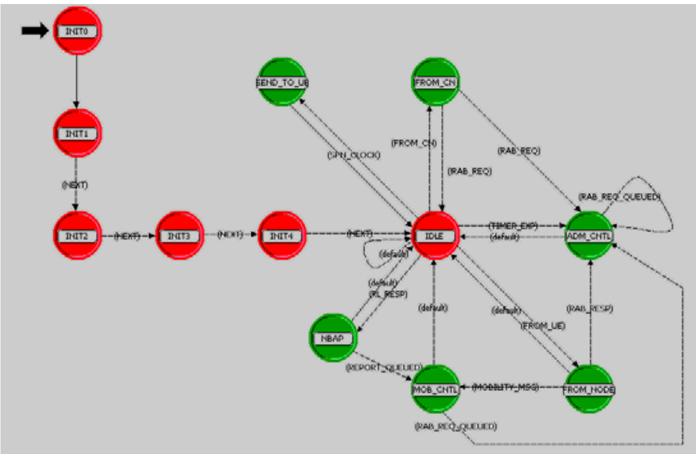
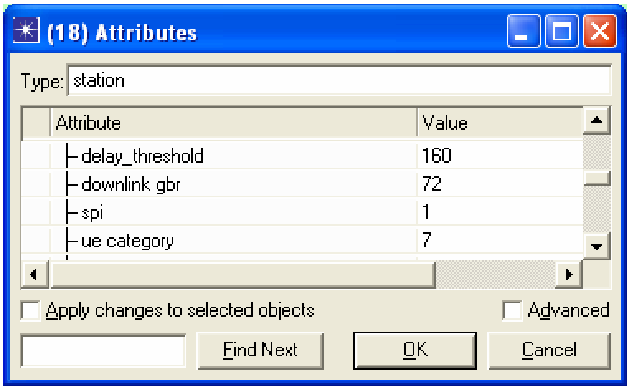


Figure 2: HSDPA RNC process model



Type: station

Attribute	Value
delay_threshold	160
downlink gbr	72
spi	1
ue category	7

Figure 3: Setting Mobile Station Attributes

$$P_{SPI}(x) = \sum_{k \in SPI(x)} P_k \frac{GBR_k}{R_k} \qquad (3)$$

, where GBR_k is the GBR for user k, and $SPI(x)$ is the set of users with SPI value x.

Now let the target bit rate and priority of the new user be denoted by GBR_{new} and SPI_{new} respectively. Let P_{HSDPA}, P_{new}, and P_{SCCH} be the allocated HSDPA transmission power, the estimated power required to serve the new user and the estimated power required for transmitting HS-SCCH respectively. Then the new HSDPA user is admitted only if

$$P_{HSDPA} \ge + \sum_{x \ge SPI_{new}} P_{SPI}(x) + P_{SCCH} + P_0$$

Here P_o is a configuration parameter, which represents a safety power offset to compensate for potential estimation errors. In

[11], the required power to serve the new user at the target bit rate GBR_{new} is estimated as

$$P_{new} = P_{Tx,DSCH} \frac{GBR_{new}}{f(\rho)}, \qquad (5)$$

where ρ is the average experienced HS-DSCH SINR at the new user as estimated at the RNC from the pilot channel measurement report, and $f(p)$ is the average throughput that a user can be served with when it is scheduled in every TTI with all the available transmission power. The function $f(\cdot)$ is assumed to be stored in a table in the RNC. However since we consider code multiplexing we use a slightly different equation to compute P_{new}. Suppose the total number of channelization codes allocated by the RNC is given by ψ. Let CQI_p be the CQI value that the UE would report when experiencing an SINR of p. Let φ_p be the number of multicodes corresponding to the transmission configuration (MCS) used by the Node-B when sending to the UE using CQI_p. Then the total transmission power used to send to the UE can be estimated as $P_{Tx,DSCH} \cdot (\varphi_p/\psi)$, and thus we have

$$P_{new} = P_{Tx,DSCH} \left(\frac{GBR_{new}}{f(\rho)}\right)\left(\frac{\phi_\rho}{\psi}\right) \qquad (5)$$

5. Packet Scheduling at NodeB

Figure 4 shows the process model for an HSDPA NodeB. When a packet destined for a UE arrives from RNC, it is added to the transmission queue of that user in state HS_ADD_TO_BUFFER. During each TTI, in state HS_SCHEDULE, the scheduling candidate set (SCS) is created comprised of all the users that have non-empty transmission queues and are elligible to receive. A user is elligible, if the number of elapsed TTIs since its last reception exceeds its min_TTI. Users are chosen from SCS to be scheduled to receive packets based on the scheduling algorithm being used.

