Enhanced Packet Scheduling Algorithm Providing QoS in High Speed Downlink Packet Access

Ohyun Jo*, Jong-Wuk Son†, Soo-Yong Jeon‡ Dong-Ho Cho‡

Department of Electrical Engineering and Computer Science
Korea Advanced Institute of Science and Technology (KAIST)
373-1, Guseong-dong, Yuseong-gu, Daejeon, Korea
Telephone: +82-42-869-3467, Fax: +82-42-867-4042
Daegu Gyeongbuk Institute of Science and Technology (DGIST)†
Email: *‡{ohjo,syjeon}@comis.kaist.ac.kr, †jwson@dgist.ac.kr, §dhcho@ee.kaist.ac.kr

***Abstract*— In High Speed Downlink Packet Access (HSDPA), Packet Scheduler is a key element for high-speed and efficient transmissions. In this paper, we propose an enhanced packet scheduling algorithm which computes the priority according to the Proportional Fair algorithm. That is, for users to need Hybrid ARQ retransmission in case of erroneous packets, the priority of the Proportional Fair algorithm is elegantly modified to reduce the Transfer Delay without degradation of system throughput.**

I. Introduction

High Speed Downlink Packet Access (HSDPA) is a packet-based data service supporting downlink data rates up to 14 Mbps over a 5 MHz bandwidth in the WCDMA (Wideband Code Division Multiple Access) system[1][2]. The HSDPA concept is applicable to a new type of transport channel, which is intended to carry interactive, background and optionally streaming traffic. The traffic is transmitted as a form of Transport Blocks on the High Speed Physical Downlink Shared Channel (HS-PDSCH). To improve the system capacity, HSDPA adopted advanced technologies such as Adaptive Modulation and Coding (AMC), Hybrid Automatic Request (HARQ), fast Packet Scheduling and advanced receiver design. These features are tightly coupled and enable adaptation of the transmission parameters to the instantaneous variations of radio channel quality[5].

HARQ is considered as a combination of forward error correction (FEC) and automatic repeat request (ARQ), and the receiving side of HARQ combines the soft information of multiple transmission of a transport block at bit level. Additional decoding gain is achieved by using HARQ[3][4].

The goal of the fast Packet Scheduling is to maximize the network throughput while satisfying the QoS of users. With the purpose of enhancing the cell throughput, the HSDPA scheduling algorithm can take an advantage of the instantaneous channel variations and temporarily raise the priority of the favorable users[5]. The packet scheduler located in the NodeB allows tracking the instantaneous variations of the user's supportable data rate according to the recent channel quality information.

In this paper, we propose an enhanced packet scheduling algorithm which computes the priority according to the Proportional Fair algorithm, in which users need HARQ retransmission, the priority is modified to reduce the transfer delay without degradation of the cell throughput.

A brief description of general MAC layer aspect of HSDPA is given in Section II. In Section III, the proposed packet scheduling algorithm is described in detail. And, The performance and evaluation of the proposed scheme are given in Section IV. Finally, some conclusions and further works are drawn in Section V.

II. MAC Protocol

The data link layer of HSDPA is composed of Radio Link Control (RLC) and Medium Access Control (MAC) layers. The RLC layer is located on top of the MAC layer and is responsible for reliable data transmissions. The MAC layer is in charge of radio resource allocation and packet scheduling. It decides on which transport channel the Protocol Data Units (MAC-d PDU) will be transmitted. For the HS-DSCH transport, the MAC has additional sub-layer called MAC-hs, which enables fast radio resource allocation. The MAC-hs at the NodeB schedules the HSDPA transmission and selects a suitable Transport Format for each Transport Block (TB) at each TTI (Transmission Time Interval).

