Scheduling Algorithms For Policy Driven QoS Support in HSDPA Networks

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
{joegomes,mirayun,hchoi}@gwu.edu, mtamer@vt.edu, jayhoon.kim@gmail.com, {jgsohn,hichoi}@snu.ac.kr

Abstract— **Increasing demand for high data-rate multimedia services has led to the emergence of high-speed data transfer features such as High-Speed Downlink Packet Access (HSDPA) for WCDMA networks. The role of the MAC-hs scheduler is vital in HSDPA in improving overall system performance. Using Opportunistic scheduling to exploit multi-user diversity for efficient transmission of best effort services or considering user fairness constraints has been the main focus of most scheduling algorithms. However the need for strict QoS support for services such as streaming, gaming, and VoIP is growing. Service operators can enforce their own policies in meeting these QoS requirements. In this paper we introduce two QoS-aware policy driven scheduling algorithms. We developed an HSDPA system in OPNET, and implemented our scheduling algorithms along with other well-known algorithms. OPNET simulations show that unlike other schedulers, our Strict and Loose Policy Scheduling (SPS and LPS) algorithms comply with the policy constraints if allowed by radio conditions and cell capacity.**

I. INTRODUCTION

High Speed Downlink Packet Access (HSDPA) is introduced to increase WCDMA downlink packet data throughput in the Release 5 of the 3GPP UTRAN specifications. HSDPA offers peak data rates of up to 14 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 have been studied in depth in [2], [3]. Some alternative approaches include scheduling schemes that consider inter-cell transmission [4] and multi-user transmission in one slot [5]. [6] proposes an algorithm that provides QoS guarantees using barrier functions. However none of the existing schemes provide policy driven QoS guarantee. 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 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, propose two different algorithms, and illustrate their effectiveness in satisfying QoS policies.

II. HSDPA SYSTEM MODEL

In HSDPA service, a new downlink transport channel HS-DSCH (High Speed Downlink Shared Channel) was introduced, that carries data to the selected UE (or UEs) in each transmission time interval (TTI) of 2 ms. The transported bits from HS-DSCHs are mapped onto up to 15 physical downlink shared channels each using a separate orthogonal cdma code. The associated High Speed Shared Control Channel (HS-SCCH) is used to communicate control information between the UE and the base station (NodeB). In the downlink direction NodeB notifies to the selected UEs the information needed to decode HS-DSCHs, such as modulation type, code rate, CDMA code etc. In the uplink direction, the UE notifies the NodeB of the Channel Quality Indicator (CQI) and a positive or negative acknowledgement pertaining to the received frame. CQI indicates the instantaneous channel quality received by the user, so that the NodeB can adjust its transmission parameters for AMC to cope with variations in the channel conditions for the user. Moving the scheduling to NodeB enables a more efficient implementation of scheduler by allowing it to work with the instantaneous channel information. Several AMC schemes are proposed including QPSK and 16QAM with coding rates of 1/3, 1/2 and 3/4. 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 upto 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 time u_i can be scheduled next is $t + min\_TTI(k)$. Category 10 can theoretically support upto 14.4 Mbps using 10 codes and has minimum inter TTI interval of 1. Therefore an ideal scheduling algorithm must consider many different parameters, such as CQI values, the capabilities of individual handsets (UE category) and quality-of-service parameters associated with the data. In addition, a cellular operator may have certain policies to maximize revenue, e.g. giving priority to the heaviest data users with the highest Guaranteed Bit Rates (GBR). As a result the scheduling algorithm can affect an operator's finance by efficient management of sprectrum and cell capacity.

III. PROBLEM STATEMENT

We make the following notations and assumptions.

- The system consists of a single Node-B and n users $u_1, \cdots, u_n$, where each user u_i is associated with a UE category $k_i \in \{1, \cdots, 12\}$, and each UE category k_i is associated with $min\_TTI(k_i) \in \{1, 2, 3\}$.
- M is the number of codes allocated by the RNC for HS-DSCH channels.
- GBR_i denotes the guaranteed bit rate for user u_i.
- $CQI_i(t) \in CQI = \{0, 1, \cdots, 30\}$ 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 HSPDSCHs codes, and the instantaneous data rate associated to CQI x.
- $\gamma_i(t)$ is the allocated CQI for user i at TTI t. If i is not scheduled to receive any data at t then $\gamma_i(t) = 0$.
- Suppose user 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 i at TTI t over the last w_o TTIs defined as follows:

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

A. Loose Channel Allocation Policy

For every TTI t find $X(t) = (\gamma_1(t), \cdots, \gamma_n(t))$ 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) if $\gamma_i(t) > 0$, then $\sum_{k=t-min\_TTI(C_i)+1}^{t-1} \gamma_i(k) = 0$
In addition an effort should be made in terms of finding a balance between the two goals of maximizing throughput (using opportunistic scheduling) and satisfying QoS constraints (GBR). Hence we call it loose policy. Notice that this is different from the goal of PF algorithm, which disregards QoS constraint.