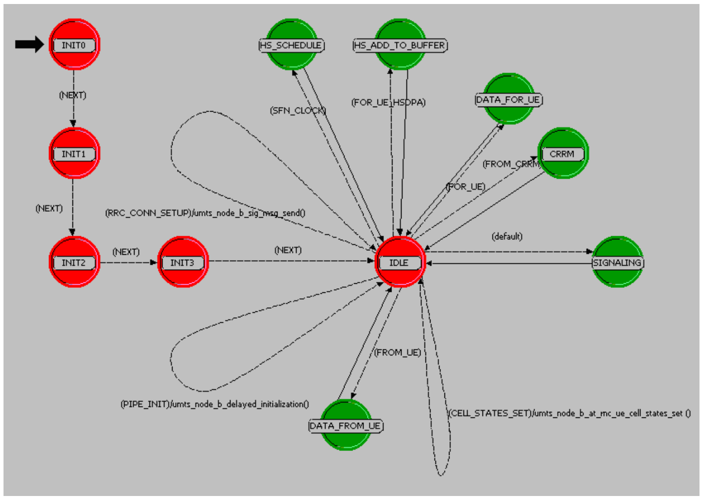


Figure 4: HSDPA NodeB process model

5.1. Notations and Assumptions

We make the following notations and assumptions to formulate the scheduling problem: a Node-B has n users, $u_1 \ldots u_n$, connected to it, where each user u_i is associated with a UE category $k_i \in \{1, \ldots, 12\}$ and each UE category k_i is associated with $min\_TTI\ (k_i) \in \{1, 2, 3\}$. GBR_i denotes the guaranteed bit rate for user u_i. Node-B is allocated M HS-PDSCH codes by the RNC for downlink transmission. $CQI_i\,(t) \in CQI = \{0, 1, \ldots, 30\}$ denotes the CQI value reported by u_i that can be used for scheduling in $TTI\ \ t$. Given $x \in CQI$, we use $blk(x)$, $\#CH\ (x)$, and $DR\ (x)$ to represent the block size (i.e., the transport channel bits), the number of required parallel HSPDSCH codes, and the instantaneous data rate associated to x. $\lambda_i\ (t)$ is the assigned CQI for u_i at TTI t. If u_i is not scheduled to receive any data at t then $\lambda_i\ (t) = 0$. Suppose user u_i entered the system at TTI s_i. Then, $\delta = t - s_i + 1$ is the number of TTIs the user has been in the system. We denote $th_i\ (t)$ to be the moving average of the throughput of user at TTI t over the last w_0 TTIs defined as follows:

$$th_i(t) = \begin{cases} th_i(t-1)(1-\dfrac{1}{w_0}) + \dfrac{DR(\gamma_i(t))}{w_0}, & \delta \geq w_0 \\ \dfrac{\sum_{j=s_i}^{t} DR(\gamma_i(t))}{t - s_i + 1}, & otherwise \end{cases}$$

5.2. Scheduling Problem Statement

For every TTI t find $X\,(t) = (\gamma_1(t), \ldots, \gamma_n(t))$ such that,

(a) $\gamma_i(t) \leq CQI_i(t)$ for all i,

(b) $\sum_{i=1}^{n} \#CH(\gamma_i(t)) \leq M$,

(c) $\sum_{k=t-\min\_TTI(C_i)+1}^{t-1} \gamma_i(t) = 0$

In other words, the goal of the scheduling problem is to select a set of receiving users and to select the instantaneious transmission data rates (format) for them such that each user's assigned CQI is less than their requested CQI, the total number of codes used is less than M, and each selected user is elligible.

