Wireless Scheduling Approaches and Practical Implementation Issues

Christoph Stepping

Institute for Data Communications Systems, University of Siegen, Hoelderlinstr. 3, 0-57068 Siegen, Germany, Tel.: +49 (271) 740-2623, Fax: +49 (271) 740-2536

EMail: stepping @nue. et-inf. mi-siegen.de

Abstract: In the recent years, well-known scheduling algorithms like Weighted Fair Queueing, Round Robin and their extended and improved variations have been developed in relation to fixed wired networks. Based on the great differences between the wired and the wireless channel, wireless scheduling algorithms have been developed to reach the aims of fairness, guaranteed throughput and delay in terms of QoS as well as a minimized loss of PDUs over a wireless channel. The general ideas of several algorithms are presented and discussed with respect to their complexity and their implementability. Related to the practical implementation of these algorithms, additional assumptions and extensions have to be taken into account, for example the retrieval of the buffer states of remote queues located at the mobile stations or the influence of non-transparent transmissions concerning the required Quality of Service (QoS). Another point is the simultaneous support of different transmission techniques like ATM or IP with the same scheduling algorithms which is also discussed. The most important issues and ideas will be addressed in brief to give a complete overview over the main concerns which must be considered with regard to the practical implementation of effective and fair scheduling algorithms in the wireless domain.

1. INTRODUCTION

Concerning wired networks, a complete set of scheduling and allocation algorithms has been developed in the recent years. Key assumptions in the wired domain are the permanent availability of the transmission channel

which possesses a constant capacity and which can be accessed by every data flow at any time. Main features in the wired domain are the fairness among different data flows, the minimization of the resulting delay and the maximization of the system throughput. The most important algorithms concerning this issue are Weighted Fair Queueing (WFQ) with its extensions WF2Q and WF2Q+ as well as Round Robin and Earliest Due Date approaches with their according variations. An additional general distinction of allocation principles in the wired domain is made in relation to the work-conserving character of the schemes.

In contrast to the fixed wired domain, the wireless channel is characterized by its time-variant capacity based on location-dependent and bursty errors. Furthermore, each mobile station (MS) experiences its own time-variant propagation conditions. While one MS cannot transmit any data because of its channel conditions, other MS may have no erroneous channel. An effective bandwidth allocation based on a base station (BS) located central scheduler can perform its task in a better way if a detailed knowledge of the current channel situation for each MS is available. This leads to the need of channel monitoring and channel estimation.

Additionally, the spatial distribution of traffic sources and intermediate buffers directed in the uplink (UL) direction at the MS lead to additional difficulties. Due to these various reasons and differences, simply transferring the algorithms from the wired to the wireless domain will fail.

The following sections are structured as follows: At first, a brief description of the most important wireless scheduling algorithms is given (section 2). Then, the common properties of the algorithms are discussed in section 3. After that, additional features and some new ideas are introduced: An approach for a unified description of transmitting traffic over ATM and IP systems will be presented in section 4. Issues concerning the practical implementation and the therefore needed additional mechanisms are discussed in section 5 on a protocol layer basis, mainly concerning the MAC layer and the LLC layer. Finally, the paper is summarized and some conclusions will be drawn.

2. WIRELESS SCHEDULING

In this section, the main allocation schemes for wireless data transmission are briefly summarized to give an impression of the way they execute their task of allocating the available bandwidth in a fair and flexible manner. Starting at an idealized fluid model concerning a single bit as minimum information unit, the allocation schemes are adapted to packet based transmission and some extensions concerning the determination of

their current channel and buffer status. [1] gives an overview over most of these schemes.

The basic principles and naming conventions are explained in connection with the first following scheme.

2.1 IWFQ

According to the development in the wired domain, the main algorithm Idealized Wireless Fair Queueing (IWFQ) [2] is based on a fluid model which is executed in a bit-wise manner. Its origin is the Weighted Fair Queueing (WFQ) [3] approach which was first considered for wired networks.