B. Loose Policy Scheduling Algorithm

The PF scheduler orders the receivers using the metric $\frac{r_i(t)}{th_i(t)}$ where $r_i(t)$ is the instantaneous data rate and $th_i(t)$ is the current throughput. However PF does not make any QoS guarantees. It considers all users to be 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 $\frac{DR(\gamma_i(t))}{\#CH(\gamma_i(t))} \cdot \frac{GBR_i}{th_i(t)}$. 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 discuss the algorithm in further detail.

Step 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 $\Delta_i(t) = w_0(GBR_i - Th_i(t-1)) + Th_i(t-1)$.

Step 2: Given $CQI_i(t)$, choose smallest $\gamma_i(t)$ such that (i) $\gamma_i(t) \leq CQI_i(t)$ and (ii) $DR(\gamma_i(t)) \geq \Delta_i(t)$.

Step 3: Sort the users in $U_{active}(t)$ in descending order of $\frac{DR(\gamma_i(t))}{\#CH(\gamma_i(t))} \cdot \frac{GBR_i}{th_i(t)}$.

Step 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 $\frac{DR(\gamma_i(t))}{\#CH(\gamma_i(t))} \cdot \frac{GBR_i}{th_i(t)}$ value 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.

Step 5: Repeat *Step 4*.

C. Strict Channel Allocation Policy

Now we will consider a stricter set of policy requirements. Suppose a service provider differentiates its users using Scheduling Priority Indicator (SPI) values, where $SPI \in C = \{1, 2.., 15\}$[1]. In general we assume higher classes (with higher SPI) have higher GBR values. Although the problem can be generalized for any number of SPI values, for simplicity let us consider only SPI values 1 and 2 with corresponding GBR values GBR_g and GBR_s. 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 i at TTI t, where $d_i(t) \in \{r_1, r_2, \cdots, r_k\}$ such that $r_1 = r_{min} < r_2 < \cdots < r_k = r_{max}$. We assume that a user belonging to a particular class is on average under good enough channel conditions 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 is enough resources, and to satisfy users from higher classes before the lower classes when there is not enough resources. Although it is the admission control algorithm's job to make sure that all admitted users can be satisfied, situations could arise due to user mobility when such guarantees cannot be made.

(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 under similar conditions according to their GBRs.

IV. 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 $i \in N$ 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 $M_c(th_i(t)) = P_c(th_i(t)) \cdot d_i(t)$ where P_c denotes the preliminary utility function which we define shortly. Then for any gold user g and silver user s the 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 $0 \leq th_g(t) < GBR_g$ and $0 \leq th_s(t) \leq r_{max}$.

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

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

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

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

Note that Condition (C1) and (C5) corresponds to policy (P1), (C2) and (C4) corresponds to (P3), and Condition (C3) corresponds to (P5).

We define functions P_{Gold} and P_{Silver} satisfying the conditions above as follows. Let $\beta = r_{max} - GBR_g$, $\alpha = \frac{r_{max}}{r_{min}}$ and x be the throughput of the user.

$$P_{Silver}(x) = \begin{cases} \alpha\beta + 1 & \text{if } 0 \leq x < GBR_s \\ r_{max} - \frac{GBR_g}{GBR_s}x & \text{if } GBR_s \leq x \leq r_{max} \end{cases}$$

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

Functions P_{Gold} and P_{Silver} are shown in Fig. 1. Notice that maximum P_{Silver} value for a silver user s with throughput $th_s(t) \geq GBR_s$ and maximum P_{Gold} value a gold user g with throughput $th_g(t) \geq GBR_g$, 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 $(\alpha\beta + 1)r_{min} > \beta r_{max}$. This satisfies conditions (C2) and (C4). Similarly if $th_s(t) < GBR_s$, $M_{Silver}(th_s(t)) \leq (\alpha\beta + 1)r_{max}$. For a gold user with $th_g(t) < GBR_g$, $M_{Gold}(th_g(t)) \geq (\alpha^2\beta + \alpha + 1)r_{min} > (\alpha\beta + 1)r_{max}$. 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 throughput 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). Conformance to (P4) can be shown using similar reasoning.

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

$$P_c(x) = \begin{cases} \alpha^c\beta + \alpha^{c-1} + \cdots + \alpha^0 & \text{if } 0 \leq x < GBR_c \\ r_{max} - \frac{GBR_{max}}{GBR_c}x & \text{if } GBR_c \leq x \leq r_{max} \end{cases}$$

A. 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.

Step 1. let i be the user such that $i = \underset{j}{argmax}\{M_{c_j}(th_j(t))\}$ where user j belongs to class c_j. Let cqi_i be its requested CQI, $codes(cqi_i)$ be the number of codes required.