5.3. Loose Policy Scheduling

In addition to the above constraints, a service provider may have other policies that impose supplementary requirements on the scheduler. These policies could be loose or strict in nature. Here, we define a loose policy scheduler to be a scheduler that tries to find a balance between the two goals of maximizing throughput (using opportunistic scheduling) and satisfying QoS constraints (GBR). Notice that this is different from the goal of PF algorithm, which disregards QoS constraint. The PF (proportional fair) scheduler orders the receivers using the metric $r_i(t)/th_i(t)$, where $r_i(t)$ is the instantaneous data and $th_i(t)$ is the current throughput. It considers all users equally important and adds a fairness property with respect to user throughput. We try to incorporate both QoS constraints and throughput maximization into our algorithm by considering the metric $\mu = \dfrac{DR(\gamma_i(t))}{\#CH(\gamma_i(t))} \cdot \dfrac{GBR_i}{th_i(t)}$. Notice that the first ratio contributes to the users with better instantaneous data rate per code and the second ratio emphasizes users who are furthest away from meeting their guaranteed bit rate. Next, we give the algorithmic steps in further detail.

1. Let $U_{active}(t)$ denote the set of users who may be scheduled for transmission in *TTI t* (i.e., constraint (c) is satisfied.) For each user $u_i \in U_{active}(t)$, define $\Delta_i(t) = w_0(GBR_i - Th_i(t-1)) + Th_i(t-1)$.
2. Given $CQI_i(t)$, choose smallest $\gamma_i(t)$ such that (i) $\gamma_i(t) \le CQI_i(t)$ and (ii) $DR(\gamma_i(t)) \ge \Delta_i(t)$.
3. Sort the users in $U_{active}(t)$ in descending order of μ.
4. Let z denote the number of unused HSPDSCHs. We are done when either z = 0 or U_{active} is empty. Otherwise, choose the user from the sorted list with the highest μ and assign $\#CH(\gamma_i(t))$ codes to the user if $\#CH(\gamma_i(t)) < z$. Otherwise, assign z codes and the highest possible CQI using z codes to the user. Remove the user from U_{active}.
5. Repeat Step 4.

5.3. Strict Policy Scheduling

Now we will consider a stricter set of policy requirements. Suppose a service provider differentiates its users using Scheduling Priority Indicator (*SPI*) values, where $SPI \in C = \{1, 2, \ldots, 15\}$[1]. In general, we assume higher classes (with higher SPI) have higher *GB*R values. Although the problem can be generalized for any number of *SPI* values, for simplicity let us consider only *SPI* 1 and 2 with corresponding *GB*R values GBR_s and GBR_g. We will call these two classes *Gold* and *Silver* classes respectively. Let $d_i(t)$ denote the maximum achievable instantaneous data rate indicated by the CQI for user u_i at TTI t, where $d_i(t) \in \{r_1, r_2, \ldots, r_k\}$ such that $r_1 = r_{min} < r_2 < \ldots < r_k = r_{max}$. We assume that a user belonging to a particular class is on average under good enough channel condition to be able to receive its *GBR*. Under this assumption, the goal of our policy is to define fair rules for governing resource allocation under all circumstances, i.e. to guarantee each user its *GBR* and fairly distribute the surplus capacity when there are enough resources, and to satisfy users from higher classes before the lower classes when there are not enough resources. The policies are as follows:

P1. A silver user can be scheduled only if all gold users have been satisfied.

P2. If there are multiple gold users with throughput less than GBR_g, the gold user with the highest value of $d_i(t)$ is scheduled.

P3. A silver user s with $th_s(t)$ less than GBR_s has higher priority than any gold or silver user that meets its *GBR*.

P4. If there are multiple silver users with throughput less than *GBR*s while all gold users have been satisfied, a silver user with the highest value of $d_i(t)$ is scheduled.

P5. When all users have met their *GBR*, surplus capacity must be proportionally distributed among the gold and silver users according to their *GBRs*.

5.3.1. Marginal Utility Function

Let N be the set of users in the system who can receive at *TTI t*. We will define marginal utility functions $M_c(th_i(t))$ for each class $c \in \{Gold, Silver\}$, whose purpose is to assign a utility value, following the policy rules described earlier, to each user u_i at *TTI* t according to its class c, current data rate $d_i(t)$ and current throughput $th_i(t)$. Eventually the utility values will be used to determine which user(s) will be scheduled at t. Let $M_c(th_i(t)) = P_c(th_i(t))\, d_i(t)$ where P_c denotes the preliminary utility function which we will define shortly. Then for any gold user, g, and silver user, s, P_{Gold} and P_{Silver} has to follow the following conditions:

C1. $P_{Gold}(th_g(t))\, r_{min} > P_{Silver}(th_s(t))\, r_{max}$ for $0 \le th_g(t) < GBR_g$ and $0 \le th_s(t) \le r_{max}$

C2. $P_{Silver}(th_s(t))\, r_{min} > P_{Gold}(th_g(t))\, rmax$ for $0 \le th_s(t) < GBR_s$ and $GBR_g \le th_g(t) \le r_{max}$

C3. $P_{Silver}(th_s(t)) = P_{Gold}\left(\frac{GBR_i}{th_i(t)} \cdot th_s(t)\right)$, for $GBR_s \le th_s(t) \le r_{max}$ and $GBR_g \le th_g(t) \le r_{max}$.