The MAC-hs at the user equipment (UE) decodes the received MAC-hs PDUs in a soft buffer. It generates an ACK (or NACK) and reorders the received MAC-hs PDUs to compound the upper layer PDU(i.e., RLC PDU). The reordering operation utilizes a timer-based mechanism as follows; Timer (T1) starts when an MAC-hs PDU with Transmission Sequence Number (TSN) bigger than the next expected TSN is correctly received. If the Timer (T1) expires, all timed-out MAC-hs PDUs will be removed from the reordering buffer. These timed-out MAC-hs PDUs will cause the retransmission of RLC PDUs in the RLC layer[6].

Due to the reordering operation in the MAC-hs sub-layer, the performance of HSDPA can be degraded unnecessarily. For example, if a UE can not decode the received packets

1-4244-0063-5/06/$20.00 ©2006 IEEE

correctly and experiences channel quality degradation caused by path loss or shadowing, the UE may not be selected for retransmission by the scheduler because of the bad channel quality. As a result, the retransmission delay will be longer. In MAC-hs, the long retransmission delay may cause upper-layer retransmissions because of the operation of T1 timer. However, the retransmissions in the upper layer should be avoided as much as possible. Upper layer retransmission will cause much longer delay and resource waste. Due to the retransmission in upper layer, the timed-out MAC-hs PDUs in the MAC layer will be transmitted one more time even though it was successful. These redundant retransmissions will waste radio resources and increases the transfer delay, therefore a more efficient scheduling algorithm is required.

III. Proposed Scheme

The goal of the proposed scheduling algorithm in this paper is to schedule the users for interactive and background services. According to [5][7], the specification does not set any absolute quality guarantees on delay for interactive bearers, but interactive users expect to get the message as soon as possible.

The Proportional Fair (PF) scheduling algorithm is used to increase fairness, while still maintaining the cell throughput. The Proportional Fair scheduler serves the user with largest relative channel quality as follows,

$$P_i(t) = \frac{R_i(t)}{T_i(t)},\ i = 1, \ldots, N \tag{1}$$

Here, subscript i distinguishes N users in the cell, $P_i(t)$ denotes the user priority at instant t, $R_i(t)$ is the instantaneous data rate experienced by user i if it is served by the Packet Scheduler, and $T_i(t)$ is the user throughput up to instant t.