(i) if $(codes(cqi_i) < codes\_left)$
    send to user i using data rate $DR(cqi_i)$
else
    send to user i at the maximum rate possible using $codes\_left$ number of codes.
(ii) update $codes\_left$ and remove i from U.

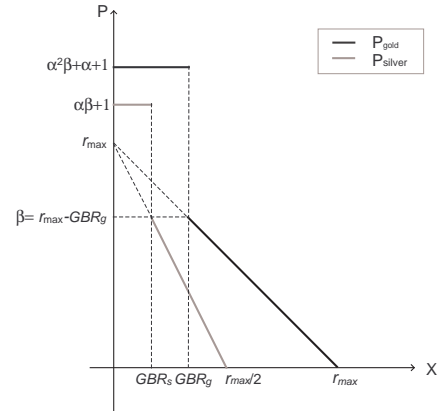Step 2. if $(codes\_left > 0$ and $U \neq \phi)$ goto Step 1.

Fig. 1. P_{Gold} and P_{Silver}

V. SIMULATIONS

We developed an HSDPA model by extending the UMTS model in the OPNET simulator. In the following section we briefly describe our simulator implementation.

A. OPNET Simultor

We implemented LPS and SPS algorithms in the $NodeB$ process model along with Max C/I and PF algorithms. $NodeB$ maintains a separate transmission queue for each mobile station (MS). An MS measures the SINR for each received packet and reports the corresponding CQI back to the NodeB. For this, actual value interface (AVI) tables[7] are used. For each UE category we use a separate AVI table that maps a CQI (or data rate) to a corresponding threshold SINR value. An UE reports the maximum possible CQI according to the received SINR for every received packet. Error decision for a received packet is also made based on the threshold SINR for the CQI

associated to the transmitted packet. We do not consider the 7.5 slot delay associated with CQI reporting, i.e. we assume perfect link adaptation. At each TTI the MAC-hs scheduler chooses and sends packets to a set of users according to their requested CQI. NodeB can fragment or concatenate packets based on the transport block sizes if necessary.

B. Simulation Setup

In order to evaluate our proposed schedulers we simulated several scenarios. We have a total of 20 HSDPA receivers (10 gold and 10 silver) in the network. 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
Max cell Tx power	20 W
Total HS-PDSCH power	8 W
Number of HS-PDSCH codes	5,10
ACK/NACK	Not modeled
CQI reporting interval	On every packet reception
HSDPA terminal category	1, 7
User receiver type	1-Rx Rake
Path loss model	Vehicular Outdoor
Shadow fading std.	10 DB
Power delay profile	ITU Veh-A
Site-to-site distance	2km
User speed	3mph
Window length for thpt measurement	1000 TTI

TABLE I

SUMMARY OF MAIN SIMULATION PARAMETERS

C. Evaluating Loose Policy Scheduling Algorithm

We compared our LPS algorithm with the Max C/I and PF scheduling algorithms under 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. 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 chooses the UE with best channel condition. Consequently the silver users receive at 128 kbps, whereas the gold users are totally 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.

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.
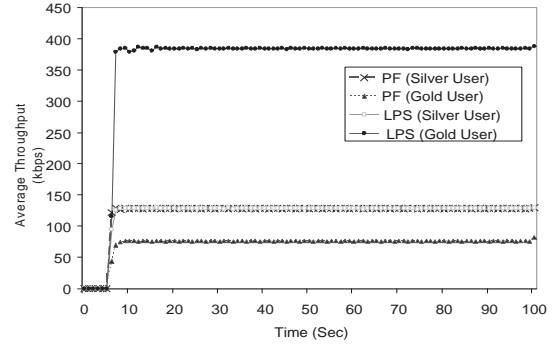

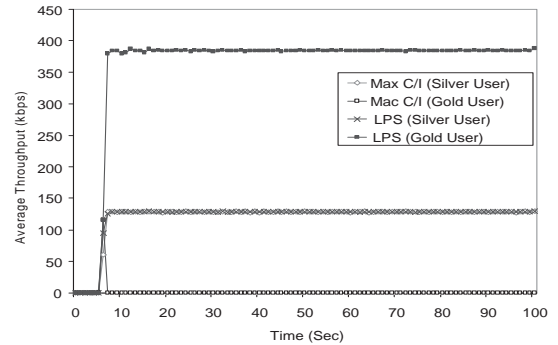Fig. 2. Comparing Loose Policy Scheduler and Proportional Fairness

