Hierarchical Scheduling Strategies for Wireless ATM MAC Protocols

Rolf Sigle, Thomas Renger#*

**University of Stuttgart, Institute of Communication Networks and Computer Engineering Pfaffenwaldring 47, D-70569 Stuttgart, sigle@ind.uni-stuttgart.de and Daimler-Benz AG, Communication Technology, D - 89013 Ulm, sigle@dbag.ulm.daimlerbenz.com*

#University of Stuttgart, Institute of Communication Networks and Computer Engineering, Pfaffenwaldring 47, D-70569 Stuttgart, renger@ind.uni-stuttgart.de

Abstract

Wireless ATM networks, which are currently under investigation, promise to support future multimedia applications. The medium access control (MAC) protocol of such systems plays a key role in providing differentiated quality of service (QoS). A variety of proposed MAC protocols use a centrally controlled flexible allocation of time slots. In order to provide a fair scheduling, the mobile stations must inform the base station about the state of their queues. This is done by sending so-called request packets. As these updates consume valuable bandwidth, their frequency should be kept as small as possible. The proposed concept of combining hierarchical scheduling with fair queueing strategies requires only a very little number of updates and is insensitive against partially outdated information.

1 Introduction

Because of the trends towards multimedia and mobile communications broadband wireless networks which support advanced multimedia applications are needed in the near future. While for the fixed network new developments like ATM promise to provide differentiated QoS nowadays wireless cellular networks like GSM or IS-54 and wireless LANs like HIPERLAN1 (HIgh PErformance Radio Local Area Network) or IEEE 802.11 are mainly single service networks which are not able to meet all these future demands. Therefore, ATM based wireless broadband networks are currently investigated by several companies like Olivetti [7] and NEC [13] among others as well as European R&D projects like the Magic WAND ACTS project [11] and wireless ATM prototype systems are build. Moreover, the wireless ATM working group of the ATM Forum and the recently formed ETSI BRAN (Broadband Radio Access Network) project (former ETSI RES10 working group) started with the specification [3, 12] of various extensions to ATM protocols which are needed to cope with the mobility of users and wireless access.

In order to provide differentiated QoS, different ATM service categories (e.g. CBR, rt-VBR, nrt-VBR, ABR, UBR) have been defined together with a framework of traffic management functions [2]. Moreover cell scheduling strategies [1] have been developed which are able to guarantee the network performance parameters [5] (e.g. cell loss ratio, cell transfer delay, and cell delay variation) specified in the traffic contract. In order to allow seamless access of wireless nodes to ATM networks, these networks must support the same service categories as defined for the fixed network. Hence, a medium access control protocol for wireless ATM, which is responsible for the resource management in the wireless network, must cooperate with the traffic management functions of the fixed ATM network and thereby constitute a robust and efficient overall control framework [14].

The remainder of this paper is organized as follows. In Section 2 we describe mechanisms and potential problems of wireless ATM MAC protocols. Then in Section 3 we outline the principle of fair queueing strategies and in Section 4 the concept of hierarchical scheduling in wireless ATM MAC protocols is introduced. In Section 5 we present first results from simulation studies and finally we conclude by giving an outline of future work.

2 Wireless ATM MAC Protocols

In order to control the allocation of radio resources in a fair and efficient manner, an appropriate wireless MAC protocol is needed. A variety of MAC protocols for wireless ATM have been proposed recently in the literature. Their common features and problems arising in these protocols are discussed in the following. Most proposals (cf. [7],[9],[11],[13], and [8]) are based on a reference model which introduces a wireless data link control (DLC) layer below the ATM layer (see Figure 1).