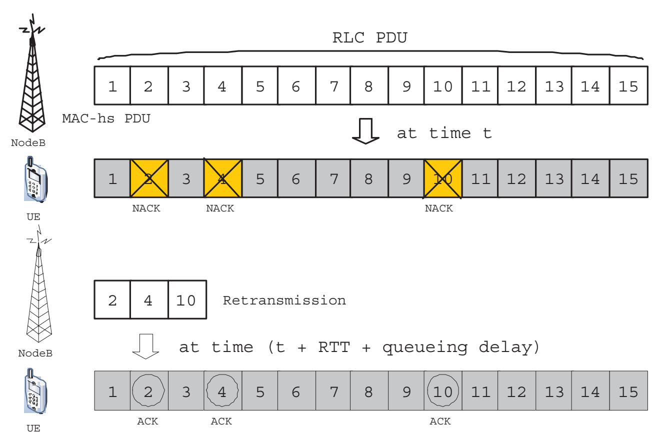

Fig. 1. Packet transmission procedure in HSDPA

It is however noticeable that the PF scheduling may cause a long retransmission delay in HSDPA. Fig. 1 shows the procedure that an RLC PDU is transmitted from a nodeB to a UE. In Figure. 1, the NodeB has an RLC PDU in the queue to send to the UE. The RLC PDU is composed of 15 MAC-hs PDUs. When the UE gets the highest priority which is calculated from Equation (1), NodeB begins transmitting the data in the queue in the unit of MAC-hs PDU. If the UE correctly received 12 packets out of 15 packets and 3 packets are broken at a certain instant t, the UE will request retransmissions for the three erroneous packets to the NodeB with NACKs. To notify the errors, it takes some time represented as RTT(Round-trip time). After RTT, NodeB puts the packets, which is broken before, into the queue for retransmission. NodeB should wait for the chance to transmit those packets until the UE gets the highest priority again. However, the opportunity for retransmission in a short time may not be guaranteed because the corresponding UE has already received many packets successfully. That means the denominator in equation (1) increased due to a number of successful packets. The increase of denominator in equation (1) results in the decrease of priority. Therefore, it may have a small opportunity to be assigned for retransmissions by the scheduler. Moreover, channel quality can be worse because the channel quality of the corresponding UE is randomly variable. So, the probability that the UE is assigned for retransmission will be decrease as the number of UEs increases. As a result, waiting time in the UE's soft buffer, namely retransmission delay, will be longer. The minimum retransmission delay could be RTT. However, it will be much longer than the lower bound in the case of the conventional Proportional Fair algorithm due to the reasons mentioned above.

In [8], Barriac proposed a modified Proportional Fair algorithm to equalize the user throughput. Thus, we propose a modified Proportional Fair algorithm, which is expressed as

$$y = \begin{cases} P_i(t) = \dfrac{R_i(t)}{T_i(t)} & \text{if } N_i = 0 \\ P_i(t) = \dfrac{R_i(t)}{T_i(t)} \cdot \dfrac{\{\overline{R_j(t)}\}_{max}}{\overline{R_i(t)}} & \text{if } N_i > 0 \end{cases} \tag{2}$$

Here, $\overline{R_t(t)}$ is the average channel quality of user i, $\{\overline{R_j(t)}\}_{max}$ indicates the maximum value of the average data rates of all users in the cell, and N_i is the number of packets to be retransmitted to user i.

The difference between the proposed and the conventional Proportional Fair algorithm is that a scaling factor is multiplied to the priority function for retransmission. The scaling factor, $\frac{\{\overline{R_j(t)}\}_{max}}{\overline{R_i(t)}}$ means the ratio of maximum value of average data rates of the all UEs in the cell to the average data rate of the corresponding UE. The scaling factor can raise the priority of the UE which has packets to be retransmitted. As a result, the UE which has some packets to be retransmitted will sometimes have the opportunity to transmit data even though the priority value calculated from the conventional Proportional Fair algorithm is not the highest. The proposed algorithm will decrease the delay for retransmission since the priority of a UE that waits for retransmission increases. Radio resource for transmission is allocated to a UE that has the maximum priority value by the scheduler at NodeB.

In this case, system performance degradation would be concerned. Because, the UE may have opportunity for data transmission even though the practical priority is not the highest. Sometimes, a UE have the best channel quality may miss the opportunity. Therefore, the error probability can be higher than before. It may cause system degradation. However, HARQ is adopted in HSDPA system. Using HARQ, the broken packet is not thrown away even though it is erroneous. The UE combines the newly arrived packet with the broken packet to get decoding gain through HARQ. Considering the decoding gain of HARQ, error probability of retransmitted packets will drastically decrease. Therefore, throughput degradation of the system is almost negligible.

IV. Performance evaluation

A. Anaylsis

The average delay for retransmission is given as

$$E[\tau] = \sum_{n=0}^{R} \{p(x=n) \cdot (ST + n \cdot (RTT + \sum_{k=0}^{n} u_k \cdot ST))\} \quad (3)$$

Here τ is the delay time which occurs during retransmission, p is the probability of error, R is the possible number of retransmission (R is four in HSDPA),RTT is Rount-trip time, ST is average service time for an user, u_k is the number of users who have higher priority than an user who waits for the chance to transmit data, and $p(x=n)$ is the probability that retransmission occurs n times. Then, the probability that retransmission occurs is given as follows.