Due to possible errors on the wireless channel, a certain additional delay can occur. In the order of maintaining the given guarantees, a flow tries to reach the promised service (e.g. the amount of data a flow is allowed to transmit) by determining the current channel state and delaying the transmission until the channel state switches back to good propagation properties. Therefore, flows with good channel conditions can transmit a larger amount of data as they have been guaranteed during connection setup. They are called “leading flows”. Flows with bad channel conditions receive less service as they have been confirmed (“lagging flows”), therefore they try to compensate their lag by taking capacity from leading flows. At last, flows which receive exactly the service they expect are called “in sync” flows.

The classification of a flow into one of these categories is based on the comparison with the according error-free service model. In this case, the error-free service is simulated by a WFQ system. Furthermore, to execute this comparison, each packet receives a start tag S which symbolizes the point of time at which its transmission starts and a finish tag F which stands for the point of time a packet has completely finished its service in the current scheduling entity. In the general case, variable packet lengths are allowed and therefore the temporal difference between both tags is variable.

The range of the allowed compensation is bounded in the IWFQ approach by an upper bound for the lead and the lag. Considering the compensation of lag, the flow with the greatest lag is selected for transmitting the next packet according to the finish times of its packets. A lagging flow can transmit as long as his own channel is clean or until the finish time of one of its packets is higher than the finish time of the head-of-line packet of another flow. Therefore, a flow with a great lag can capture the complete transmission capacity and cause starvation for other flows with a smaller lag. In sync or leading flows will experience starvation, too, which is very disadvantageous.

2.2 WPS

Wireless Packet Scheduling (WPS) [2] tries to approximate the general issues addressed by IWFQ with concern to packet based transmission. The error-free service model is based on Weighted Round Robin (WRR) in contrast to the more complex WFQ as described in section 2.1. WPS assumes to work on a frame-based structure with time slots.

Additional mechanisms used for compensation in contrast to IWFQ are:

- Spreading: If a flow can use e.g. 3 slots for its transmission, the slots are not following one after another directly. Slots of other flows are interleaved between them.
- Swapping: According to the current error situation, slots of different flows can swap their slots. The goal is that every slot can be used because it is assumed that one of the MS will perceive a clean channel.
- Credit system: If swapping is not possible due to the slot position in the frame (e.g. the last slot is allocated and no swapping partner is available), additionally allocated or missing slots are summed up with means of credits and debits for the next frame. The background is to reach a long-term fairness.
- Prediction: For an effective swapping process, the different channel states must be known at the scheduling entity. Therefore, channel prediction mechanisms are introduced.

2.3 WFS

A more sophisticated approach is given by Wireless Fair Service (WFS) [4]. It tries to exclude in-sync flows from the compensation process between leading and lagging flows and it achieves a decoupling of bandwidth and delay by implementing slot queues and packet queues. Slot queues determine the amount of capacity which is allocated to a flow while packet queues are used to keep the transmission order of the packets.

The error-free service of WFS is realized by an extended WFQ version. Beside the rate weight r_i known from the standard WFQ, an additional delay weight Φ_i is introduced for computing the finish tag of the head-of-line packet (HOL packet).

To cope with error effects, swapping of slots is allowed within a certain amount of time specified by the look-ahead parameter ρ . Depending on the value of ρ , the error-free service model can change between WF2Q [3] for $\rho = 0$ (no swapping between HOL packets) and EDD [5] for $\rho \rightarrow \infty$.

Another important matter of WFS is the graceful degradation of leading flows. This means, the lead of a flow is decreased proportionally to its value. The other way round, lagging flows have to compensate their lag over a

longer period. Thus, the problem of starvation is solved by introducing proportional compensation.

2. CIF-Q

Channel-condition Independent Fair Queueing (CIF-Q) [6] represents also an approximation of WFQ. A reduction of complexity concerning the error-free service model is achieved by using Start-Time Fair Queueing (STFQ) [7]. This approach simplifies the computation of the virtual time which has to be executed after serving the HOL packet of a flow.

