Performance of Scheduling Algorithms for HSDPA

Matthias Malkowski, Andreas Kemper, Xiaohua Wang
Communication Networks, RWTH Aachen University, Faculty 6
Kopernikusstr. 16, 52074 Aachen, Germany
Email: {mal|kem|xwg}@comnets.rwth-aachen.de

***Abstract*— The High Speed Downlink Packet Access (HSDPA) is one of the newest extensions to the Universal Mobile Telecommunications System (UMTS). HSDPA allows for higher data rates due to new Adaptive Modulation and Coding (AMC) techniques, a Hybrid Automatic Repeat Request (H-ARQ) protocol and a fast scheduling algorithm. The fast scheduling, taking into account the current channel conditions of each user, contributes significantly to the performance of HSDPA by the so called multiuser diversity gain.**

Within the scope of this paper several scheduling algorithms are introduced and are analyzed with respect to both, the overall system performance and the individual user Quality of Service (QoS) constraints. Results illustrate the trade-off between the conflicting targets of high cell throughput and the delay requirements for real-time services.

I. Introduction

In order to meet the increasing demand for high data rate multimedia services, the 3rd Generation Partnership Project (3GPP) has developed a new high speed data transfer feature named High-Speed Downlink Packet Access (HSDPA) in Release 5 specifications.

The HSDPA concept has been designed to increase the downlink packet data throughput by means of fast physical layer retransmission and transmission combining, as well as fast scheduling and link adaptation, controlled by the Node B. Among all the new features, the fast scheduling is a key component that has a significant impact on the performance of the whole system.

The Node B based scheduling principle is shown in figure 1. The scheduler is located in Node B rather than in RNC, as it was the case for Release 4. In this way, the delays for the scheduling process are minimized. Additionally, the radio channel condition is measured and reported as the Channel Quality Indicator (CQI) [1] by each mobile, which allows for the channel aware scheduling so that mobiles with temporarily good channel conditions can be favored. This exploitation of multiuser diversity can significantly increase the system capacity. On the other hand, the issues of fairness and QoS among different users should be considered. While there are several publications on the theoretical benefits of different scheduling mechanisms in general, hardly any experience on their behavior in a typical HSDPA scenario is known.

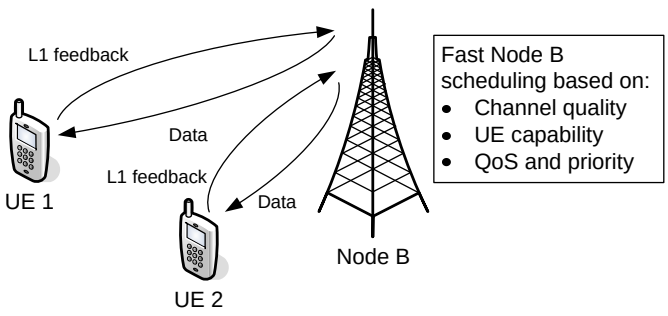

Fig. 1. Node B scheduling principle

This paper is structured as follows. In section 2 we introduce our system model used for performance evaluation, while the analyzed scheduling algorithms are explained in section 3. Section 4 presents our results and the paper is concluded by section 5.

II. System model

A. System overview

The system we use for the performance evaluation of the scheduling algorithms is modeled with the relevant MAC-hs [2] protocol, a physical layer and a traffic generator. The Transmission Time Interval (TTI) equals 2 ms and one High Speed Shared Control Channel (HS-SCCH) is configured in our simulations. Hence, up to one user can be scheduled in each TTI. The model for the physical layer is simplified with a fixed BLER equal to 10 % when AMC is applied. For each user, the mobility has a normal distribution with the mean speed value equal to 3 m/s and variance equal to 1. All the users are confined to a fixed area and move in a Brownian manner. According to [3], an analytical formula can be used for the CQI generation in AWGN channel conditions:

$$SNR = \frac{\sqrt{3} - log_{10}(CQI)}{2} \cdot log_{10}(BLER^{-0.7} - 1) \ldots + 1.03 \cdot CQI - 17.3 \quad (1)$$

B. QoS classes and service differentiation

The UMTS standard defines four QoS classes [4] which are the conversational class, streaming class, interactive class and background class. These classes mainly differ in the way they balance between the transmission delay and reliability. According to [5], HSDPA focuses on the last three classes.