$$\begin{aligned}
p(x=0) &= (1-p)^y \\
p(x=1) &= g(y) \\
p(x=2) &= \sum_{t_1=1}^{y} f(y,t_1) g(t_1) \\
p(x=3) &= \sum_{t_1=1}^{y} f(y,t_1) \big(\sum_{t_2=1}^{t_1} f(t_1,t_2) g(t_2) \big) \\
p(x=4) &= \sum_{t_1=1}^{y} f(y,t_1) \Big(\sum_{t_2=1}^{t_1} f(t_1,t_2) \sum_{t_3=1}^{t_2} f(t_2,t_3) g(t_3) \Big) \\
f(x,y) &= \binom{x}{y} p^x (1-p)^{x-y} \\
g(x) &= \sum_{y=1}^{x} \binom{x}{y} p^y (1-p)^x
\end{aligned}$$

Here, y is the number of MAC-hs packets composing an RLC PDU.$f(x,y)$ and $g(x)$ are functions to shorten the expressions.

If the number of retransmission goes over R, the packet will be finally considered as an error. And it will be dropped. The packet drop rate is numerically calculated as

$$p_{drop} = 1 - \sum_{n=1}^{R} p(x=n) \quad (4)$$

Using equation (3), the latency performance of the conventional PF scheduling algorithm and the modified PF scheduling algorithm can be estimated. If we assume the channel is randomly changing and the distribution is normal distribution, the expectation of u_k in equation (3) is just half of total number of all the other UEs, which is $\frac{N-1}{2}$, in the case of the conventional PF scheduling algorithm.

In the proposed scheme, the priority is raised due to the scaling factor, namely $\frac{\{\overline{R_j(t)}\}_{max}}{R_i(t)}$. So, u_k decreases as the priority increases. Following analysis shows the efficiency of the proposed scheduling algorithm compared to the conventional algorithm.

If we assume the channel quality index $R(t)$ is a random variable which is normally distributed, we can estimate the expectation of $\overline{R(t)}_{max}$. $\overline{R(t)}_{max}$ is also a random variable and means the maximum value of average data rates of the all UEs in the cell. Window Size(WS) should be defined for average data throughput. WS is defined as a multiple of transmission time interval(TTI). Setting WS to be 3000, $\overline{R(t)}$ is the average delay during 3000 TTIs. If we assume that the variance of $R(t)$ is σ and $R(t)$ is stationary, the variance of $\overline{R(t)}$ with window size WS will be $\frac{\sigma}{\sqrt{WS}}$. Because $\overline{R(t)}$ is the average of $R(t)$, $R(t-TTI)$, $R(t-2TTI)$, $\cdots$, $R(t-WS \cdot TTI)$. However, each data throughput value at each TTI during a short period such as 3000 TTIs is usually correlated with each other. That means the channel qualities within a short time are pretty similar. The correlation of the channel quality in the period depends on the speed of channel variation. If the channel of a UE is changing slowly, the correlation is strong. Otherwise, the channel quality is totally independent at each TTI, and there is no correlation in channel quality at each time. So, in this paper we define one more variable, namely α, which is a variable to consider the speed of channel variation. If the speed of channel variation is slow, the channel quality at each TTI will be more correlated. So, the variance of the average data rate of each user will be larger than $\frac{\sigma}{\sqrt{WS}}$. Let α be a variable to represent the increase of the variance. So, α can be used to indicate of the speed of channel variation. Considering the channel variation, the variance of $\overline{R(t)}$ will be represented as $\alpha\frac{\sigma}{\sqrt{WS}}$, rather than $\frac{\sigma}{\sqrt{WS}}$.

