Advanced Scheduling and Admission Control Techniques for Cellular Packet Radio Networks

Peter Stuckmann, Oliver Möller

Chair of Communication Networks, Aachen University
Kopernikusstr. 16, D–52074 Aachen, Germany
E-mail: pst@comnets.rwth-aachen.de

Abstract

In this paper we analyze the performance of scheduling algorithms applied in the Radio Link Control/Medium Access Control (RLC/MAC) layer of the (E)GPRS Base Station Subsystem (BSS). At first, scheduling algorithms are presented for an EGPRS best-effort service. While today's GPRS scheduler designs are based on the logical split between RLC and MAC, we propose to use information from the RLC layer for radio block scheduling on MAC level. Additionally we propose the usage of information on the actual link quality for adaptive scheduling. The performance of the proposed algorithms is compared to that of basic implementations typical for today's GPRS networks. In the next step we introduce scheduling algorithms for the support of different traffic classes. Both priority scheduling and fair scheduling approaches are extended by adaptive scheduling concepts. The interworking with admission control in the Serving GPRS Support Node (SGSN) is discussed and proposals for parameterization of the weights defined in the MAC scheduler and the maximum number of admitted flows for each traffic class in the admission control entity are presented.

1 Introduction

Operational cellular packet radio networks based on *General Packet Radio Service* (GPRS) and *Enhanced GPRS* (EGPRS) only include very basic MAC scheduling algorithms based on simple round robin and only realize best-effort services without *quality of service* (QoS) support for different applications and subscribers with their specific QoS requirements. Admission control to avoid overload situations is also not realized. In the last few years, algorithms were proposed for adaptive best-effort scheduling in wireless networks making use of the information on the link quality for different flows [3, 5]. Protocol aspects, e.g., the interworking of RLC and MAC protocols have not been addressed yet. For traffic class scheduling several algorithms have been proposed. They can be subdivided into priority schemes and bandwidth sharing schemes. For bandwidth sharing schemes several concepts are available designed for ATM switches and IP routers such as *Weighted Fair Queuing* (WFQ), *Weighted Round Robin* (WRR) and *Deficit Weighted Round Robin* (DWRR). Hybrid scheduling approaches have been developed to support real-time applications together with background traffic composed of several traffic classes [6]. Older scheduling algorithms for queuing systems based on the job duration or queue length of each connection, e.g., *Shortest Jobs First* (SJF) currently cannot be applied effectively for Internet traffic, since the TCP flow control hides the information how much data has to be transmitted for each session from the data link layer.

The contribution of this paper is the application and extension of these concepts for best-effort scheduling and traffic class support to be deployed in the base station RLC/MAC layer of cellular packet radio networks. We evaluate the performance of the proposed integrated scheduling schemes based on the example of the EGPRS standard. These schemes can be implemented in GPRS and EGPRS networks without changing the standard and can also be employed in other cellular packet radio networks.

For each proposed scheduling algorithm we present simulation results for EGPRS that are compared to performance results typical for algorithms that are presently implemented in operational GPRS networks. They are gained with the simulation tool GPRSIM that in fact is an emulator for GPRS and EGPRS. It comprises load generators for typical GPRS usage and a prototypical implementation of the GPRS protocols [8, 10]. These results can be regarded as representative because results of the GPRSIM have been validated by traffic performance measurements in operational GPRS networks [4].

2 QoS Architecture of Cellular Packet Networks

To define a QoS contract between the *Mobile Station* (MS) and the network, *Packet Data Protocol* (PDP) contexts containing QoS profiles are negotiated between the MS and the *Serving GPRS Sup-*