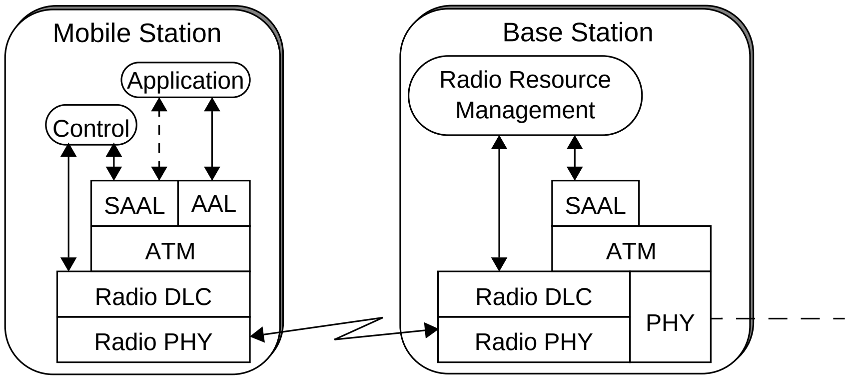

Figure 1: WATM Reference Model

The DLC layer is composed of the MAC sub-layer and logical link control (LLC) sublayer which deals with transmission errors. The up- and downlink are separated either by different time slots or frequency channels, i.e. either time division duplex (TDD) or frequency division duplex (FDD) is applied. Due to the higher available channel bandwidth in the TDD case a higher multiplexing gain can be achieved. Furthermore, in case of asymmetric traffic (e.g. higher bandwidth in downlink than in uplink direction) the FDD scheme cannot utilize the available bandwidth and therefore leads to a waste of resources. The major drawback of TDD is the radio turn-around time needed between any downlink and uplink time slot to switch from receiving to transmitting mode. Thus, the number of radio turn-arounds should be kept small.

The radio medium access is centrally controlled by the MAC instance of the base station and uses a TDMA scheme. In order to achieve high flexibility time slots are allocated on a slot by slot basis. The base station acts as an ATM multiplexer with distributed input queues located inside the mobile stations and in the base station itself (only in the TDD case). In contrast to concentrated ATM multiplexers the scheduler inside the base station has only imperfect and partially outdated information about the states of the input queues located in the mobile stations. In order to update this information, the mobile stations send control information to the base station either piggybacked on data slots, in special control slots or in both. The access policy on the uplink control slots is based on polling, random access, or a combination of both. An example of a TDMA/TDD frame structure is depicted in Figure 2.

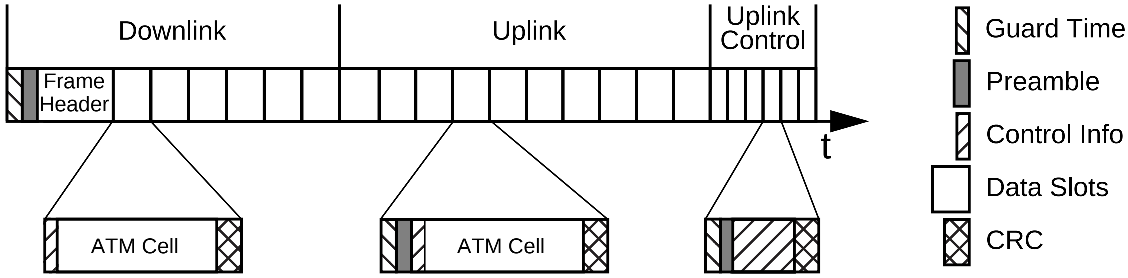

Figure 2: TDMA/TDD frame structure

The base station initiates a new frame by sending a frame header including information about the allocation of downlink and uplink slots. The number of downlink and uplink slots as well as the total frame length may be variable or fixed. Downlink data slots containing ATM cells, downlink control information (e.g. for ARQ purposes), and error control information, follow the frame header. Data slots allocated to the uplink direction are similar to downlink data slots. As uplink slots are sent by different stations, each packet needs an individual preamble to allow synchronization. Moreover, due to imperfect compensation of signal transfer delay a guard time is needed between each individual uplink packet (control and data packets). In order to reduce the overhead caused by guard times and preambles, some protocols introduce the concept of cell trains, i.e. if more than one data slot is allocated for one mobile station in a frame they may be concatenated to one uplink packet [11]. As mentioned before, mobile stations may send control information (i.e. queue state information, and ARQ information) in specific uplink control slots which are located either before or after the uplink data slots.