When classifying services according to their delivery requirements, the concept of Real Time (RT) and Non Real Time (NRT) services is introduced. Usually, RT services have been considered to impose strict delay requirements on the end-to-end communication. As a result, the involved network nodes in the RT traffic have to transfer the packets within a maximum tolerable delay. Due to these severe delay constraints, the error correction possibilities of this type of communication

are very limited. On the other hand, NRT traffic is commonly considered as error sensitive, though with less demanding delay constraints than RT traffic. These characteristics of NRT traffic allow for link and also end-to-end level recovery mechanisms, enabling an error free delivery of the payload.

The above mentioned QoS classes in UMTS can be grouped by these two categories. Namely, the conversational and streaming traffic can be identified with RT services, whereas the interactive and background services belong to a NRT traffic pattern. For clarity and structuring purposes, this paper distinguishes between NRT and RT traffic. NRT (i.e. best effort) services require payload to be transferred error free, whereas delay requirements still allow for end-to-end error recovery mechanisms such as carried out by TCP. In contrast, RT services have delay requirements which exclude end-to-end retransmission mechanisms. Hence, they are using unreliable transport protocols like the UDP.

C. Traffic modeling and performance metrics

During the simulation, a traffic generator is employed to create MAC-d PDUs whose size equals to 336 bits. For NRT services, a full queue model is applied. It clamps the buffer occupancy of the corresponding MAC-hs priority queue at a constant level. Accordingly, there is always sufficient data available for transmission to each mobile. For RT services, a constant data rate model is realized by creating the same amount of MAC-d PDUs periodically, e.g. every 5 TTIs.

The performance metrics for NRT services are mostly user and system throughput. For RT services, the delay experienced by MAC-d PDUs and the packet loss rate due to the discard timer [6] are the main evaluation metrics. In addition, the inter-scheduling interval is measured to compare the fairness of different scheduling algorithms. The interval refers to the time period between two consecutive scheduling events for each user.

III. Scheduling algorithms

A. Maximum SINR (maxSINR)

This scheduling algorithm serves in every TTI the user with best channel conditions and, therefore, the highest instantaneous supportable data rate. The serving principle has obvious benefits in terms of cell throughput. Consequently, under idealized conditions it is the system throughput optimal scheduler. Mathematically seen, it schedules user

$$j = \arg\max_i \{R_i(t)\} \tag{2}$$

at time t. $R_i(t)$ is the instantaneous data rate experienced by user i if it is served by the packet scheduler. The main disadvantage of this approach is the inherent unfairness. For instance, when a User Equipment (UE) is far away from the base station and its mobility is low, it may never be scheduled.

B. Proportional Fair (PF)

The PF scheduling algorithm was initially proposed in [7] and further analyzed in [8] and [9]. According to [10], the PF scheduler serves the user with best relative channel quality:

$$j = \arg\max_i \{\frac{R_i(t)}{\lambda_i(t)}\} \tag{3}$$

Here, $R_i(t)$ is as defined above and $\lambda_i(t)$ is the average data rate for user i. This rule ranks the users according to their instantaneous channel quality relative to their own average channel conditions. Accordingly users with a higher average throughput are not necessarily privileged. In this way, not only the multiuser diversity can be exploited, but at the same time also the issue of fairness among the users is taken into account.

Up to now, the presented scheduling methods do not take into account the delay experienced by each individual user. As a result, they are not suitable for scheduling of RT services. In order to meet this requirement, several QoS based scheduling methods have been proposed. Relevant examples of them are introduced below.

C. Modified Largest Weighted Delay First (M-LWDF)

M-LWDF as proposed by [11] is an algorithm to keep the probability of delayed packets exceeding the discard bound below the maximum allowed SDU error ratio

$$Pr(D_i > T_i) \leq \delta_i \ , \tag{4}$$

where D_i indicates the Head of Line (HOL) packet delay of user i, T_i represents the delay bound and δ_i is the allowed percentage of discarded packets. This M-LWDF scheduler selects the user

$$j = \arg\max_i \{a_i \cdot \frac{R_i(t)}{\lambda_i(t)} \cdot D_i(t)\} \ , \tag{5}$$

where the term a_i is a constant used for QoS differentiation. Consequently, varying services can have different δ so that the priority between users with different demands in terms of error rate can be adjusted. According to the suggestion of [12], an approximating practical rule for choosing a_i is $a_i = \frac{-\log(\delta_i)}{T_i}$. The term $\frac{R_i(t)}{\lambda_i(t)}$ is derived from the PF algorithm and $D_i(t)$ means the HOL packet delay.

