Medium Access Control Protocol for transparent ATM Access in MBS

Dietmar Petras

Dept. of Communication Networks, Aachen Univ. of Technology
EMail: petras@dfv.rwth-aachen.de, WWW: http://www.dfv.rwth-aachen.de

Abstract: *This paper presents investigations on a medium access control protocol enabling transparent ATM access to mobile users. The* Dynamic Slot Assignment *protocol has been developed. The results of performance evaluations and improvements of the protocol are described.*

1 Introduction

Goal of MBS is to offer mobile applications a transparent access to fixed ATM networks. "*Transparent*" in this sense should be interpreted as "*service independent*" and therefore "*transparent for the ATM adaption layer (AAL)*".

To enable transparent ATM access, the protocols of the AAL are not allowed to be executed on the air interface. This results in a layering shown in figure 1. As comparison, figure 2 shows the solution of UMTS offering a more efficient usage of the channel resources for some specific, commonly used services.

The transmission of ATM cells by means of virtual channels over the air interface leads to a MAC protocol, that extends the statistical multiplexing from ATM multiplexers to the specific situation characterized by multiple access of not easy to co-ordinate terminals. The MAC scheme has to allocate the medium to mobile stations on demand dependent on the current transmission requirements. Especially, the transmission of ATM cells from real-time oriented services requires a highly dynamic priority controlled MAC protocol to enable short delays whenever necessary.

The delays on the air interface are caused by competition of multiple mobile stations on the uplink and on the other hand by necessary retransmissions caused by a higher BER compared to fibres. The retransmissions have to be controlled by a special ARQ protocol [2]. The uplink is the critical part for the MAC scheme because of contention and danger of collisions. A MAC protocol for transparent, mobile ATM access has to coordinate the capacity allocation under consideration of the maximum delay of each virtual channel and the current queueing delay of ATM cells. This results in a transmission of ATM cells in the order of their residual life time. Therefore a priority mechanism in combination with efficient signalling methods and a fast collision resolution algorithm is necessary. The collision resolution algorithm requires a fast synchronous mechanism for transmitting positive as well as negative acknowledgements for collision detection.

Problems occur when using a decentral coordinated priority mechanism. This way all stations are treated equally so that a preferred treatment of certain ATM cells cannot be guaranteed, but only made likely. This leads to the requirement of a central coordinated mechanism.

2 Dynamic Slot Assignment Protocol

The basic idea behind the *Dynamic Slot Assignment* (DSA) protocol [1] is the extension of the statistic multiplexing of ATM onto the air interface. The uplink of one cell behaves like an ATM multiplexer. Within a mobile station the concurrently existing *virtual channels* (VC) and *virtual paths* (VP) are multiplexed onto a *traffic channel* (TCH). The TCHs of all mobile stations in one cell are multiplexed onto the same *physical channel* (PCH). Thereby the maximum number of mobile stations or TCH's per

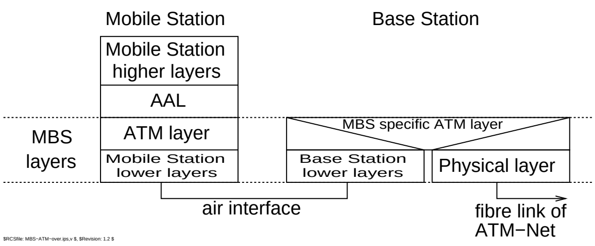

Figure 1: Layering for transparent, mobile ATM access

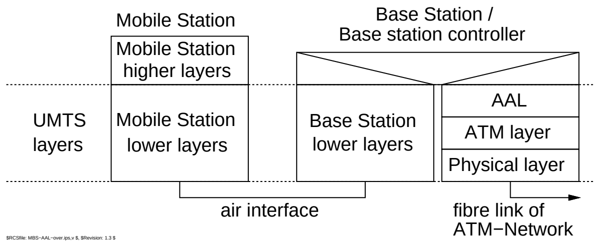

Figure 2: Layering with service specific transmission

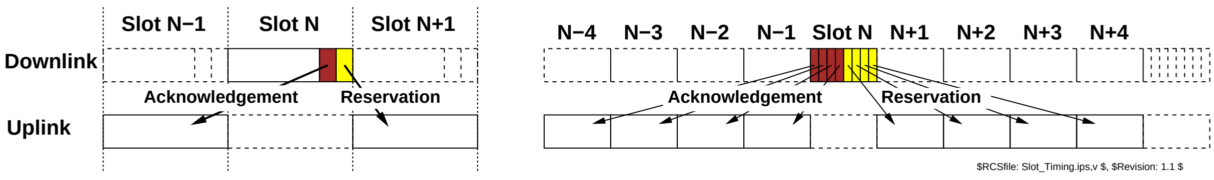

Figure 3: Timing relation between uplink and downlink when signalling in every slot header or special broadcast slots

PCH depend on their capacity requirements. Because of the statistic multiplexing the throughput of a PCH increases with the number of mapped TCHs. Therefore the capacity of a PCH should be as large as possible. In the most favourable case a carrier is not divided into several PCH's by TDMA. The DSA protocol requires a FDD system with a fixed time relation between uplink and downlink. Within MBS one PCH offers 36Mbit/s on the uplink as well as on the downlink.