The different protocols are distinguished by the scheduling algorithm used, the exchanged queue state information, the access policy on uplink control slots, and the size of guard times and preambles which strongly depend on the physical layer. The ATM cell transfer delay and cell delay variation are heavily influenced by the amount and frequency of exchanged status information and the scheduling algorithm. Therefore, traffic conforming to the traffic contract at the mobile station may be altered to non-conforming traffic which may therefore cause cell discards by the usage parameter control (UPC). In order to deal with this problem the MAC algorithm may be adapted to the conformance definition of the UPC function, i.e. the generic cell rate algorithm (GCRA). Additional changes of the traffic profile, caused by ARQ protocols at the LLC layer, cannot be prevented by this approach and may therefore be handled by traffic shaping, e.g. at the output of the base station. Besides QoS guarantees the MAC protocol must support the isolation of different traffic flows, i.e. traffic flows conforming to the traffic contract must not be violated by other flows that try to send more than they are allowed to. Fair queueing strategies, developed for ATM fixed networks, support these properties and allow an efficient and fair sharing of the available resources. Therefore, we propose to adapt these strategies to wireless ATM MAC protocols.

3 Fair Queueing Buffer Scheduling Strategies

A variety of different fair queueing buffer scheduling strategies have been developed for ATM fixed networks. These scheduling mechanisms provide guarantees on end-to-end delay, cell loss, and fairness for flows conforming to the GCRA. They are based on the generalized processor sharing (GPS) which was originally used for scheduling in multiprocessor systems. GPS uses a fluid flow model where all active flows are served parallel according to their assigned service share.

Packet-by-packet generalized processor sharing (PGPS) described in [10] tries to adapt GPS to a packet network environment where only one packet is in service at any time. PGPS approximates GPS by scheduling packets based on their finishing times in the corresponding GPS system. A separate FIFO queue and a weight ϕ_i (service share) which is based on the traffic parameters, negotiated during connection set-up, is assigned to each flow. The scheduling algorithm shares the available bandwidth among all active connections (queue length greater than zero) according to their specified weights. Hence, independent of all other flows each flow is guaranteed at least its corresponding service share and spare bandwidth is fairly shared among all active connections. In order to determine the sequence of packets the notion of virtual time $\hat{v}$ is introduced and each packet *k* of flow *i* is assigned a service tag F_i^k (virtual finishing time). With $F_i^0 = 0$:

$$F_i^k = \frac{1}{\phi_i} L_i + max(F_i^{k-1}, \hat{v}(a_i^k)) \quad (1)$$

The virtual finishing time of a packet is determined by the length of the packet (L_i, constant in ATM networks) divided by the corresponding weight (ϕ_i), the virtual finishing time of the preceding packet, and the virtual time at the arrival of the *k*th packet, a_i^k (see Eq. (1)). The virtual time is calculated according to Eq. (2) where *C* is the link capacity and $B(t)$ is the set of currently active flows.

$$\frac{d}{dt}\hat{v}(t) = \frac{C}{\sum_{j \in B(t)} \phi_j} \quad (2)$$

The major drawback of this algorithm is its implementation complexity due to frequent calculations of the virtual time function which are necessary upon changes of the set of active flows. Therefore, self-clocked fair queueing (SCFQ) was introduced in [4]. It approximates the virtual time function by the service tag of the *l*th packet of flow *j* receiving service at arrival time of a new packet:

$$\hat{v}(t) = F_j^l \quad (3)$$

Note, if a new packet arrives at a queue with queue length greater than zero the service tag of the preceding packet is always greater than or equal the current virtual time in SCFQ. Hence, in this case the exact arrival time is not significant and the service tag can be calculated when the packet reaches the first position in the queue (Eq. (4)).

$$F_i^k = \frac{1}{\phi_i} L_i + max(F_i^{k-1}, F_l^j) \quad (4)$$

The adoption of this mechanism to wireless ATM MAC protocols is discussed in Section 4.