By combining the PF metric and the HOL delay, this algorithm not only takes advantage of the multiuser diversity available in the shared channel through the PF algorithm. Furthermore, it also increases the priority of flows with HOL packets close to their deadline violation. However, the value of the HOL delay $D_i(t)$ has a significant impact on the total scheduling priority. An extreme case occurs for the priority equal to zero with $D_i(t) = 0$, which means that all the SDUs have to wait for the increase of the priority. This wait duration is a kind of intrinsic delay experienced by each SDU, i.e. MAC-d PDU in HSDPA.

D. Expo-Linear (EL)

To avoid the intrinsic delay in M-LWDF, some other algorithms have been proposed. One of the examples is the Expo-Linear algorithm proposed in [13]. It schedules user

$$j = \arg\max_i \{a_i \cdot \frac{R_i(t)}{\lambda_i(t)} \cdot e^{a_i D_i(t)}\} \ , \tag{6}$$

where the a_i, $R_i(t)$, $\lambda_i(t)$ and $D_i(t)$ have the same meaning as in M-LWDF mentioned above.

This algorithm introduces an exponential term to better equalize the weighted delay. When the HOL delay is low, mostly the PF metric dominates the scheduling decision. When the HOL delay approaches the delay bound, the total priority increases in an exponential manner. In the following we use the EL algorithm as a candidate for those scheduling algorithms which takes the delay constraints into account.

IV. Results

Basic differences in the behaviour of PF and EL algorithms are at first shown for RT only users. Relevant details of the simulation scenario are listed in table I

The 9 UEs are separated in 3 groups with different pathloss scopes as indicated by their in average perceived CQI values in table II. In contrast, all UEs have a constant rate data source, generating in parallel 3 MAC-d PDUs every 5 TTIs (i.e. 10 ms), each containing 336 bits of data. Consequently, the resulting source data rate per UE is 100.8 kbps.

A. Scheduler comparison for RT only services

As previously stated, for RT services in particular the packet delay has to be limited to a maximum tolerable value, otherwise packets are discarded. Hence, especially packets from users experiencing bad pathloss conditions should be transmitted within the 400 ms delay bound. Comparing the behaviour of the PF and the EL algorithms in figure 2 and figure 3 it instantly turns out that EL distributes queuing delays more evenly among users than PF does. As presented later this is realized by adapting the inter-scheduling intervals for the different groups.

The detailed parameters regarding the performances of PF and EL are compared in table III. From the mean values of queuing delay with the two scheduling algorithms listed in the table, it can be seen that PF scheduler does not take the delay into consideration. Actually, since the frequency of MAC-d PDU generation is 300 PDUs per second, there is a critical packet loss in UE 1-6 due to the expiration of the discard timer. In contrast, the EL scheduler provides for most of the UEs a better QoS in terms of service delay. With this scheduler, UE 4-9 have almost no packet loss during the whole simulation time. Packet loss of UE 1-3 become much less than that of PF scheduler.

Apart from disruptions by retransmitted packets, the individual queuing delay is largely dependent on the corresponding inter-scheduling interval as shown in figure 4 and figure 5. Contrary to the delay distributions, the inter-scheduling interval graduation depends on the chosen algorithm. Different to