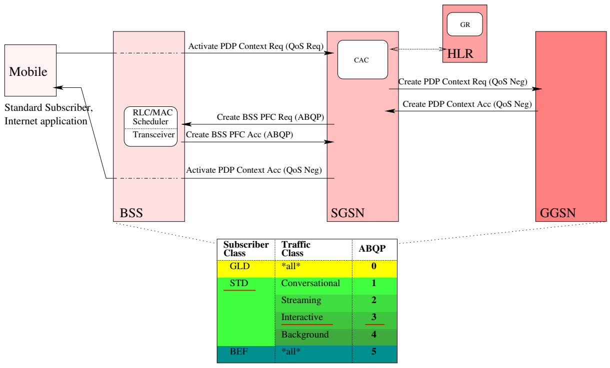

Figure 1: QoS negotiation

port Node (SGSN) [2]. In ETSI Release 99, the *Base Station Subsystem* (BSS) is provided with a *Packet Flow Context* (PFC) containing an *Aggregate BSS QoS Profile* (ABQP) (see Figure 1) and is responsible for resource allocation on a *Temporary Block Flow* (TBF) base and scheduling of packet data traffic with respect to the relevant QoS profiles negotiated. Moreover, it regularly informs the SGSN about the current load conditions in the radio cell. The tasks of the *Gateway GPRS Support Node* (GGSN) comprise mapping of PDP addresses as well as classification of incoming traffic from external networks based on downlink *Traffic Flow Templates* (TFTs). The *GPRS Register* (GR) holds the QoS-related subscriber information and delivers it on demand to the SGSN [7, 8].

From a time-scale point of view, the mechanisms for QoS management in GPRS can be regarded as a three-stage model. On PDP context activation the QoS parameters are negotiated. As long as the PDP context remains active, these parameters should be guaranteed unless there is a QoS renegotiation. The QoS profile is considered both for each TBF and for each radio block period. At TBF setup, radio resources like a set of *Packet Data Channels* (PDCHs) usable for this TBF are assigned according to the negotiated QoS parameters. During the TBF, radio blocks are scheduled at the BSS in competition with other existing TBFs in the radio cell. This scheduling function has to be performed considering the QoS profiles of the PDP contexts associated with the TBFs.

2.1 Scheduling

RLC/MAC scheduling in the (E)GPRS BSS can be subdivided into three steps: the selection of the traffic class, scheduling of the next TBF inside the selected traffic class and scheduling of the next RLC block of the selected TBF (see Figure 2.1).

Traffic Class Scheduling The MAC scheduler classifies the incoming radio resource requests of established TBFs regarding the application and subscription of the MSs. For example, the TBF can be classified in one of three subscriber classes, *Gold* service, *Standard* service and *Best-effort* service. In case the TBF belongs to a *Standard* subscriber it is additionally classified according to the application QoS profile to one of the four standard traffic classes, *Conversational*, *Streaming*, *Interactive*, or *Background*. Within the resulting six traffic class queues which are also called *TBF queues* only the identifiers of the TBFs are registered. The traffic class scheduler only has the information about TBFs which are requesting a data transfer and does not have information on the amount of data to be transmitted for each TBF. The traffic class scheduler selects a traffic class queue to be served by applying a class scheduling algorithm. This algorithm is not specified in the standard and can be optimized by the system designer. The algorithm can be, e.g. a priority algorithm or a bandwidth sharing algorithm.

TBF Scheduling Once a traffic class queue containing all TBF identifiers of this traffic class has been selected by the traffic class scheduler, the TBF scheduler selects one TBF of this TBF queue applying the TBF scheduling algorithm. This algorithm is also implementation-specific. As an example a *round robin* (RR) algorithm can be applied. The TBF scheduler only has the information that a TBF is established and has neither information on the amount of data to transmit nor if the TBF actually has data available. So the scheduler starts with the first TBF listed in the queue and checks if it has been allocated to the regarded PDCH. If not, the scheduler continues with the following TBF. In case the TBF is able to use the regarded *Packet Data Channel* (PDCH) the related RLC entity is polled for data until it reaches the predefined RR quantum or there are no more radio blocks to transmit. Then the following TBF of the same class queue is served if the same traffic class is still selected by the traffic class scheduler. Typically the RR quantum is in the order of 1-20 radio blocks.

RLC Block Scheduler In the third step, the RLC entity which has been polled for data by the MAC scheduler, checks if there are any data blocks available in the transmit buffer.