The compensation model of CIF-Q is based on two different assumptions:

1. Only the flow with the greatest lag is compensated.
2. Leading flows relinquish a slot with a certain probability α . If a slot is relinquished, it is allocated to the flow with the greatest lag.

α is a system parameter which influences the compensation behavior and leads to a proportional compensation for leading flows. CIF-Q is able to address most of the desired issues concerning fairness and delay for channel access.

2.5 SBFA

The Server Based Fairness Approach (SBFA) [8] is based on the principle that different flows should influence each other as less as possible. In contrast to the compensation model of most of the other mentioned allocation procedures, it is assumed that flows with an error-free channel should receive the guaranteed total amount of capacity. This has to be seen in contrast to the other already mentioned schemes in which in-sync or leading flows release a certain part of their capacity for compensation issues.

Therefore, SBFA reserves a determined part of the bandwidth for compensation. By allocating this excessive bandwidth to the flows which currently need additional capacity, the long-term fairness can be guaranteed. To reach this aim, dedicated compensation sessions are maintained in parallel to the original data flows. These sessions therefore are titled Long-Term Fairness Servers (LTFS). A distinction in packet queues PQ_i and slot queues SQ_i is made furthermore.

There is no scheme explicitly stated to use for error-free service, but the performance of SBFA depends on the choice of the scheme heavily.

2.6 CSDPS

Channel State Dependent Packet Scheduling (CSDPS) [9] can use any scheduling discipline designed for error-free transmission. As an example, the simple standard WRR scheme may be used.

If a flow should receive a slot, CSDPS checks its channel state. If the channel is error-free, the flow will receive its slot, if not, the flow is skipped in this round and the next flow will be served. CSDPS makes no use of measuring lead or lag and requires policed incoming traffic. No compensation model is foreseen.

Guarantees can only be given for the error-free case, so CSPDS is a rather simple scheme with rather simple guarantees.

3. DISCUSSION

After presenting the main issues of the main scheduling schemes for wireless purposes, the already reached properties should be discussed in brief in this section. The next section provides an overview of additional requirements which should also be taken into account for wireless scheduling schemes.

Each of the presented schemes except CSDPS requires the realization of an error-free model which has to be permanently maintained in parallel to the data flows itself. As far as the error-free model is based on WFQ, each HOL packet is assigned a start tag S and a finish tag F . Additionally, the virtual time $V(t)$ must be updated according to the service duration of the last transmitted packet. In sum, for each transmitted packet of a flow, at least three values have to be updated only for simulating the error-free situation.

The allocation process itself causes the computing of the current lead or lag situation of a flow after a packet has been transmitted and thus causes additional effort.

Compensating the lag of lagging flows requires to determine the flow with the maximum lag by means of sorting algorithms. This task has also to be executed after each packet to compensate the right flow in the next step. If a central scheduling entity executes the allocation task, as it is implemented in almost all wireless transmission systems, additional information has to be transmitted in UL direction to keep the scheduling entity informed about the present situation at the different MS. This additional effort is not negligible.

Considering the complete set of parameters which have to be updated after the transmission of a single packet within a radio cell, a tremendous complexity arises. Especially in connection with mobile broadband systems, the processing time for the execution of these steps is strongly limited because of the high achievable data rates. Therefore, some simplifications should be introduced while maintaining the general properties of the most elaborated wireless scheduling schemes.

Other topics which have not been addressed by the presented scheduling schemes are the following:

Whereas all of the schemes operate with a simple weight for each flow representing the average data rate, in real transmission systems a traffic flow cannot be specified by only one value. In general, traffic sources show a time-variable behavior which leads to time-variant data rates and therefore some burstiness is introduced. Even for services with originally constant data rates, transmission in non-transparent mode causes retransmissions which leads to a time-varying behavior [10].