4 Hierarchical Scheduling in Fair Queueing MAC Protocols

A mobile station (MS) running multimedia applications on high speed networks may very likely use several active connections in parallel. If these connections have different QoS requirements, they cannot simply be multiplexed in a common MS FIFO queue. Hence, a scheduler must determine the service order. In wireless ATM MAC protocols connections may be treated either separately by the scheduling mechanism in the base station (BS) or several connections may be pooled and handled as one flow by the central scheduler. In order to realize the second approach an additional scheduler is needed in the mobile station. We refer to this as hierarchical scheduling.

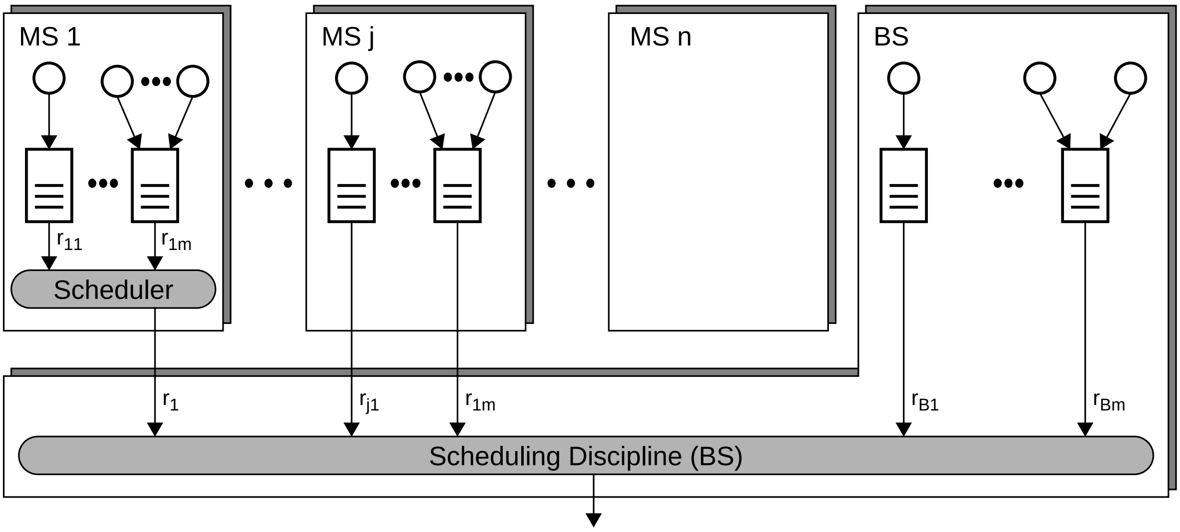

Figure 3: Hierarchical Scheduling Strategy

In Figure 3 a mixture of hierarchical and non-hierarchical scheduling is depicted. While the connections of MS 1 are jointly treated in the base station scheduler, the connections of MS j are treated separately. Depending on the QoS requirements, several connections may use a common queue, e.g. several UBR connections. In the non-hierarchical approach the scheduling algorithm inside the base station can determine for each connection individually when it will be served. In order to make the right decision, the base station must have detailed information about the status of all queues in the mobile stations. This causes a high overhead for the exchange of status information and hereby decreases the throughput of the system. Moreover, the information in the base station is very likely to be outdated and therefore the „wrong“ connection may be served. As the BS scheduler has direct access to the queues located inside the base station they can simply be served separately.

In the hierarchical approach the mobile station scheduler informs the base station scheduler only about the total number of waiting packets. Hence, a smaller amount of bandwidth is needed to update this information. Furthermore, if using SCFQ scheduling there is no need to send queue state information via random access as long as there is at least one cell waiting in the mobile station. The scheduler in the mobile station may use different scheduling strategies, it may use the same as the base station scheduler or a different one, e.g. a simple priority based scheduling or round-robin which we use in the following.