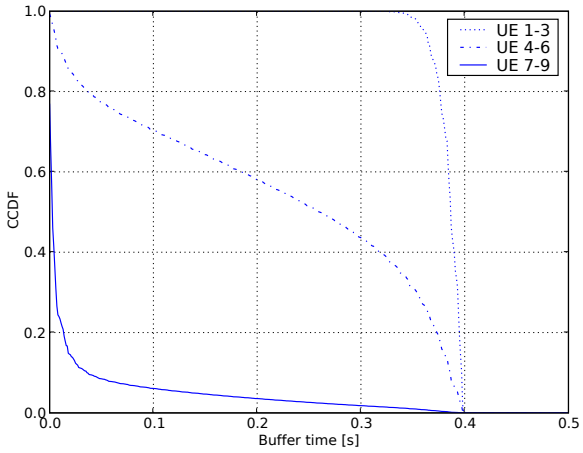

Fig. 2. MAC-d PDU queuing delay for Proportional Fair

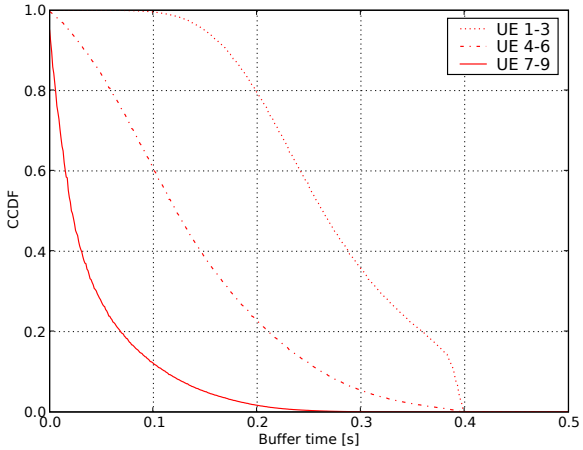

Fig. 3. MAC-d PDU queuing delay for Expo-Linear

TABLE I
Scenario details and configuration parameters

Parameter	Value
Simulation time	10000 s
Number of UEs	9
Pathloss	Variable, grouped
Traffic model	Constant, 100.8 kbps
MAC-d PDU arrival interval	5 TTIs
Number of HS-PDSCH codes	5
Number of HS-SCCH codes	1
UE category	6
BLER	10 %
Maximum number of retransmissions	4
Transmit windows size	12
Receive window size	12
Release timer	140 ms
Maximum delay	400 ms
CQI feedback cycle	2 ms
CQI repetition factor	1
Filter length	50 TTIs
δ (allowed fraction of discarded packets)	0.01
Throughput measurement interval	50 TTIs

TABLE II
Pathloss in different UE groups

	UE 1-3	UE 4-6	UE 7-9
Pathloss	High	Average	Low
Mean CQI	11	15	19

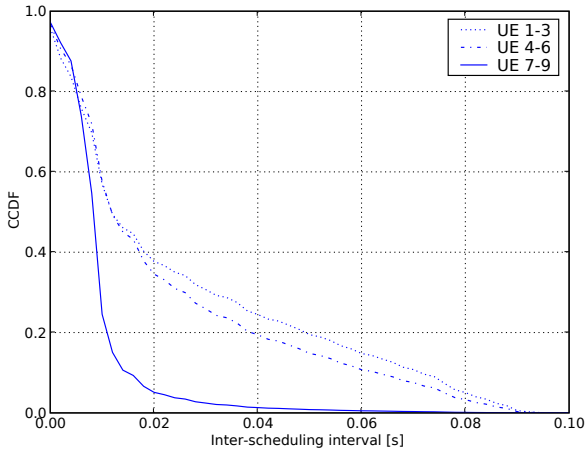

Fig. 4. MAC-hs inter-scheduling interval for Proportional Fair

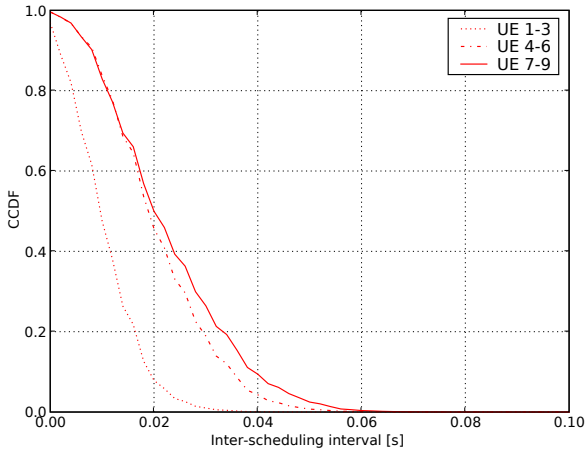

Fig. 5. MAC-hs inter-scheduling interval for Expo-Linear

PF, EL tries to compensate larger delays of distant users at the expense of near users. According to the comparisons of PF and EL scheduler in the previous sections, we can see that the PF scheduler is not suitable for the delay sensitive services.

B. Performance for mixed services

In this section, the performance of PF and EL with a mixture of RT and NRT users are compared. The important differences in the simulation scenario, compared to the previous set-up, are listed in table IV. We group the 9 users by different traffic models. The first 3 users have a full queue model, which is used for simulation of NRT services. The rest of the users have constant data rate services. For UE 4-6, these are realized by

TABLE III
COMPARISON OF MEAN QUEUING DELAY AND PACKET LOSS

Algorithm	Mean queuing delay [ms]			PDUs per second		
	UE 1-3	UE 4-6	UE 7-9	UE 1-3	UE 4-6	UE 7-9
PF	385.2	224.8	21.7	140	257	299
EL	275.1	140.4	44.2	283	300	300