In case of *RLC acknowledged mode* the elements in V(B) indicate the acknowledgement status of related RLC data blocks. There are three possible states for each RLC data block:

- `NACK` indicates an RLC block which has not been transmitted yet, which has been negatively acknowledged or which has an expired timer
- `PENDING_ACK` indicates an RLC block which has been sent, but no acknowledgement has been received for this block yet
- `ACK` indicates data which has been sent and has already been acknowledged

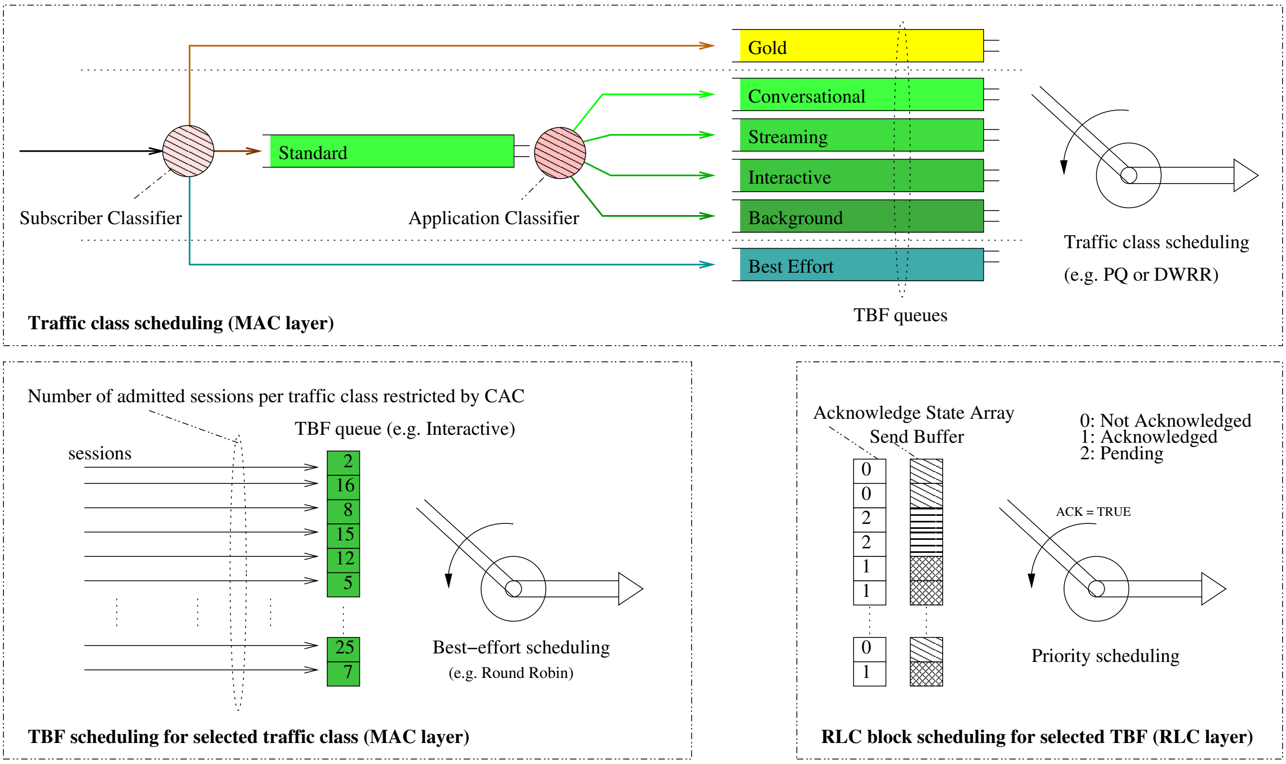

Figure 2: Scheduling of traffic classes, TBFs and RLC blocks

The RLC block scheduling algorithm determines the order of transmission of the RLC blocks inside the RLC send buffer of a regarded TBF. The RLC data blocks in the RLC transmit window with the acknowledge state `NACK` are forwarded to the MAC starting with the oldest one. If no `NACK` data block exists, the oldest RLC data block with the acknowledge state `PENDING_ACK` is retransmitted.

The priority of `NACK` blocks to `PENDING_ACK` blocks inside one RLC entity is specified in the standard [1]. It is also specified that `PENDING_ACK` blocks should be transmitted if a radio block period is scheduled for the regarded TBF and if no `NACK` block exists for this TBF. A decoupled implementation of RLC and MAC leads to the transmission of `PENDING_ACK` blocks, while other TBFs still could have `NACK` blocks to transmit that are more urgent. This gives the motivation to implement an RLC/MAC layer with a MAC TBF scheduler that serves TBFs with `NACK` RLC blocks ahead of TBFs with only `PENDING_ACK` blocks, which is consistent to the GPRS standard.