C4. For any two silver users, s and s', $P_{Silver}(th_{s'}(t))\, r_{min} > P_{Silver}(th_s(t))\, r_{max}$ for $0 \le th_{s'}(t) < GBR$s and $GBR_s \le ths(t)$ r_{max}

C5. For any two gold users g and g', $P_{Gold}(th_{g'}(t))\, r_{min} > P_{Gold}(th_g(t))\, r_{max}$ for $0 \le th_{g'}(t) < GBR_g$ and $GBR_g \le thg(t) \le r_{max}$.

Note that Condition C1 and C5 correspond to policy P1, C2 and C4 correspond to P3, and Condition C3 corresponds to P5. We define functions P_{Gold} and P_{Silver} satisfying the conditions above as follows. Let $\beta = r_{max} - GBR_g$, $\alpha = r_{max}/r_{min}$ and x be the throughput of the user.

$$P_{Silver}(x) = \begin{cases} \alpha\beta + 1, & 0 \le x < GBR_s \\ r_{\max} - \dfrac{GBR_g}{GBR_s} x, & GBR_s \le x \le r_{\max} \end{cases}$$

$$P_{Gold}(x) = \begin{cases} \alpha^2\beta + \alpha + 1, & 0 \le x < GBR_g \\ r_{\max} - x, & GBR_g \le x \le r_{\max} \end{cases}$$

Functions P_{Gold} and P_{Silver} are shown in Fig. 1. Notice that maximum P_{Silver} value for a silver user with throughput $th_s(t) \ge GBR_s$ and maximum P_{Gold} value a gold user g with throughput $th_g(t) \ge GBR_g$ is β. So the maximum marginal utility (M_c) for these users is βr_{max}. The minimum marginal utility for a silver user s' with $th_{s'}(t) < GBR_s$ is $(\alpha\beta + 1)\, rmin > \beta r_{max}$. This satisfies conditions C2 and C4. Similarly if $th_s(t) < GBR_s$, $M_{Silver}(th_s(t)) \le (\alpha\beta + 1)\, r_{max}$. For a gold user with $th_g(t) < GBRg$, $M_{Gold}(th_g(t)) \ge (\alpha^2\beta + \alpha + 1)\, r_{min}\, (\alpha\beta + 1)\, r_{max}$. This satisfies conditions C1 and C5. It can also be noticed from the figure that when all users meet their GBR, condition C3 is also satisfied. Also since two gold users with throughtput less than GBR_g has the same P value, the one with the higher data rate will have a higher marginal utility, which conforms to P2. For similar reasons P4 is also satisfied.

Suppose there are more than two classes, i.e. $|C| > 2$. Let GBR_{max} denote the *GB*R associated to the highest class or SPI. Then P_c can be generalized for any *SPI* value $c \in C$ as

$$P_c(x) = \begin{cases} \alpha^c \beta + \alpha^{c-1} + \ldots + \alpha^0, & 0 \le x < GBR_c \\ r_{\max} - \dfrac{GBR_{\max}}{GBR_c} x, & GBR_c \le x \le r_{\max} \end{cases}$$

5.3.1. Strict Policy Scheduling Algorithm

Suppose *U* is the list of users that are elligible to receive in TTI *t*. A user is elligible if there are packets to be sent to that user and minimum inter TTI interval for that user has elapsed since its last reception. Let *M* be the number of codes allocated for the HS-DSCH channels and *codes left* be the number of codes left to be assigned. During each TTI *t*, the following steps are used to produce the list of users that are scheduled to receive data.

1. Let *i* be the user such that $i = argmax_j \{M_{c_j} (th_j (t))\}$ where user j belongs to class c_j. Let cqi_i be its requested *CQI*, *codes*(cqi_i) be the number of codes required.
    a. If (*codes* (cqi_i) < *codes_left*)
        Send to user u_i using data rate *DR* (cqi_i)
       Else
        Send to user u_i at the maximum rate possible
        Using *codes_left* number of codes.
    b. Update *codes left* and remove from *U*.
2. If (*codes_left* > 0 and $U \ne \varnothing$) goto Step 1.