In our approach we use the frame structure depicted in Figure 2 and the SCFQ scheduling strategy in the base station scheduler which serves the base station queues on a per connection basis. As described in Section 3, if a cell arrives at a non-empty queue the scheduler does not have to know the exact arrival time as it can directly derive the service tag from the preceding cell of this flow. Therefore, a mobile station does not have to send an update packet in the uplink control slots but will send the new queue length information piggybacked on its next data packet. If a new cell arrives at a previously empty queue, the mobile station sends a request packet in an uplink control slot. Access to these slots is contention based according to the splitting algorithm described in [6] which polls individual terminals if they collided twice. Hence, it guarantees the successful transmission of the queue state information at latest after three frames. In order to get a non-decreasing virtual time function the scheduler in the base station uses the current virtual time at frame header generation to determine the service tag for the first cell of the corresponding flow.

At the end of each frame the base station determines the allocation of slots for the next frame. For this it saves the number of waiting cells of each flow and maintains an ordered list containing a service tag for the first element of all active flows. Scheduling is done in two steps:

1. Up to a maximum number of iterations (maximum frame length) or until all queue length counters are zero the scheduler removes the first element of its service tag list, generates a transmission permit, decreases the corresponding queue length counters by one, and, if it is not zero, a new service tag is calculated using Eq. (4) and inserted into the service tag list.
2. Permits are then separated in uplink and downlink permits and included in the frame header. Note, that this separation is necessary to limit the overhead due to turn-around and guard times.

In parallel, the virtual time is updated according to the service tag of the cell currently in service. If the base station scheduler in the proposed hierarchical scheduling strategy schedules a cell of a mobile station, this mobile station can still decide which of its connections must be served first. Therefore, even if the base station doesn't know about the arrival of an urgent cell at the mobile station this cell can still overtake other cells at the mobile station and be transmitted in the next mobile station data slot.

5 Simulation Results

In computer simulations we compared two different schemes, the hierarchical scheme where all connections of the mobile station are jointly treated by the base station scheduler and the non-hierarchical case in which all connections are served separately. In the hierarchical case a round-robin scheduler is used in the mobile station to serve its queues. We assumed a scenario with five mobile stations where each of these sets up five connections in either direction. Four on-off sources and one background Poisson source generate ATM cells. During the on periods the on-off sources equidistantly emit cells at peak cell rate. The number of cells in this period is derived from a shifted geometric distribution. The duration of the off period is negative exponentially distributed. The selected values for the source parameters as well as their corresponding weights ϕ_i are depicted in Table 1. For the Poisson background traffic lower weights are selected as this traffic should have lower priority. In the hierarchical case the sum of all weights is assigned to a mobile station.

Table 1: Source Parameters

Source	peak rate	mean burst size	mean rate	ϕ_i
on-off1-4	2 Mbit/s	10 cells	300 kbit/s	2,5
Poisson	-	-	500 kbit/s	1,0