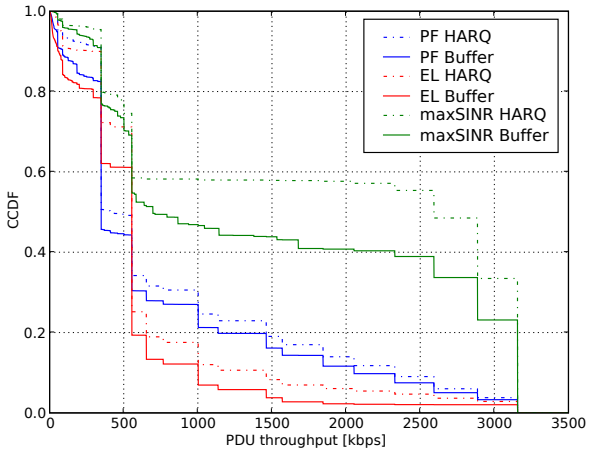

Fig. 6. MAC-hs PDU throughput with mixed services

generating 2 MAC-d PDUs with a size of 336 bits every 5 TTIs (equals to 10 ms). The same procedure is used for the simulation of services for UE 7-9, but with 3 instead of 2 PDUs every 10 ms. In contrast to the previous scenario now the average pathloss is the same for all UEs.

Since the number of HS-SCCH codes is set to 1, the MAC-hs PDU throughput can be considered as the aggregate cell throughput. Thus figure 6 shows the throughput measured before and after the reordering buffer. Here it can be observed that in general the cell throughput of PF scheduler is higher than that of the EL scheduler. The corresponding mean values are listed in table V. The maxSINR scheduler which always schedules the UE with the best channel conditions achieves the highest throughput.

Since UE 1-3 are NRT users, here the throughput is the critical factor for evaluation. While looking into figure 7, it turns out that the PF scheduler again provides higher throughput for the NRT users. The mean UE throughput with PF is 129 kbps, while the mean UE throughput with EL is only 46 kbps.

On the contrary, as shown in figure 8, the EL scheduler provides a better QoS for the RT users, whose services are delay sensitive. The mean MAC-d PDU queuing delay and number

TABLE IV
SIMULATION SCENARIO MIXED SERVICES

	UE 1-3	UE 4-6	UE 7-9
Traffic model	Full queue	67.2 kbps	100.8 kbps
Delay [ms]	5000	1000	400

TABLE V
MEAN CELL THROUGHPUT

Algorithm	HARQ throughput [kbps]	Buffer throughput [kbps]
maxSINR	1905	1534
PF	859	768
EL	695	550

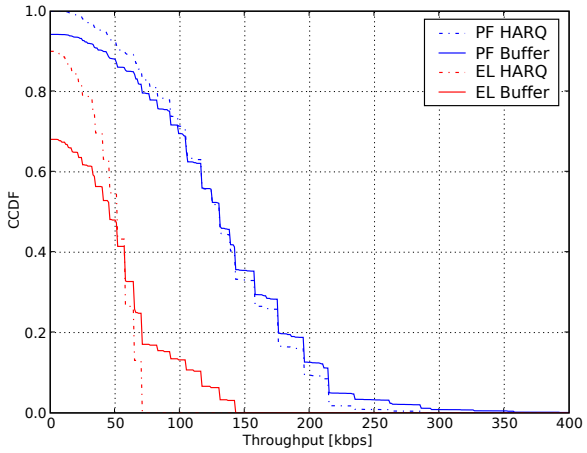

Fig. 7. MAC-hs UE throughput of NRT service

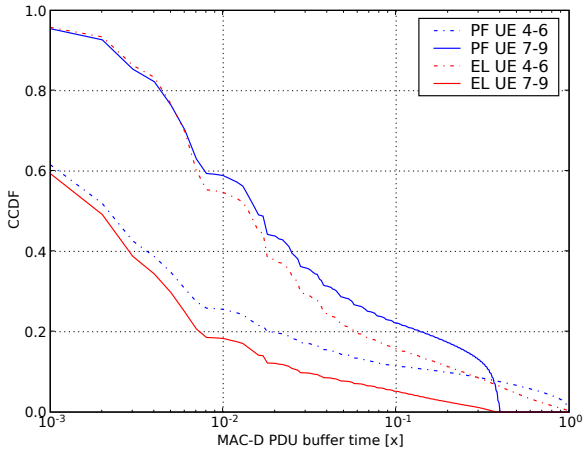

Fig. 8. MAC-d PDU queuing delay of RT users

of transmitted MAC-d PDUs per second on UTRAN side are listed in table VI. It can be observed from these results that there are packet losses due to discard timer expiration when PF scheduler is applied. Compared with PF, EL scheduler ensures that there is no packet loss due to delay bound for the delay sensitive services. Furthermore, the most delay sensitive services (UE 7-9) get the lowest MAC-d PDU queuing delay, which means they are the most prioritized during scheduling. The maxSINR scheduler showed to be not applicable for RT services because it neither takes delay bounds into account nor it generates a fair share of the available resources.