2.2 Admission Control

Connection Admission Control (CAC) in GPRS networks is part of the session management functions and is performed during the PDP context activation phase. It can be based on the number of available radio resources, the number of admitted sessions per traffic class and on the load situation in the regarded radio cell. The load situation can be characterized using load monitoring or load prediction based on the active sessions and the session to admit. Since measures for the current load situation of each radio cell will not be available at SGSNs in the short term, admission control will be based on the number of active sessions in the regarded radio cell and the available radio resources, namely the number of PDCHs in the cell.

A CAC policy has to be defined in a way that all active traffic flows can be served according to the QoS profiles negotiated. Preferably, there is only a limited number of privileged connections allowed simultaneously.

Furthermore, any *Standard* traffic should receive the resources necessary to meet its QoS requirements. Thus, it might be preferable to rather reject a PDP context activation request than to endanger the quality of all sessions. On the other hand, it might be advantageous to displace background or even interactive traffic flows to allow for an additional *Conversational* traffic flow to be admitted.

In our proposed CAC implementation, PDP context activation requests are differentiated according to their subscribed QoS profile and their traffic class. If additional PDP contexts are allowed for the requested QoS profile, the CAC decides on the basis of the radio resources available to admit the session or not.

A possible implementation for the traffic and subscriber classes Gold, Interactive and Background is shown in Figure 3.

To avoid a total withdrawal of resources from the *Standard* traffic classes with lower QoS requirements,

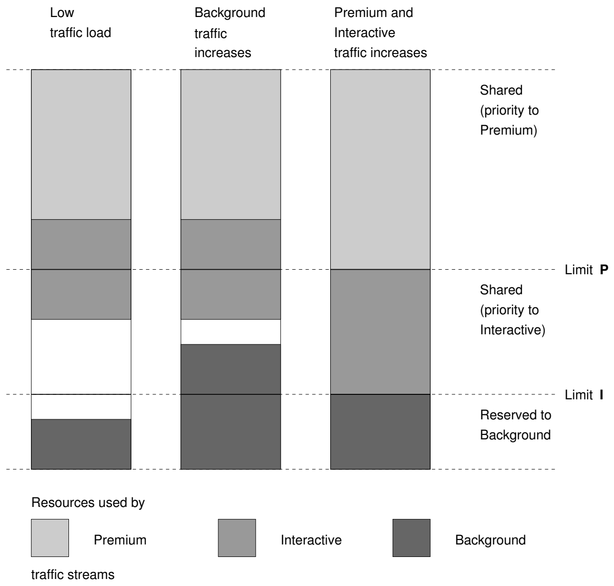

Figure 3: Admission control policy (example)

e.g., other than *Streaming*, there is a share reserved for Interactive traffic from the pool of radio resources in the cell. In times of high load, traffic flows with more demanding QoS requirements are allowed to displace flows belonging to applications with lower QoS requirements, but only up to a certain limit. The limits are specified by the maximum allowed number of active sessions for the regarded traffic class. When this limit is reached, the requested QoS is not accepted, but degraded to the next-lower-prioritized class.

3 Advanced Scheduling Algorithms for Best-effort Services

To support best-effort services all TBFs served by one base station are stored in one TBF queue. They are usually served by a RR strategy (see Section 2.1) to ensure a fair service for all active sessions. In this paper we propose two extensions of the RR scheduling strategy:

- the use of acknowledge state information of RLC blocks in the RLC layer (see Section 2.1) for TBF scheduling (*Displaced Pending Acknowledge Round Robin* (DPARR))
- the use of information on the actual link quality for adaptive TBF scheduling (*Link Quality-based Deficit Weighted Round Robin* (LQDWRR))

3.1 Displaced Pending Acknowledge Round Robin (DPARR)

In this scheduling scheme we schedule TBFs that have RLC blocks to transmit with the acknowledge state `NACK` always ahead of TBFs that have only RLC blocks with acknowledge state `PENDING_ACK`. With this modification it is avoided that `PENDING_ACK` RLC blocks are transmitted even if other TBFs have `NACK` blocks to transmit that are more urgent. With this feature both the system throughput and the user throughput performance can be significantly increased.