To be able to fulfill the quality requirements of all VC, the assignment of transmission slots to a TCH has to be done in dependency of the transmission requirements of all other TCH's. Besides the static Quality of Service (QoS) parameters determined during connection setup (*connection specific parameters*), especially the short term, fast changing requirements (*dynamic parameters*) have to be taken into account. Therefore, the base station as central instance collects the connection specific and dynamic parameters of all mobile stations. Based on this data, single transmission slots are assigned (*Dynamic Slot Assignment*, DSA) and establishment of new TCHs is controlled. The capacity assignment is executed on a slot–by–slot basis leading to very dynamic reactions.

The reservation of uplink slots is done by short signalling messages on the downlink. Each downlink burst contains a reservation address in its header for the following uplink slot or several such addresses are grouped together in one downlink burst. Together with the reservation addresses for the next following N slots acknowledgement messages for the last N slots are transmitted. The sending of single slot reservations is illustrated in figure 3 (left). This permits very short reaction times. But all downlink bursts have to be sent in broadcast mode, since the receivers of reservation and acknowledgement messages and payload in general are not the same. This prevents the usage of power control mechanisms. A lower dynamic is achieved by grouping several signalling messages to a special signalling burst (figure 3, right), which can be transmitted in broadcast mode. Then in case of normal ATM cell bursts power control mechanisms can be applied.

It turned out to be difficult to update the central data base inside the base station, since it has to be informed of the changing capacity requirements during all times. Therefore each uplink burst contains the dynamic parameters of the sending mobile station in its header. Examples of dynamic parameters are:

- Number of waiting ATM cells
- smallest residual lifetime of all queued ATM cells

But often situations occur, in which the mobile station's necessity for transmission is increased without it being able to inform the base station. Two cases are possible:

1. If a transmitted ATM cell is the last one in its queue, this mobile station will not request further slots. Since it only makes sense to assign a slot to a mobile station if a transmission in that slot can be guaranteed or at least is very likely, this mobile station cannot inform the base station on an arrival of a new ATM cell.
2. The smallest residual lifetime of all queued cells in a mobile station can decrease very fast if, e.g. a mobile station has several parallel VCs with different maximal delays and a new ATM cell with a very small maximum delay is generated. In this case the base station will handle the mobile station with a too low priority.

In these cases an alternate way of transmitting dynamic parameters is necessary. It is done by cyclic insertion of special slots, which are intended for random access. For simplification a sequence of these contention slots is called *random access channel* (RACH). To resolve collisions a collision resolution algorithm is necessary. To prevent unnecessary collisions, only those mobile stations may access the RACH, which state corresponds to one of the two cases stated above. Such a mobile station is said to be in *contention mode*. Likewise, a mobile station, which has sent its dynamic parameters to the base station and according to its need of transmission can expect a reserved slot, is said to be in *reservation mode*. With this combination of reservation and contention it is not necessary for mobile stations with bursty services in times of low data rates to request reservation slots. Section 2.3 explains, how the efficiency of the RACH and with it the dynamic of the DSA protocol can be increased by dividing contention slots into subslots and only submitting the dynamic parameters to the random access.

2.1 Slot assignment strategy in the base station

The base station collects the connection specific and dynamic parameters of all mobile stations. Based on this data base it determines a reservation address for each slot (the address of the mobile station, which may transmit in this slot) and broadcasts it on the downlink. The reservation address can be obtained in an efficient way by usage of a priority mechanism. A priority is calculated for each mobile station and the next slot is assigned to the station with the highest priority. For mobile stations in contention mode there has to be an alternate way of access. Therefore sufficient contention slots have to be inserted. This is solved in a flexible way by calculating a priority for the RACH taking into account the amount of previous collisions in order to resolve collisions. Different methods for optimization of random access are examined in sections 2.2 and 2.3.

Calculation of a mobile station's priority depends upon its request for more slots or its message of switching to contention mode. Priority of mobile stations in reservation mode is calculated by formula 1.

$$P = \log\left(\frac{a}{\tau_r}\right) + b \cdot l \qquad \begin{array}{rl} \tau_r: & \text{residual lifetime} \\ l: & \text{queue length} \\ a, b: & \text{scaling coefficient} \end{array} \tag{1}$$

If a mobile station has informed the base station about its change to contention mode, its queue length is zero and residual lifetime invalid. Such mobile stations must transmit using the RACH. Because the throughput on the RACH is low as a result of collisions (Slotted Aloha: max. 37%), it is useful to reduce access to the RACH by polling. With the usage of polling the maximum delay can also be guaranteed (if no permanent overload situation occurs requiring handover to another frequency channel) and thus prevents instability of the RACH. Therefore, priority for mobile stations with $l = 0$ is calculated by formula 2.