None of the considered schemes explains how the weights are determined nor if they have to be interpreted in a static way (fixed weights determined at connection setup) or in a dynamic way (changing of weights during connection lifetime).

Furthermore, the issue of handling the effects of ARQ protocols are not mentioned in the proposed schemes. PDUs which are waiting for retransmission should receive a higher priority because the continuation of the flow's transmission strongly depends on their successful transmission. Considering compensation models, the different handling of retransmission data must also be taken into account (retagging of retransmitted packets, discarded packets).

Furthermore, a great set of transmission systems does not work on a pure packet basis but on a structured MAC frame. Therefore, the retrieved information used for the allocation process is valid for the whole MAC frame and not only considering the next packet to transmit.

Another difference which has to be taken into account concerns the meaning of the scheduling strategy: Whereas in a packet based scheduling approach the decision which HOL-packet of the different connections has to be transmitted next defines the number of packets to transmit implicitly (in this case one packet), frame based schemes have to deal with the capacity of a whole frame. So the matter in frame based approaches is not only which connections should transmit next but also how much capacity each of them should receive to maintain the confirmed amount of service.

Another important topic is the provision of fairness under temporary overload situations. A complete scheduling scheme should also deal with strategies to determine which connection has to give up how much of its service while maintaining the promised QoS in the highest possible degree.

Whereas the presented scheduling schemes deal with data transmission in a very general way, a more detailed relation to existing transmission systems like ATM or IP based systems may be useful. Especially the parameters and mechanisms introduced in these both systems should be considered.

A common scheduling architecture for ATM and IP systems would be very useful and can be achieved in a certain way as will be shown in the next section.

4. UNIFIED DESCRIPTION OF QOS FOR ATM AND IP TRAFFIC

Concerning the support of QoS, ATM systems and IP networks can be classified in a certain way by the same parameters. In the following paragraphs, the support of QoS in the Internet domain is related to the stringent requirements made by the architecture of “Integrated Services” (IntServ) [11], whereas the ATM support of QoS is described in [14].

In connection with the “ReSerVation Protocol” (RSVP) [12][13], a specification of traffic flows with regard to a so called TSpec is given. The relation of the ATM traffic parameters with the values defined by a TSpec is shown in Table 1.

Table 1. Comparison of ATM traffic parameters and TSpec parameters of IntServ

Meaning	ATM parameter	TSpec parameter
Mean data rate	SCR	r
Peak data rate	PCR	P
Maximum burst size	MBS	b

At first, the source description used in both systems can be unified. Basically, controlling and monitoring the amount of data which is injected to a network leads to a common view for both systems.

In the ATM area, the permanent cell rate is defined as sustainable cell rate (SCR) [14]. This parameter represents an upper bound of the average cell rate, built out of the total number of cells transmitted during the connection lifetime. A certain variability or burstiness can be allowed by

specifying values for the maximum burst size (MBS). The MBS represents the number of cells which can be transmitted with a rate higher than the SCR, but without violating the QoS for this connection which has been defined by a traffic contract. A third parameter defines the maximum allowable cell rate, the peak cell rate (PCR). Cells are not allowed to be sent with a temporal spacing smaller than $1/\text{PCR}$.

Concerning Internet traffic, a description of the source behavior is given by the already mentioned TSpec [13]. A TSpec comprises parameters which have the same meaning as the ATM parameters: a peak data rate r , a mean data rate b and a bucket size b , which is explained in the next paragraph. Since IP packets show different sizes, additional parameters are required in the TSpec as minimum policed unit m and maximum datagram size M .

4.1 Shaping and policing entities

Guarantees can only be fulfilled completely as long as the underlying network is prevented from overload situations. Therefore, some shaping and policing mechanisms have to be introduced (Figure 1). Due to the according source descriptions, a unified architecture can be used to monitor the given constraints regarding the bandwidth and burstiness of the injected traffic.