The mechanism is consistent with the GPRS standard, since RLC and MAC need not necessarily be implemented decoupled and the MAC scheduling of TBFs for RLC data block transmission is not explicitly specified in the standard.

3.2 Link Quality-based Deficit Weighted Round Robin (LQDWRR)

LQDWRR is based on the scheduling algorithm *Deficit Weighted Round Robin* (DWRR) proposed in [6]. The original aim was to accurately assign capacity to class queues and to provide nearly perfect fairness in terms of throughput. Fairness is guaranteed even if the packet flows contain data packets of different length. This algorithm is extended by the preferred service of TBFs with good channel quality. Fair scheduling of backlogged TBFs is ensured by the deficit counter that adapts the service share dynamically.

The link quality is represented by the optimal *Modulation and Coding Scheme* (MCS), which is reported by the EGPRS *Link Quality Control* [9]. In case the MCS is below a predefined value MCS_{LIMIT}, the quality of the link is classified as "*bad*", otherwise it is classified as "*good*". We introduce a *scheduling state variable* $SV_i(n-1)$ that represents one of three possible states:

Normal (**N**): The considered TBF_i was able to send data during the last scheduling cycle and the radio link quality was *good*:
$MCS_i[n-1] > MCS_{LIMIT}$

Backlogged (**B**): The regarded TBF_i was not able to send data within the previous scheduling cycle and the reported optimal MCS was below the limit:
$MCS_i[n-1] < MCS_{LIMIT}$

Lagged (**L**): The according TBF_i was backlogged for the maximum duration and has a service lag of Q_{max} and therefore was allowed to send data within the previous scheduling cycle, although the reported optimal MCS was below the limit:
$MCS_i[n-1] < MCS_{LIMIT}$

The deficit counter indicates the number of blocks to be transmitted in the scheduling cycle. It is increased by the RR quantum after each scheduling cycle and is decreased by 1 for each transmitted radio block. In this way backlogged TBFs keep the RR quantum for each scheduling cycle for the next one.

3.3 Performance Analysis

In this section the performance of the proposed best-effort scheduling algorithms compared to the basic

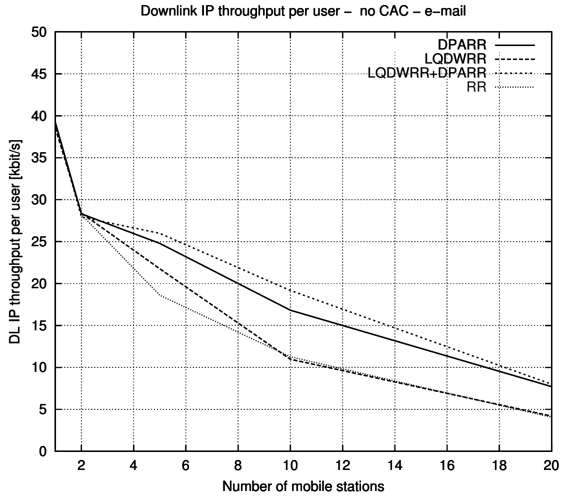

Figure 4: E-mail throughput performance for different best-effort scheduling algorithms

RR algorithm is evaluated. The scenario is characterized by an EGPRS radio cell with four fixed PDCHs, cluster size 3 and cell radius of 300 m and a traffic mix of 10% Streaming, 62% e-mail and 28% WWW sessions [7, 8].

Figure 4 shows the throughput performance of e-mail traffic for the discussed scheduling algorithms over the number of active stations per cell. While the performance in situations with low traffic load is similar for all four scheduling algorithms, DPARR achieves a performance gain of up to 50% compared to RR because no capacity is wasted in form of preferred `PENDING_ACK` RLC blocks. For WWW and Video Streaming traffic a similar gain is achieved.

The consideration of the link quality in the algorithm LQDWRR with an MCS_{LIMIT} of MCS-4 only increases the performance by maximum 10% in the regarded scenario. The reason is that in EGPRS often high MCSs are chosen even in situations with lower C/I values and the throughput is still acceptable because of incremental redundancy [9]. Additionally TCP timeouts occur for backlogged stations, which limits the performance gain achievable with LQ-based scheduling.