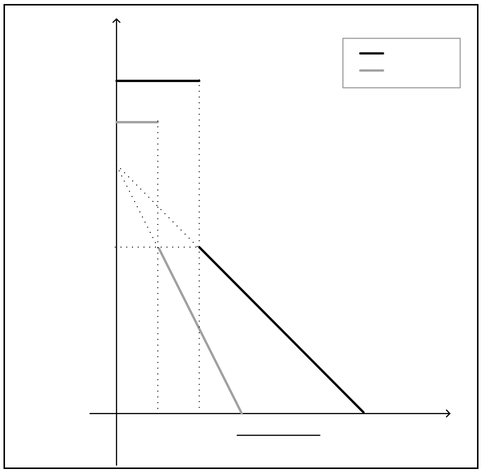


Figure 5: P_{Gold} **and** P_{Silver}

6. Performance Evaluation

In order to evaluate our proposed schedulers we simulated several scenarios. The GBR for each class, the traffic load, and capacity of the network (HS-PDSCH codes) are varied based on the scenarios. The main parameter settings of our simulations are shown in Table I.

Parameter	Setting
Total HS-PDSCH power	8 W
Number of HS-PDSCH codes	5,10
H-ARQ	Chase Combining
Number of H-ARQ channels	6
Max number L1 transmissions	4
CQI reporting interval	On every TTI
HSDPA terminal category	7
Path loss model	Vehicular Outdoor
Shadow fading std.	10 DB
Site-to-site distance	2km
User speed	0-120 km/h
Throughtput window	1000 TTI

P

$\alpha^2\beta+\alpha+1$

$\alpha\beta+1$

$r_{\max}$

Table 2: Summary of Main Simulation Parameters

6.1. Evaluating Loose Policy Scheduling Algorithm

We compared our LPS algorithm with Max C/I and PF, using the following scenario. There are ten gold and ten silver users all with UE category 7. On average, the gold users are at a worse (by around 5 db) radio condition than the silver users. Ten codes are allocated at NodeB for downlink transmission. Each gold user has a guaranteed bit rate of 384 kbps, whereas for the silver users it is 128 kbps. Packets are being generated destined to each gold and silver user at the same rate as their GBR.

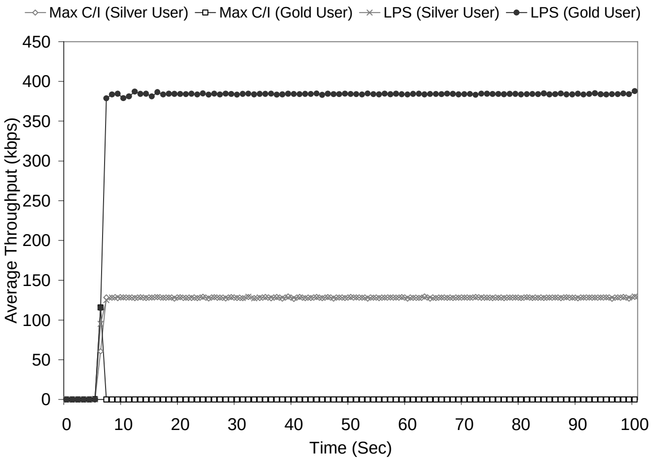


Figure 6: Comparing Loose Policy Scheduler and Proportional Fairness

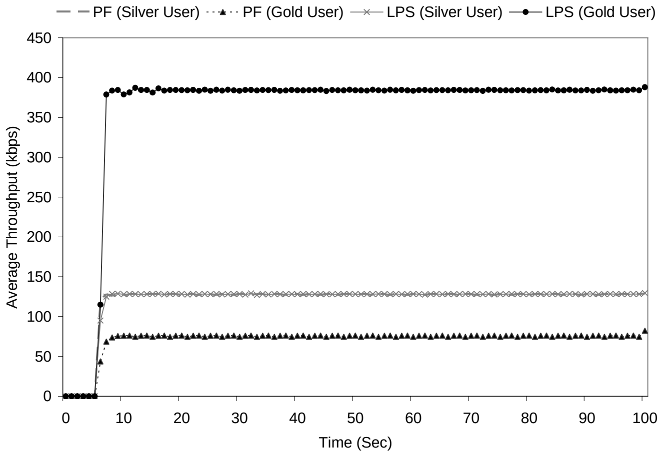


Figure 7: Comparing Loose Policy Scheduler and Max C/I

Figures 2 and 3 show the average throughputs of a typical gold and silver user. Notice in Figure 2 the PF scheduler provides the silver users with a higher throughput (around 128 kbps) than the gold users (75 kbps), since they have better channel condition. At every TTI, Max C/I (Fig. 3) always choose the UE with best