The average delay of the proposed algorithm is numerically calculated from the estimation of u_k^*, where u_k^* means the number of users that have higher priority than the user who waits for the chance to transmit data and is served by the proposed scheduling algorithm. Since we assume that $R(t)$ is normally distributed, error function, $erf(x)$, will be useful for the estimation of u_k^*. The expectation of $\overline{R(t)}_{max}$ can be estimated as

$$E[\overline{R(t)}_{max}] = erf^{-1}(1 - \frac{2}{N}) \cdot \sigma^* + \mu \quad (5)$$

Here, μ is the average of $R(t)$ and $\overline{R(t)}$, and σ^* is the variance of $\overline{R(t)}$, that is, $\alpha\frac{\sigma}{\sqrt{WS}}$. Using the relation of the variances of $\overline{R(t)}$ and $R(t)$, u_k^* is calculated as

$$E[u_k^*] = \left(1 - \frac{erfc\left(\frac{E[\overline{R(t)}_{max}] - \mu}{\sigma}\right)}{2}\right) \cdot (N-1) \qquad (6)$$

In equation (6), $E[\overline{R(t)}_{max}]$ is larger than μ. So, $\frac{erfc\left(\frac{E[\overline{R(t)}_{max}] - \mu}{\sigma}\right)}{2}$ is less than $\frac{1}{2}$. Thus, $E[u_k^*]$ which is calculated in the proposed algorithm is obviously less than u_k for conventional algorithm..

B. Numerical results

Numerical results from the analysis are as following. In equation (3), we assume that the number of MAC-hs PDUs that compose an RLC PDU is 50, round-trip time is 12ms, and average service time is 20ms. Figure 2 shows the increase of delay as error probability increases. Moreover, it shows that an increase of α, which means slowly changing channel condition, leads the better performance.