Figure 1. Shaping and policing entities

The burstiness is monitored by a token bucket which is based on a source generating so called tokens with a constant rate according to the SCR or the mean rate r . Tokens are inserted to a bucket of fixed size which represents the maximum burst size MBS respectively b . Each PDU which is inserted to the network decreases the token buffer state by one and removes a token from the bucket. If the token bucket is empty, additional PDUs are stored in

the intermediate buffer and experience a certain delay (shaping) until new tokens are generated.

To prevent the PDUs of a burst from being transmitted with the link rate, an additional policer guarantees the peak rate PCR respectively p in the Internet case. Finally, the same shaping and policing entities can be used to manage traffic from ATM as well as from Internet sources. The main difference between the application of both systems is the possibility of differing PDU sizes.

4.2 Delay and jitter issues

Besides the already mentioned parallels concerning bandwidth requirements, some parallels concerning the allowable delay can be considered, too.

[15] defines an upper border for the acceptable end-to-end-delay which consists of a fixed part caused by the physical transmission and a variable, queueing-dependent part. This definition can be related to the ATM area via the maxCTD parameter which represents the maximum period of time which is allowed to pass by between the injection of a PDU at the transmitter to the reception at the receiver.

Therefore, the region of the achievable delay is covered in both systems, too. One major difference is related to the jitter, which is specified as cell delay variation in the ATM case. So far, there is no parameter with the same meaning defined for the Internet domain.

5. IMPLEMENTATION ISSUES

In the following, a wireless scenario which simultaneously supports both already mentioned transmission systems over a common, shared link is presented as a basis for a typical wireless transmission system. The scenario is based on the HIPERLAN/2 standard which describes a short-range, broadband wireless LAN [16] using a TDMA/TDD-frame at MAC layer.

After introducing the general protocol structure, some additional points which should be taken into account concerning scheduling purposes and some new ideas regarding this area are proposed.

5.1 Overview: Protocol stack

HIPERLAN/2 (H/2) is designed to support different transmission systems in parallel. The adaptation to the applied system is achieved via a convergence layer which consists of a service specific and a common part.

Figure 2 shows the adaptation of cell-based ATM traffic and packet-based IP traffic towards the H/2 system.

Figure 2. Applied protocol stack

The common part of the convergence layer inserts uniform, equally sized PDUs into the link layer, independent of the originally used transmission system. Therefore, the control of the MAC layer executed by the scheduler can be simplified if a uniform description of the traffic and QoS parameters for ATM based traffic and IP based traffic is available.

As shown in Figure 2, the originally intended set of traffic and QoS parameters specified at the Service Access Point (SAP) of the network layer have to be adapted with respect to the radio based transmission. Due to the occurring errors on the radio channel, a modified set of parameters has to be taken into account for scheduling purposes (see section 5.4). For example, additional bandwidth may be required to execute the retransmission of formerly corrupted PDUs properly.

A possible approach of splitting the allocation process is shown in Figure 3.

Figure 3. Call admission and allocation process

At connection setup, which is always necessary for ATM connections as well as for IntServ flows, a certain amount of capacity can be allocated in addition to the capacity required for transparent and error-free transmission (long-term behavior). This is done by the call admission control (CAC) entity. The allocation of the available capacity within a MAC frame depends on the current error situation and represents the highly dynamic short-term behavior which is handled by the scheduling algorithm.

The shown model consisting of two stages assures the general availability of required resources and enables a very flexible and dynamical allocation process on the other hand. It should be explicitly mentioned that the CAC procedures make use of the complete set of parameters as shown in Table 1 and are not only based on a mean data rate like the previously discussed wireless scheduling schemes.

5.2 PHY: Channel estimation

As shown in the presented scheduling approaches, the exact knowledge of the current channel situation for each MS can increase the performance of the data transmission significantly. Normally, the schemes try to resolve the channel situation by executing measurements on PHY level.