P_{gold}

P_{silver}

channel condition. Consequently, the silver users receive at 128 kbps, whereas the gold users are much starved. On the other hand, LPS meets both gold and silver users' throughput requirements. Moreover, since PF and Max C/I choose one user for each TTI, the capacity cannot be fully utilized if the scheduled user's queue does not have enough data to match the chosen CQI. Since the silver users, that have lesser data destined for them, are being chosen most of the time (always by Max C/I), the cell capacity is being under utilized. On the other hand, as LPS chooses the minimum CQI that supports the GBR, it can send to multiple users at the same time, when M is large enough. This, together with the high GBR of the gold users, increases capacity utilization. Notice in Fig. 4 that LPS has an average downlink throughput of 5 Mbps compared to Max C/I's 1.28 Mbps and PF scheduler's 2 Mbps.

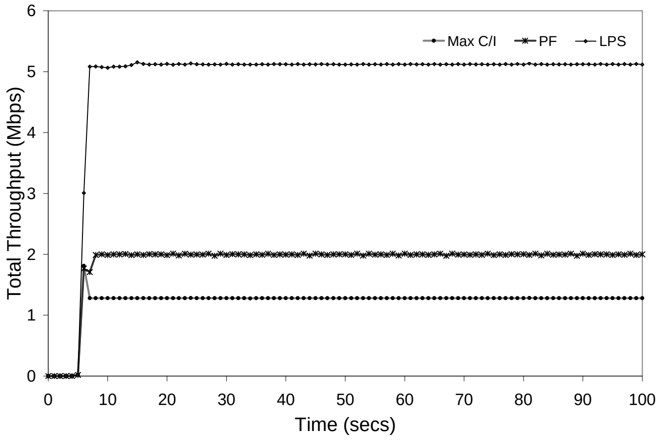


Figure 8: Comparing total received throughput at the mobile stations

In a similar scenario, we moved one of the gold users further away from the NodeB than the rest of the users to worsen its radio conditions (20 db). Consequently, LCS reduced its average throughput to around 200 kbps, while still maintaining the GBR for all the other users. This indicates that LCS can relax the QoS constraints in order to increase throughput.

6.1. Evaluating Strict Policy Scheduling Algorithm

1) Scenario 1: In this scenario, GBR_g and GBR_s are set to 250 kbps and 100 kbps respectively. The traffic load destined to the Gold and Silver users are 300 and 150 kbps. 5 codes were allocated for the HS-PDSCH channels. A new gold user is added to the network after 50 seconds. For UE category 1 the downlink capacity using 5 codes is around 3.6 Mbps under good channel conditions.

From Fig. 5 we can see that before adding the new user, all users are receiving at their GBR and the total downlink throughtput in the cell reaches cell capacity. After the new gold user came in its throughput quickly went upto 250 kbps; however each silver user's throughput went down by about 25 kbps to make room for the new gold user. However, Fig. 6 illustrates the disability of PF Scheduler in maintaining any QoS requirement. In this case, both gold and silver users receive similar throughput (150 kbps) as they are experiencing similar radio conditions (SINR).