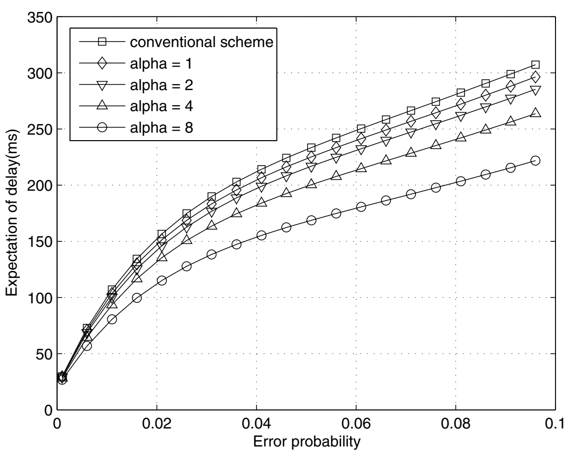

Fig. 2. Expectation of delay vs packet error probability

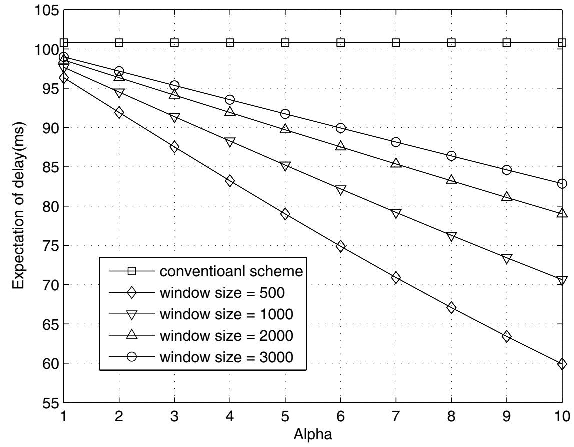

Fig. 3. Expectation of delay vs Alpha value

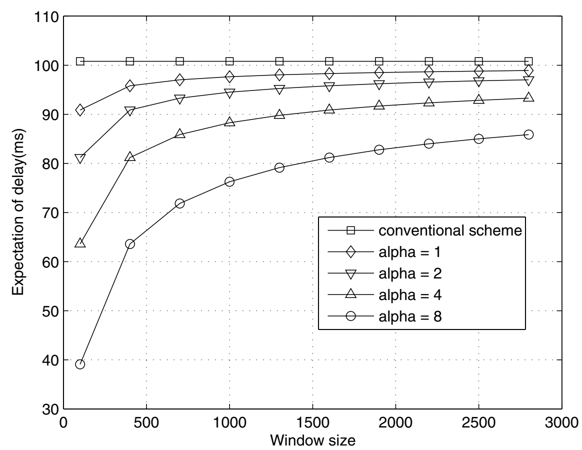

Fig. 4. Expectation of delay vs window size

In Figure. 3 and Figure. 4, we observe the effect of α and window size. Error probability is assumed to be 0.01 in Figure. 3 and Figure. 4. As α increases and window size decreases, it shows the better performance. That is because large α or small window size makes $\frac{\{\overline{R_j(t)}\}_{max}}{\overline{R_i(t)}}$,which is the scaling factor in equation (2), increase. As a result, UEs which have some packets to be retransmitted can have higher priority. It reduces waiting time for retransmission, which makes queuing delay longer.

C. Simulations

The simulation parameters are determined as follows. We employed 19-hexagonal cell model, where each cell is not sectionized(no sector) and site to site distance is 2800m. The number of UEs in each cell is 20, and the velocity of each UE is 10km/h. We assigned the position of each UE within the cell with a uniform distribution. The number of codes for HS-PDSCH is 15, and the carrier frequency is 2.19GHz. 5 level of MCS are considered, which are QPSK with R=1/4, 1/2, 3/4 and 16QAM with R=1/2, 3/4. Maximum number of retransmission is four. The channel estimation is assumed to be ideal and data transmission through HS-SCCH and HS-DPCCH is considered to be error free. Roundtrip time delay(RTT) is 12ms. Also, channel model is 1 path rayleigh fading, and pass loss is proportional to $1/d^4$, where d is distance. FER data in [9] is used for the execution of link layer simulation.

Figure. 5 shows the delay which is measured in the simulation according to traffic load. The proposed Proportional Fair scheduler gives about 30% reduction of retransmission delay. Average retransmission delay means the waiting time in soft buffer until all MAC-hs PDUs which compose an RLC PDU are transmitted successfully.

Figure. 6 shows the increase of error rate according to the increase of scaling factor in equation (2). Two cases are observed. First, we observe the increase of error rate when data packets are firstly transmitted. In this case, we can not

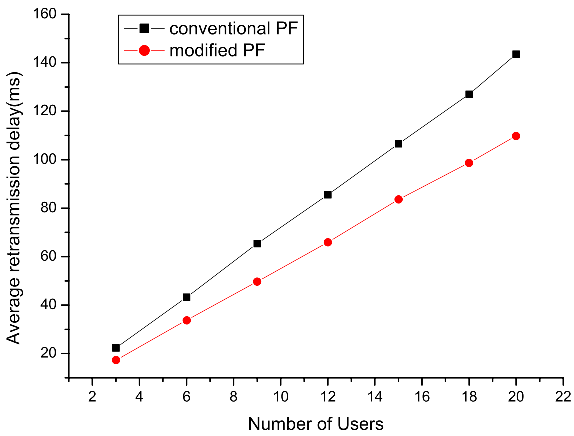

Fig. 5. Delay vs number of UEs

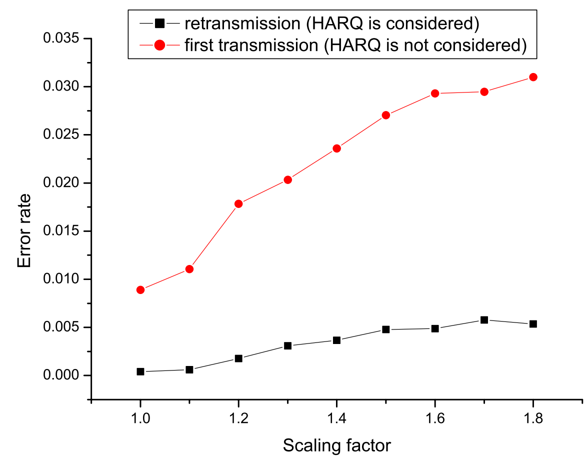

Fig. 6. Error rate vs scaling factor

get the decoding gain from HARQ. As a result, the error rate increases seriously as the scaling factor increases. Otherwise, if we take a look at the error rate of retransmitted packets, the increase of error rate is much less severe than the increase of error rate in the case of first transmission. That is because there is a decoding gain of HARQ through chase combining.