TABLE VI
MAC-D PDU DELIVERY STATISTICS

Algorithm	Buffer delay [ms]		PDUs per second		Discarded packets	
	UE 4-6	UE 7-9	UE 4-6	UE 7-9	UE 4-6	UE 7-9
maxSINR	98.8	22.0	82	110	59.2 %	63.4 %
PF	74.8	80.0	197	279	1.7 %	7.0 %
EL	76.9	17.5	200	300	0 %	0 %

V. CONCLUSION

Simulation results have shown that the EL scheduler behaves similar to the PF scheduler when applied to NRT services. Additionally, it turned out that the PF scheduler is not suitable for RT services. The delay requirement of RT users is not taken into consideration by the PF scheduler. Consequently, there is a severe packet loss when the PF algorithm is employed for RT services. In contrast, the EL scheduler calculates the user priority not only based on the PF metric, but also the delay bound. Therefore, it is able to meet the different QoS requirements of RT users.

Considering a mixture of RT and NRT services, there is a trade-off between the throughput of NRT users and the delay requirement of RT users. The PF scheduler outperforms the EL scheduler with a higher aggregate throughput. However, it can not guarantee the delay requirement of RT users. The EL scheduler provides a relatively low cell throughput, but it meets the delay requirement of RT users. Hence, the EL scheduler is a better option for supporting the mixed services.

REFERENCES

[1] 3GPP TS 25.214, "Physical layer procedures (FDD)," 3rd Generation Partnership Project, Technical Specification Group Radio Access Network, Technical Specification V6.9.0, 2006.

[2] 3GPP TS 25.321, "Medium Access Control (MAC) protocol specification," 3rd Generation Partnership Project, Technical Specification Group Radio Access Network, Technical Specification V6.8.0, 2006.

[3] F. Brouwer et al., "Usage of link-level performance indicators for HSDPA network-level simulations in E-UMTS," in *2004 IEEE Eighth International Symposium on Spread-Spectrum Techniques and Applications*, Sydney, Australia, September 2004, pp. 844–848.

[4] 3GPP TS 23.107, "Quality of Service (QoS) concept and architecture," 3rd Generation Partnership Project, Technical Specification Group Radio Access Network, Technical Specification V6.4.0, 2006.

[5] 3GPP TR 25.855, "High speed downlink packet access; overall utran description," 3rd Generation Partnership Project, Technical Specification Group Radio Access Network, Technical Report V5.0.0, 2001.

[6] 3GPP TS 25.433, "UTRAN Iub interface Node B Application Part (NBAP) signalling," 3rd Generation Partnership Project, Technical Specification Group Radio Access Network, Technical Specification V6.9.0, 2006.

[7] J.M. Holtzman, "CDMA forward link waterfilling power control," in *Proceedings of IEEE Vehicular Technology Conference*, vol. 3, September 2000, pp. 1663–1667.

[8] ——, "Asymptotic analysis of proportional fair algorithm," in *Proceedings of IEEE Personal Indoor Mobile Radio Communications*, 2001, pp. F33–F37.

[9] R.C. Elliot et al., "Scheduling algorithms for the CDMA2000 packet data evolution," in *Proceedings of IEEE Vehicular Technology Conference*, vol. 1, Vancouver, Canada, September 2002, pp. 304–310.

[10] A. Jalali et al., "Data throughput of CDMA-HDR a high efficiency-high data rate personal communication wireless system," in *Proceedings of IEEE Vehicular Technology Conference*, vol. 3, 2000, pp. 1854–1858.

[11] M. Andrews et al., "Providing quality of service over a shared wireless link," *IEEE Communications Magazine*, vol. 39, pp. 150–154, February 2001.

[12] A. L. Stolyar and K. Ramanan, "Largest weighted delay first scheduling: large deviations and optimality," *Annals of Applied Probability*, 2001.

[13] A. Gougousis and M. Paterakis, "Scheduling with QoS support for multirate wireless systems with variable channel conditions," 2005, technical University of Crete, Greece.