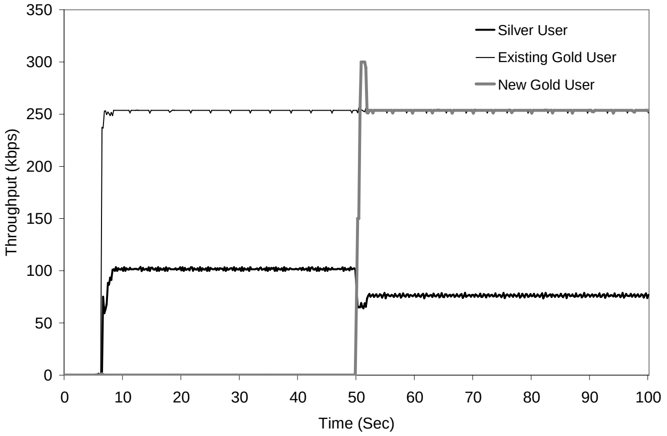


Figure 9: Change in throughput when extra user added

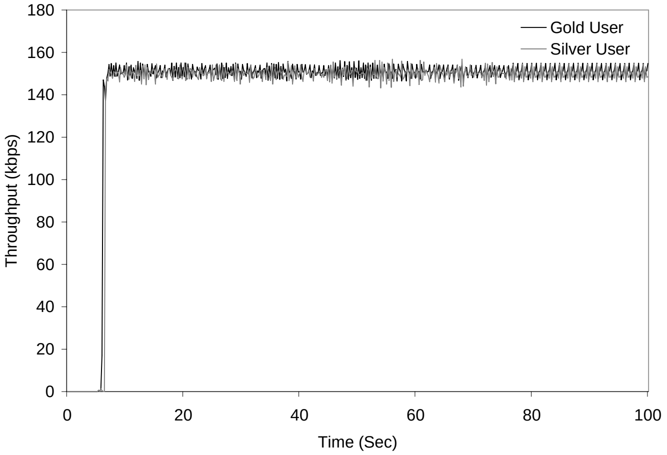


Figure 10: Gold and Silver throughput with PF

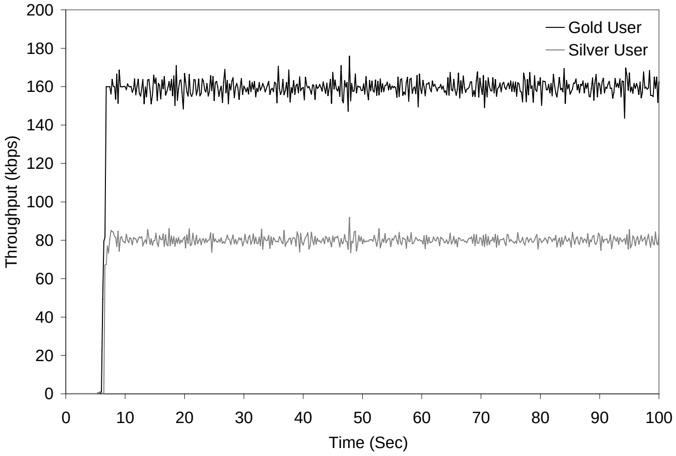


Figure 11: SPS throughput when traffic load exceeds GBR

2) Scenario 2: For this case, GBR_g and GBR_s were set to 128 and 64 kbps respectively. The traffic load destined to the Gold and Silver users were 160 and 80 kbps. Ten codes were allocated for the HS-PDSCH channels. Notice that both users are receiving at a higher rate than their GBR at all times. The additional throughput is being distributed proportionally after satisfying all users' guaranteed bit rates.

3) Scenario 3: In this scenario, we increase GBR_s for gold and silver users to 256 kbps and 128 kbps respectively and set the number of codes allocated to five to create a shortness of resources. The traffic load destined to the Gold and Silver users were 300 and 150 kbps. In Figure 8, notice that the gold user is receiving at its required GBR at all times. However due to capacity constraints the silver user is receiving at a lower rate (93 kbps) than its GBR since the total combined throughput reaches the cell capacity.

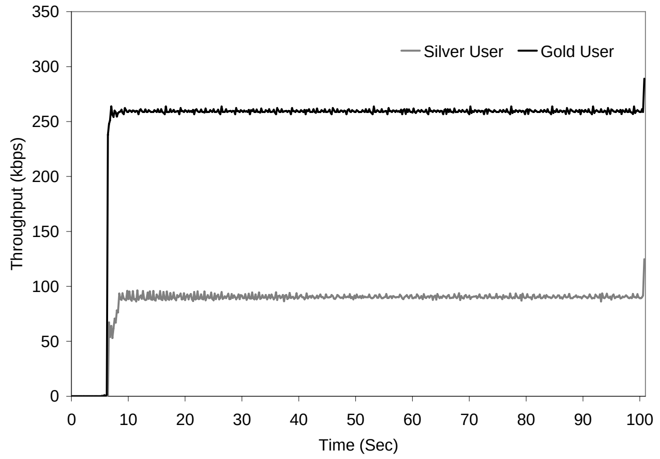


Figure 12: User throughput in SPS when total requested GBR exceed capacity

7. Conclusion

We developed and investigated two scheduling algorithms for HSDPA systems based on loose and strict enforcement of QoS aware policy constraints. We compared our algorithms against well-known schedulers using the HSDPA simulator that we developed in OPNETand showed that both algorithms perform significantly better in terms of supporting QoS constraints. They also significantly increase system throughput by scheduling multiple users in each slot. Our loose policy scheduler can also relax the QoS enforcement under bad radio conditions to increase throughput.