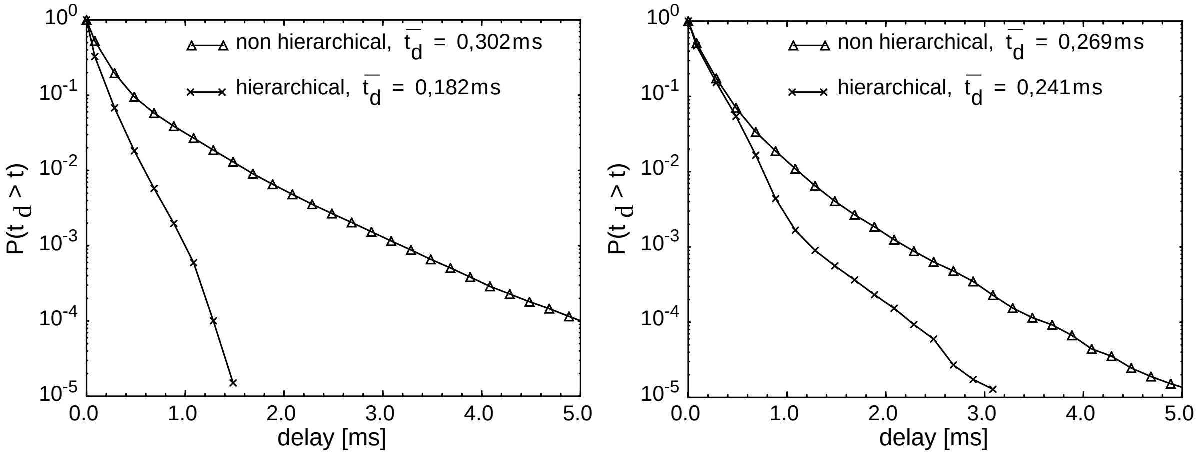

Figure 4: Uplink direction, Poisson and on-off sources

For the MAC protocol we assume a raw data rate of 30 Mbit/s including all overhead. We used flexible frame length with a maximum of 20 data slots and an adaptive number of request slots according to the algorithm described in [6]. The size of the frame header consist of a fixed part and a variable part which is dependent of the number of permits granted for this frame. The guard times, described in Section 2, are set to 1µs between uplink slots and 5µs between uplink and downlink phases (transceiver turn-around time). The size of the uplink and downlink data slots were chosen as 60 Bytes including preamble and control information. The size of request packets is 9 Bytes. Assuming a maximum frame length of 20 data slots and three request slots this results in a gross system rate of 23 Mbit/s (including ATM and AAL header overhead). The system load was chosen to 74%, i.e. 17 Mbit/s.

First results obtained from computer simulations are shown in the following figures. First we envisage the uplink direction. In Figure 4 the complementary delay distribution functions experienced by the Poisson as well as the on-off sources are depicted. As shown there, the traffic strongly profit from the proposed hierarchical scheme and also the mean value of the delay ($\bar{t}_d$) is lowered. Compared to the on-off sources the Poisson source benefits more from the hierarchical approach. This results from the round-robin scheduling strategy in the mobile station where all connections are treated equally in contrast to the non-hierarchical case where the weight of the Poisson source is lower compared to the on-off sources.

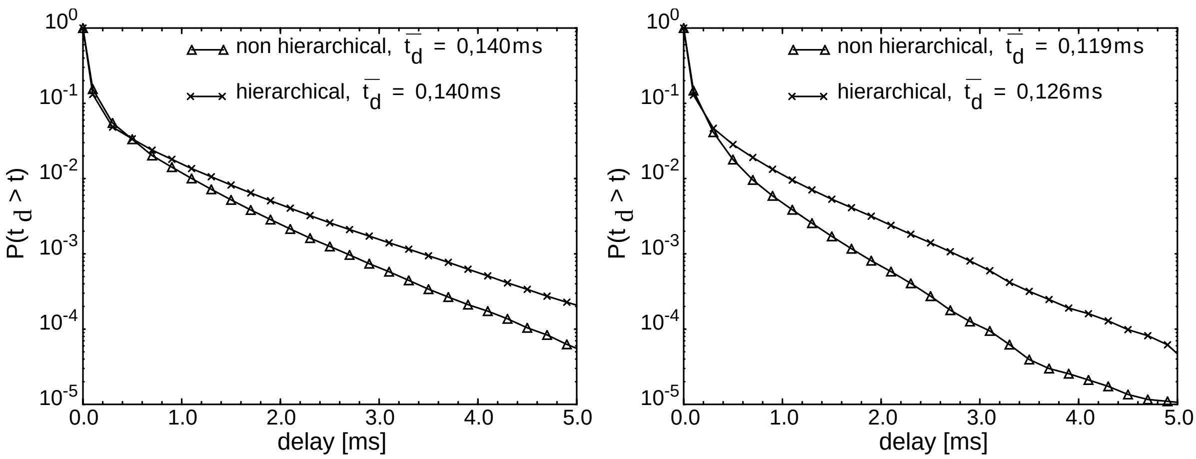

Figure 5: Downlink direction, Poisson and on-off sources

As the results obtained for the downlink direction (see Figure 5) show, the connections slightly suffer from a higher delay in the hierarchical case. In the hierarchical scheme the weight of a mobile station was chosen as the sum of the weights of its connections. Therefore, even if only one uplink connection of a mobile station is currently active it will be served at this higher rate. Hence, the performance of the downlink connections is decreased.