As soon as hybrid ARQ schemes are involved, the correction process of the FEC component of these hybrid scheme can improve the link quality. Thus, the important matter for the control of the MAC frame is no longer the situation at the PHY level, but the situation at the LLC level.

To gather some information about the link quality at LLC level, the LLC entities could measure the number of corrupted PDUs which occur between correctly received ones. This means the LLC entity can estimate the average burst error length as well as the maximum error length. Of course the approach depends on the underlying traffic structure, but the higher the occurring data rates are, the more accurate the values are.

Especially in connection with constant data rate services (equally spaced PDU distances) transmitted under real-time conditions in transparent mode, assumptions about the current channel situation can be made. In this case,

the LLC entity is somehow misused as measurement instrument by validating the PDU sequence numbers without correcting corrupted PDUs by requesting retransmissions. However, the FEC part of the hybrid ARQ scheme should fulfill its task.

5.3 MAC: Information feedback for transparent connections

The spatial distribution of UL directed buffers among the different MS within a radio cell leads to the necessity that the central scheduler located at the BS must be permanently kept informed about the situation at each MS.

As shown in Figure 4, the scheduling entity has full knowledge about the buffer states for DL orientated connections.

Figure 4. Gathering feedback information for UL connections

In contrast to this, each PDU transmitted in UL direction carries information about the buffer state and/or the delay situation of the HOL PDU which are examined by the scheduler. These pieces of information are furtheron specified as feedback information.

Since PDUs can be corrupted or delayed by one MAC frame in UL direction, the scheduler must implement strategies to cope with the total loss of feedback information during error bursts as well as strategies to cope with the relative age of the feedback information he receives. For example, if a flow with a high data rate is considered, the feedback information transmitted with a PDU at the beginning of a MAC frame can describe a different buffer state as it is available at the end of the MAC frame when the

scheduling process considering the next MAC frame is executed. The possible deviations have also to be taken into account.

Another important matter is how often the feedback information is multiplexed into the UL PDU stream (Figure 5).

Figure 5. Effects of missing buffer state information

If the feedback information is always entered at the end of a PDU train, a loss would have severe effects on the accuracy of the scheduling process because the scheduler has to use the last available data resulting from one of the previous PDU trains. The other way round, the insertion of the according information causes additional overhead but leads to a higher accuracy.

An improvement could possibly be made by the estimation of the behavior of the traffic source by using the given source description and by learning from the already experienced traffic behavior.

5.3.1 MAC: Information feedback for non-transparent connections

Concerning non-transparent transmission, the scheduling entity has to take into account that positive (ACK) and/or negative (NACK) acknowledgements must be transmitted in the backward channel of a connection. Especially if unidirectional DL connections are considered, the amount of acknowledgements waiting for UL transmission cannot be transmitted via piggybacking because of a missing backward transmission. In this case, additional capacity must be allocated for this transmission. Ideally, in connection with variable slot sizes, PDUs carrying acknowledgements show a smaller size than PDUs used for normal user traffic.

Figure 6. Feedback information in the non-transparent case

In consequence, the scheduler has to recognize not only the current buffer state but also the current state of waiting acknowledgements.

5.4 LLC: Consideration of non-transparent transmission

While during transparent transmissions the packet order is kept in mind, the retransmission of PDUs due to the execution of Logical Link Control (LLC) protocols (ARQ protocols) causes a more complicated situation. At first, the complexity of data flows is increased by one order. This means, flows based on constant data rates are transformed into flows with variable data rates. Respectively, transparent services with variable data rates are changed into highly variable services if they are transmitted in non-transparent mode. In consequence, the Complexity of the parameter set describing the traffic increases tremendously through the exploitation of LLC protocols which makes it very difficult to handle these parameters for general scheduling strategies.

5.4.1 Protocols

Usual protocols which are intended to assure a reliable link are protocols from the Go Back N (GBN) protocol family and the Selective Repeat (SR) protocol family which are well-understood in their basic versions [17].