4 QoS Support for Traffic Classes

For traffic class scheduling we compare a priority scheduling algorithm with bandwidth sharing. In the priority algorithm a traffic class queue is only served if all queues of higher priority are empty. This can lead to poor performance and high session blocking rates for low priority classes, if the traffic load for higher priority classes is too high. To guarantee a certain minimum capacity for all classes the DWRR algorithm can be used [6]. In our scenario we have assigned the traffic class weights to 58% for Streaming (Video Streaming), 35% for Interactive (WWW) and 7% for Background (e-mail), which is related to the predicted offered traffic per traffic class and the QoS that is desired to be achieved for the classes. For TBF scheduling inside the traffic class queues the algorithms LQDWRR and DPARR are applied, since the combination has shown the best performance for best-effort scheduling in Section 3.3. For CAC at the SGSN we consider two strategies: *Soft CAC*, where a larger number of sessions with higher priorities are admitted, and a *Hard CAC*, where only a smaller number of sessions with high priorities are admitted to be able to guarantee the QoS for all traffic classes. The maximum number of allowed sessions per available PDCH are shown in Table 1.

Table 1: CAC Parameterization

Traffic class	Hard CAC	Soft CAC
Streaming sessions	1 per PDCH	2 per PDCH
Interactive sessions	2 per PDCH	3 per PDCH

Figure 5 shows the user throughput performance for Video Streaming over the number of mobile stations for the same scenario as of Section 3.3. The performance for priority queuing with soft CAC (see Table 1) achieves the best performance for Streaming services. While the throughput performance remains also acceptable for Interactive and Background traffic, the session blocking rate increases dramatically for Interactive (see Figure 6) and Background sessions. Even if the hard CAC algorithm is used blocking can not be avoided, since sessions that are not admitted for the requested QoS are degraded to lower traffic classes. However, the blocking rate can be reduced by up to 20% as we have seen in sample simulations, since traffic classes with lower priority are served more often. Session blocking is one reason for the decrease in system throughput, when traffic class scheduling is applied (see Figure 7). When the bandwidth sharing scheme DWRR is used the blocking rate can be decreased and the system through-

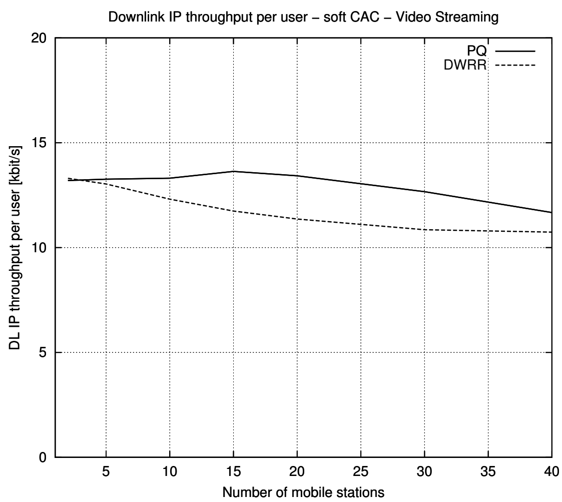

Figure 5: Streaming throughput performance for different traffic class scheduling algorithms

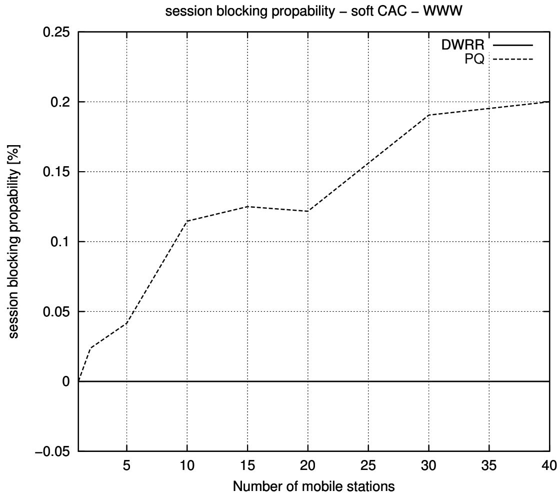

Figure 6: WWW session blocking probability for different traffic class scheduling algorithms