6 Conclusions

In this paper we have proposed the application of fair queueing strategies to TDMA/TDD wireless ATM MAC protocols. Moreover, the concept of hierarchical scheduling is introduced and evaluated by computer simulations. From system architecture point of view this concept allows to distribute the scheduling complexity over base stations as well as mobile stations. As a consequence, this leads to a rather simple and general base station architecture, whereas specific requirements for services are taken care of in the mobile station. Furthermore, the results of our performance investigations have shown that the hierarchical scheme lowers the delays of uplink traffic while at the same time only slightly affecting the performance of the downlink connections. Currently, we are investigating the use of more advanced queueing strategies in the mobile station. Moreover, in order to improve the performance of the downlink connections sophisticated assignment schemes for weights in the hierarchical scheduling case will be introduced.

References

[1] ACTS AC094 EXPERT, WPG4: „Initial Specification of Bandwidth Management“, ACTS AC094 Deliverable 4A, April 1996.

[2] ATM Forum: „Traffic Management Specification Version 4.0“, AF-TM-0056, ATM Forum Technical Committee, April 1996.

[3] ETSI BRAN project (former ETSI RES 10 working group): „HIgh PErformance Radio Local Area Networks (HIPERLANs): Requirements and Architecture“, working document no. 10-07.

[4] S. J. Golestani: „A Self-Clocked Fair Queueing Scheme for Broadband Applications“, Proceedings Infocom’94, Toronto, Canada, June 1994.

[5] ITU-T: „B-ISDN ATM Layer Cell Transfer Performance“, Draft Recommendation I.356, Geneva, May 1996.

[6] D. Petras, A. Krämling: „Fast Collision Resolution in Wireless ATM Networks“, Proceedings 2nd MATHMOD, Vienna, Feb. 1997.

[7] J. Porter, A. Hopper, D. Gilmurray, O. Mason, J. Naylon, A. Jones: „The ORL Radio ATM System, Architecture and Implementation“, ORL Technical Report No. 96.5 (ftp.orl.co.uk:/pub/docs/ORL/tr.96.5.ps.Z), 1995.

[8] M. J. Karol, Z. Liu, K. Y. Eng: „Distributed-Queueing Request Update Multiple Access (DQRUMA) for Wireless Packet (ATM) Networks“, ICC ‚95 IEEE International Conference on Communications, 1995.

[9] A. Krämling, G. Seidel, M. Radimirsch, W. Detlefsen: „Performance Evaluation of MAC schemes for wireless ATM systems with centralized control considering processing delays“, The Second European Personal Mobile Communications Conference (EPMCC‘97), pp. 173-180.

[10] A. K. Parekh, R. G. Gallager: „A Generalized Processor Sharing Approach to Flow Control in Integrated Services Networks - The Single Node Case“, Proceedings Infocom’92, 1992, pp. 915-924.

[11] N. Pronios, I. Dravopoulos, et al: „Wireless ATM MAC Overall Description“, public CEC Deliverable No. 3D1 of the Magic WAND ACTS project, 1996.

[12] K. Rauhala, et al: „Baseline Text for Wireless ATM specification“, WATM working group of the ATM Forum.

[13] D. Raychaudhuri: „WATMnet: A Prototype Wireless ATM System for Multimedia Personal Communication“, IEEE Journal on Selected Areas in Communications, Jan. 97, pp. 83-95.

[14] R. Sigle, T. Renger: „Impact of Wireless Access on Traffic Management in ATM Networks“, Proceedings 1st Workshop on wmATM’98, Hangzhou, April 1998.