In connection with real-time transmissions, ARQ protocols have to be adapted and extended to these features:

1. The number of retransmissions must be constrained due to the existing upper bounds concerning the transmission delay. PDUs which arrive too late have no value for the application.
2. If the lifetime of a PDU is expired, it needs not to be transmitted anymore and should be discarded due to transmission continuity. The discard has to be signalled with the next PDU. This can be achieved by setting a single flag in the PDU header which informs the receiving LLC entity that the sequence number of this PDU should be interpreted as the next valid one. All sequence numbers with a smaller value are ignored furtheron.
3. Concerning SR protocols, the receiving LLC entity can decide whether to play out PDUs even if the sequence is not complete or whether a PDU sequence should be discarded as soon as a single PDU is missing. Depending on the application, eventually a valuable gain can be realized by playing out incomplete PDU sequences.

Furthermore, the reaction time of ARQ protocols should be increased: As soon as a missing or corrupted PDU is detected by receiving an uncorrupted following one, the appropriate acknowledgement should be sent out by the receiving LLC entity without taking the transmission window into account. This means, the receiving entity should not wait until the transmission window is completely filled before signalling the missing PDUs.

Another effect concerning the transmission window appears in connection with bursty sources: Since the receiving entity waits until its window is filled before sending acknowledgements, additional delay can be introduced if a burst does not fill the window completely and the next burst occurs after a certain idle period. Retransmissions should be executed during the idle period to reach the confirmed overall throughput.

Due to this fact, the receiving entity monitors the period of time after the last received PDU and acknowledges the received PDUs after a timer expired. This mechanism generates additional traffic to assure a real-time behaviour for non-transparent services and must also be considered by designing the scheduler strategies.

5.4.2 Extended parameter set

The introduction of ARQ protocols requires the extension of the given parameters as already mentioned. As an example, the relationship between

the parameters defined for the fixed, wired networks and the effects of the non-transparent transmission caused by ARQ protocols should be explained.

If a transparent flow is described by its parameters SCR, PCR and MBS (in ATM) respectively its equivalents r , p and b (in IP IntServ), the according description for the non-transparent case has to be extended: For example, a transparent CBR flow is defined by its PCR whereas the values of SCR and MBS are set to zero. With regard to possible retransmissions, the permanent data rate of this connection SCR^* must be set to PCR, whereas the new peak rate PCR^* is greater than PCR.

The same happens with regard to the MBS which is set to zero in the transparent case and is defined by the maximum amount of data the LLC protocol can retransmit within one burst in the non-transparent case. For example, the worst case in relation to the simple Go-Back-N protocol is given if the first PDU in a transmission window is corrupted and the receiver detects the error at the end of the window. In this case, the MBS for the non-transparent case has to be set to the size of the transmission window.

The extended parameters should be used as a basis for the execution of the allocation process concerning connections in non-transparent mode.

5.4.3 Extended parameters for real-time traffic

Concerning real-time traffic, a more complicated constellation occurs: In this case, a non-transparent flow with realtime properties requires LLC protocols which support the discard of already expired PDUs to prevent the transmission from being blocked permanently.

At this point, the defined transmission delay $maxCTD$ has to be extended to $maxCTD^*$ for defining a maximum number of allowed retransmissions before PDUs are discarded. Finally, the number of retransmissions also influences the value of PCR^* because the PDUs must be retransmitted before they expire and therefore influence the maximum achievable data rate.

Therefore, the parameter PCR^* depends on different parameters:

$$PCR^* = f(SCR^*, maxCTD^*, T_{MACFrame})$$

where $T_{MACFrame}$ represents the fixed length (2ms) of a MAC frame in the H/2 system and therefore symbolizes the available transmission capacity in relation to delay.

6. CONCLUSIONS

After introducing the most important approaches for scheduling in the wireless domain, the common properties of these schemes are discussed. Several additional features have to be taken into account if the schemes are implemented in real transmission systems.