In fact, the error rate is somehow raised inevitably in the proposed scheduling algorithm, because a UE which has some packets to be retransmitted may have higher priority that any other UEs which have better channel quality. However, the increase of error rate due to the modified Proportional Fair algorithm is almost negligible in the case of retransmission. That is why the proposed algorithm is not suffered from performance degradation.

V. Conclusions and Further Works

This proposed algorithm is efficient in view of increasing the priority of users with packets to be retransmitted. The scheduler at the NodeB selects a UE that has the maximum priority value, and allocates radio resources to the UE for transmission. Although the priority of user to need HARQ retransmission is increased, the system throughput will not be degraded because of the decoding gain of HARQ. On the other hand, the number of RLC retransmission will be reduced and the transfer delay will be reduced, too.

By using the proposed algorithm, we can expect that the proposed scheduling algorithm reduces the transfer delay without the degradation of system throughput.

It will be interesting to investigate the optimal scaling factor which can reduce the delay as much as possible while the system performance is not degraded.

VI. Acknowledge

This research was supported in part by University IT Research Center Project.

References

[1] 3GPP, Technical Specification Group Services and System Aspects. "UTRA High Speed Downlink Packet Access (HSDPA); overrall desciption; Stage 2. (3GPP TS 25.308 version v6.3.0)."

[2] Kolding T. et al. "High Speed Downlink Packet Access: WCDMA Evolution," IEEE Vehicular Technology Society News, Volume 50, pp.4-10. Feb. 2003.

[3] Pal Frenger, Stefan Parkvall, Erik Dahlman, "Performance Comparison of HARQ with Chase Combining and Incremental Redundancy for HSDPA," Vehicular Technology Conference, 2001 VTC 2001 Fall. IEEE VTS 54th, Volume 3, pp.7-11 Oct. 2001.

[4] A. Das, F. Khan, A. sampath, H. Su, "Adaptive, asynchronous incremental redundancy (A/sup 2/IR) with fixed transmission time intervals (TTI) for HSDPA," Personal Indoor and Mobile Radio Communications, 2002. The 13th IEEE International Symposium on, Volume 3, pp.15-18 Sep. 2002.

[5] Pablo José, Ameigeiras Gutiérrez, "Packet Scheduling And Quality of Service in HSDPA," Ph.D Thesis. Oct. 2003

[6] 3GPP. Technical Specification Group Services and System Aspects. "Medium Access Control (MAC) protocol specification (3GPP TS 25.321 version v6.2.0)"

[7] 3GPP. Technical Specification Group Services and System Aspects. "Quality of Service (QoS) concept and architecture. (3GPP TS 23.107 version v6.2.0)"

[8] Barriac G., et al. "Introducing Delay Sensitivity into the Proportional Fair Algorithm for CDMA Downlink Scheduling," Spread Spectrum Techniques And Applications, 2002 IEEE Seventh International Symposium on, Volume 3,pp.652-656, 2002.

[9] M. Dottling, J. Michel, and B. Raaf, "Hybrid ARQ and adaptive modulation and coding schemes for high speed downlink packet access," in Proc. 13th IEEE International Symposium on Personal, Indoor, and Mobile Radio Communication, Volume 3, pp.1073-1077, Sept. 2002.

[10] T.B. Sorensen, M. Pons, "Performance Evaluation of Proportional Fair Scheduling Algorithm with Measured Channels," Vehicular Technology Conference, 2005 VTC 2005 Fall, Volume 4, pp.25-28 Sept. 2005.