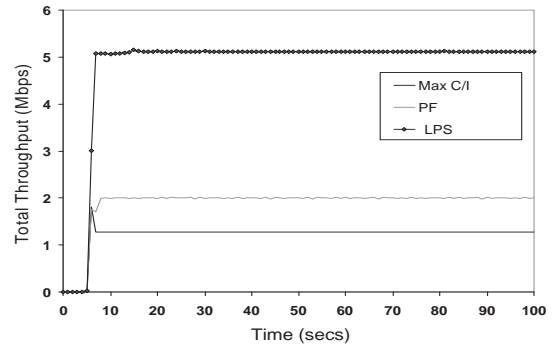Fig. 3. Comparing Loose Policy Scheduler and Max C/I

Fig. 4. Comparing total received throughput at the mobile stations

D. 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).
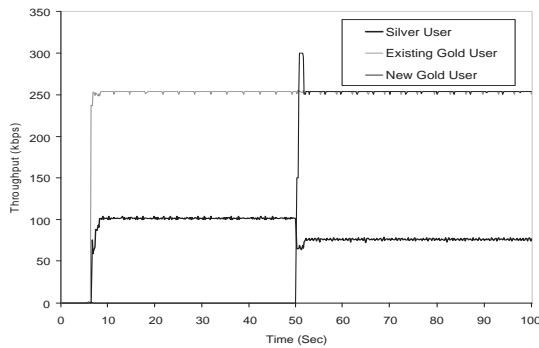

Fig. 7.   Traffic load is more than requested GBR

lower rate (93 kbps) than its GBR since the total combined throughput reaches the cell capacity.

Fig. 5.   Change in throughput when extra user added

Fig. 8.   Total requested GBR exceeds capacity

VI. 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 OPNET and showed that when radio conditions permit both the 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.
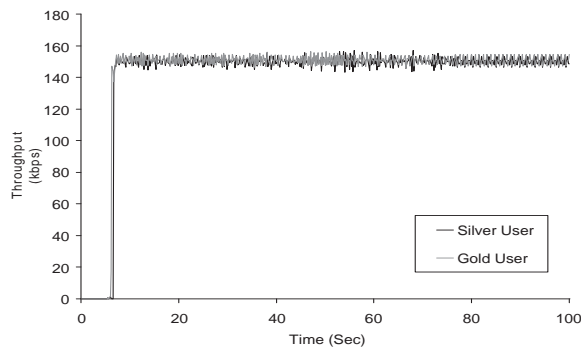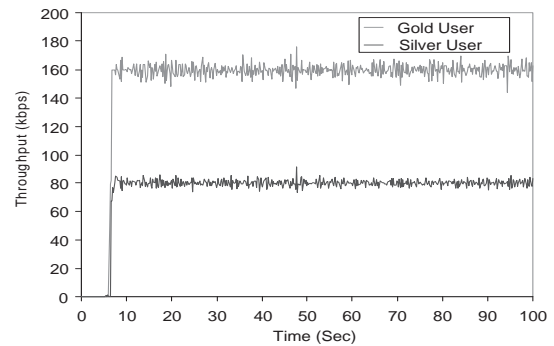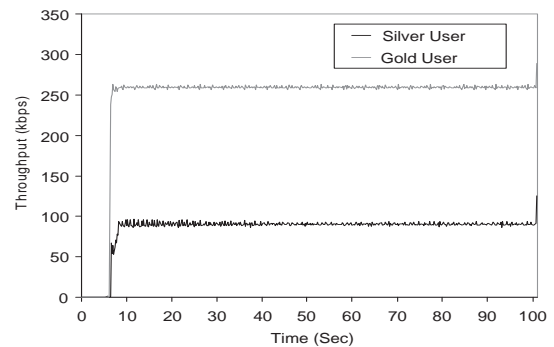

Fig. 6.   Gold and Silver throughput with Proportional Fair Scheduling

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. 10 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 GBRs for gold and silver users to 256 kbps and 128 kbps respectively and set the number of codes allocated to 5 to create a shortness of resources. The traffic load destined to the Gold and Silver users were also 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

REFERENCES

[1] H. Holma and A. Toskala, Eds., *HSDPA/HSUPA for UMTS.*   Wiley.
[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] T. Bonald, S. C. Borst, and A. Prouti`ere, "Inter-cell scheduling in wireless data networks," in *Proc. European Wireless Conference 2005*.
[5] F. Berggren and R. Jantti, "Multiuser scheduling over rayleigh fading channels," in *Proc. IEEE Globecom*, December 2003.
[6] P. Hosein, "Qos control for wcdma high speed packet data," in *4th International Workshop on Mobile and Wireless Communications Network*, 2002, pp. 169 – 173.
[7] S. Hmlinen, P. Slanina, M. Hartman, A. Lappetelinen, H. Holma, and O. Salonaho, "A novel interface between link and system level simulations," in *ACTS Mobile Telecommun*, Oct. 1997, pp. 599–604.