A unified description of ATM traffic and IP traffic based on the IntServ architecture is proposed to simplify the simultaneous support of both transmission systems through scheduling schemes.

Some additional mechanisms and new ideas concerning the realization of effective wireless scheduling schemes are presented. They comprise additional tasks like error estimation on LLC level, the effects of frame-based scheduling approaches in contrast to packet-based ones, the retrieval of feedback information in UL direction as well as the effects of non-transparent transmission and the necessary parameter variations for this case.

REFERENCES

- [1] T. Nandagopal, S. Lu, V. Bhargavan:
"A Unified Architecture for the Design and Evaluation of Wireless Fair Queueing Algorithms", Wireless Networks, 1999
- [2] S. Lu, V. Bhargavan, R. Srikant:
"Fair Scheduling in Wireless Packet Networks", ACM SIGCOMM 97, August 1997
- [3] Jon C.R. Bennett, Hui Zhang:
„WF²Q: Worst-case Fair Weighted Fair Queueing“, Proceedings of IEEE INFOCOM '96, pages 120- 128, San Francisco, CA, March 1996
- [4] S. Lu, T. Nandagopal, V. Bhargavan:
"Design and Analysis of an Algorithm for Fair Service in Error-Prone Wireless Channels", Journal of Wireless Networks, February 1999
- [5] M. Vishnu, J.W. Mark:
"HOL-EDD: A flexible Service Scheduling Scheme for ATM", in Proceedings of IEEE INFOCOM, pages 647--654, March 1996
- [6] T. S. Eugene, I. Stoica, H. Zhang:
"Packet Fair Queueing Algorithms for Wireless Networks with Location-Dependent Errors", IEEE INFOCOM '98, March 98
- [7] P. Goyal, H. M. Vin, H. Cheng:
„Start-time Fair Queueing: A Scheduling Algorithm for Integrated Services Packet Switching Networks“, IEEE/ACM Trans. on Networking, Vol. 5, No. 5, October 1997
- [8] P. Ramanathan, P. Agrawal:
"Adapting Packet Fair Queueing Algorithms to Wireless Networks", Proceedings of MOBICOM '98
- [9] R. Bhagwat, P. Bhattacharya, A. Krishna, S. Tripathi:
"Enhancing throughput over wireless LANs using Channel State Dependent Packet Scheduling", IEEE INFOCOM '96, April 1996

- [10] C. Stepping:
“Bandwidth allocation for non-transparent CBR services”, Proceedings of EPMCC 2001, Vienna, February 2001
- [11] R. Braden, D. Clark, S. Shenker:
Request For Comment 1633: „Integrated Services in the Internet Architecture: an Overview”, IETF Juni 1994, available at the Internet, e.g. <ftp://ftp.isi.edu>
- [12] R. Braden, L. Zhang, S. Berson et. al.:
Request For Comment 2205: “Resource ReSerVation Protocol (RSVP)”, IETF Sept. 1997, available at the Internet, e.g. <ftp://ftp.isi.edu>
- [13] J. Wroclawski:
Request For Comment 2210: „The Use of RSVP with IETF Integrated Services“, IETF Sept. 1997, available at the Internet, e.g. <ftp://ftp.isi.edu>
- [14] ATM Forum: „Traffic Management Specification 4.1“, af-tm-0121.000. March 1999
- [15] S. Shenker, C. Partridge, R. Guerin:
Request For Comment 221 2: „Specification of Guraranteed Quality of Service“, September 1997 available at the Internet, e.g. <ftp://ftp.isi.edu>
- [16] ETSI Technical Report TR 101 683: „BRAN HIPERLAN Type 2; System Overview“, ETSI, 2000
- [17] Bertsekas, D.; Gallager, R.: „Data Networks“, 1992, Prentice-Hall Englewood Cliffs, NJ.