References

[1] H. Holma and A. Toskala, Eds., *HSDPA/HSUPA for UMTS*. Wiley, April 2006.

[2] T. Kolding, F. Frederiksen, and P. Mogensen, "Performance aspects of wcdma systems with high speed downlink packet access (hsdpa)," in *Vehicular Technology Conference*, 2002, pp. 477 – 481.

[3] X. Liu, E. Chong, and N. Shroff, "A framework for opportunistic scheduling in wireless networks," *Computer Networks Journal*, vol. 41, no. 4, pp. 451– 474, 2003.

[4] F. Kelly, "Charging and rate control for elastic traffic," *European Transactions on Telecommunications*, vol. 8, pp. 33 – 37, January 1997.

[5] A. Jalali, R. Padovani, and R. Pankaj, "Data throughput of cdma-hdr a high efficiency-high data rate personal communication wireless system," in *Vehicular Technology Conference Proceedings, 2000. VTC 2000-Spring Tokyo. 2000 IEEE 51st*, vol. 3, pp. 1854 – 1858.

[6] T. Kolding, "Link and system performance aspects of proportional fair scheduling in wcdma/hsdpa," in *Vehicular Technology Conference, 2003. VTC 2003-Fall. 2003 IEEE 58th*, vol. 3, pp. 1717 – 1722.

[7] Y. Ofuji, A. Morimoto, S. Abeta, and M. Sawahashi, "Comparison of packet scheduling algorithms focusing on user throughput in high speed downlink packet access," *The 13th IEEE International Symposium on Personal, Indoor and Mobile Radio Communications*, vol. 3, pp. 1462 – 1466, 2002.

[8] M. Andrews, K. Kumaran, K. Ramanan, A. Stolyar, P. Whiting, and R. Vijayakumar, "Providing quality of service over a shared wireless link," *Communications Magazine, IEEE, 2005.*, vol. 39, no. 2, pp. 150 – 154, Feb. 2001.

[9] P. Hosein, "Qos control for wcdma high speed packet data," in *4th International Workshop on Mobile and Wireless Communications Network*, 2002, pp. 169 – 173.

[10] F. Berggren and R. Jantti, "Multiuser scheduling over rayleigh fading channels," in *Proc. IEEE Globecom*, December 2003.

[11] K. Pedersen, "Quality based hsdpa access algorithms," in *Vehicular Technology Conference*, 2005, pp. 2498 – 2502.

[12] S. Elayoubi, T. Chahed, and G. Hbuterne, "Resource management in umts: Admission control and packet scheduling," in *ICON 2003*, October 2003.

[13] S. Pal, M. Chatterjee, and S. K. Das, "Call admission control and scheduling policies for umts traffic for qos provisioning," in *High Speed Networks and Multimedia Communications 7th IEEE International Conference, HSNMC 2004*, 2004.

[14] S. Hmlinen, P. Slanina, M. Hartman, A. Läppetelinen, H. Holma, and O. Salonaho, "A novel interface between link and system level simulations," in *ACTS Mobile Telecommun*, Oct. 1997, pp. 599 – 604.

[15] F. Frederiksen and T. Kolding, "Performance and modeling of wcdma/hsdpa transmission/h-arq schemes," in *Vehicular Technology Conference*, 2002, pp. 472 – 476.

[16] J. S. Gomes, M. Yun, H.-A. Choi, J.-H. Kim, J. Sohn, and H. I. Choi, "Scheduling algorithms for policy driven qos support in hsdpa networks," in *IEEE 65th Vehicular Technology Conference*, 2007. VTC 2007-Spring.

[17] J. S. Gomes, H.-A. Choi, J.-H. Kim, J. Sohn, and H. I. Choi, "Integrating admission control and packet scheduling for quality controlled streaming services in hsdpa networks," *Accepted for publication in IEEE Broadnets*, 2007.