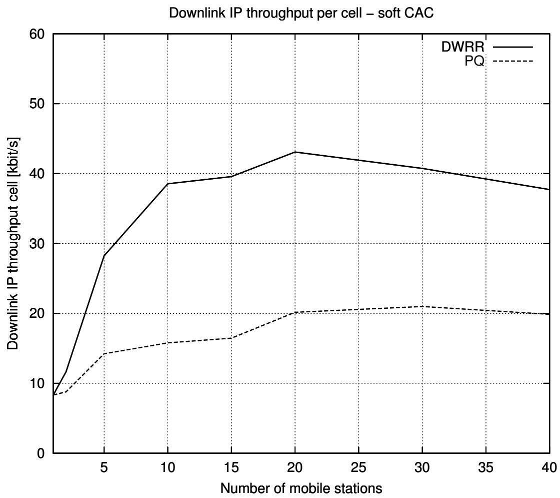

Figure 7: Downlink system throughput per cell for different traffic class scheduling algorithms

put can be increased by 100% compared to priority scheduling (see Figure 6 and Figure 7). Furthermore a hard CAC policy can increase the system throughput since blocking for traffic classes with lower priorities is reduced.

5 Conclusions

For a best-effort scheduler design we propose DPARR, since it is simple to implement and has a great effect on the throughput performance especially in situations with high traffic load. LQDWRR can be implemented in addition to realize an optimized scheduler that has no great effect in normal conditions where all stations have sufficient coverage, but can avoid the waste of capacity by serving stations with very bad channels. For the introduction of quality of service support in EGPRS several scheduling and admission control strategies are introduced. If operators want to maximize the performance of prioritized subscribers or applications and session blocking for lower traffic classes is acceptable, a priority scheduling algorithm should be applied together with hard CAC. If a certain capacity should be guaranteed also for lower traffic classes, a bandwidth sharing scheme should be implemented. The CAC strategy should then be chosen depending on the guarantee the operator wants to offer for prioritized users and on the offered traffic predicted for each service class.

References

[1] 3GPP TSG GERAN. GPRS; MS – BSS interface; Radio Link Control / Medium Access Control (RLC/MAC) protocol (3GPP TS 44.060 version 5.2.0 Release 5). Technical Specification, 3rd Generation Partnership Project, July 2002.

[2] 3GPP TSG SSA. GPRS; Service description; Stage 2 (3GPP TS 23.060 version 5.0.0 Release 5). Technical specification, 3rd Generation Partnership Project, January 2002.

[3] M. R. Jeong et altera. "Fair Scheduling Algorithm for Wireless Packet Networks", in Proceedings of the International Workshops on Parallel Processing, September 1999, pp. 280–285.

[4] P. Stuckmann et altera. "GPRS Traffic Performance Measurements", in Proc. of the IEEE Vehicular Technology Conference (VTC '02 fall), September 2002.

[5] T. Yoshimura et altera. "Wireless Packet Scheduling for Adaptive Service over IEEE 802.11", in Proceedings of the The First International Symposium on Wireless Personal Multimedia Communications (WPMC '98), Yokosuka, November 1998, pp. 314–319.

[6] M. Shreedhar and G. Varghese. "Efficient Fair Queueing Using Deficit Round Robin", in Proceedings of ACM SIGCOMM '95, August 1995, pp. 636–646.

[7] P. Stuckmann. "Quality of Service Management in GPRS-based Radio Access Networks" Telecommunication Systems 19:3, Kluwer Academic Publishers, March 2002, pp. 515–546.

[8] P. Stuckmann. "The GSM Evolution - Mobile Packet Data Services" John Wiley & Sons, Chichester, 2002.

[9] P. Stuckmann and J. Franke. "Performance Characteristics of the Enhanced General Packet Radio Service for the Mobile Internet Access", in Proc. of the Second International Conference on 3G Mobile Communication Technologies, London, UK, March 2001, pp. 636–646.

[10] B. Walke. "Mobile Radio Networks — Networking, Protocols and Traffic Performance" 2nd Edition, John Wiley & Sons, Chichester, 2001.