$$P = c \cdot \overline{r} \cdot \tau_f \qquad \begin{array}{rl} \overline{r}: & \text{mean rate} \\ \tau_f: & \text{time passed since last successful transmission} \\ c: & \text{scaling coefficient} \end{array} \tag{2}$$

2.2 Random access and collision resolution

Calculation of priority for the RACH is based on following parameters:

$\sum \overline{r}_{RACH}$: sum of mean rates of all stations in contention mode
$\tau_{\text{RACH}_{free}}$: time passed since last contention slot
$\tau_{\text{RACH}_{max}}$: maximum duration between two consecutive contention slots

By using a collision resolution procedure it is not possible to calculate the RACH priority by a simple formula. The algorithm has to minimize the probability of collisions and in case of occurrence of a collision will keep the resolving period short. Unlike conventional MAC protocols the DSA protocol offers two control possibilities:

1. The behaviour of mobile stations can be controlled by periodically broadcasted parameters. E.g. the probability of transmission can be set according to the current load of the RACH.
2. Density of contention slots can be increased. After collisions (implying a mechanism for collision detection in the base station) several additional contention slots can be inserted to resolve collisions fast. This can easily be done by increasing priority depending on the state of the collision resolution algorithm.

A heuristically deducted algorithm uses the number of sending mobile stations as measure for the load on the RACH estimated by the sequence of collisions and free slots. Priority is calculated dependent on the current state by formula 3:

$$\begin{array}{l} \texttt{if}\ (\tau_{\text{RACH}_{free}} < \tau_{\text{RACH}_{max}}) \\ \qquad P_{Normal} = \sum \overline{r}_{RACH} \cdot \tau_{\text{RACH}_{free}} \\ \texttt{else} \\ \qquad P_{Normal} = \infty \\ \\ P_{Collision} = \infty \end{array} \tag{3}$$

Simulations have shown that the algorithm adjusts its behaviour according to the dynamically changing capacity requirements on the RACH. In future studies an optimization is to be done by using analytical methods.

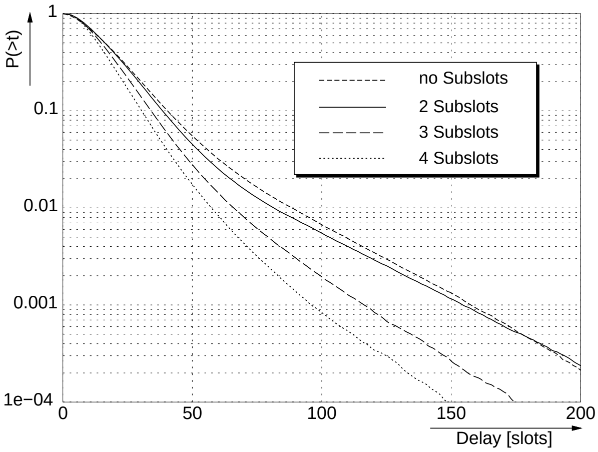

Figure 4: Complementary distribution function of delay for different numbers of subslots

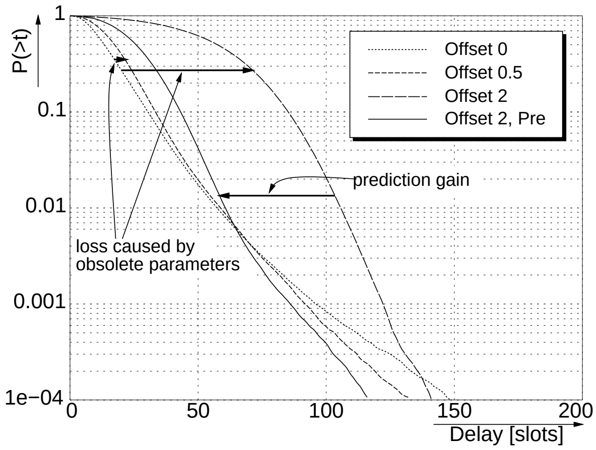

Figure 5: Influence of Offset in the physical layer on the distribution of delay

2.3 Improved random access with subslots

Efficiency of the RACH can be increased if the amount of data lost in a collision is minimized. In the version of the DSA protocol previously examined, normal bursts with ATM cells as payload are used in contention slots. But it is sufficient to transmit only the dynamic parameters, since the base station will then schedule a reservation slot for transmitting the ATM cell. The dynamic parameters can be transmitted in short bursts so that a contention slot may be divided into several subslots in order to provide more contention periods. Because of more contention subslots and constant input the collision probability decreases. Therefore the mobile stations switch faster into reservation mode leading to shorter delays.

Because of the overhead for guard time, training sequence etc. in each subburst, a contention slot cannot be subdivided indefinitely. Based on the burst structure of MBS [3] four subslots seem to be realistic.

The algorithm for subslots is similar to that of normal contention slots, but there are two major differences:

- A mobile station doesn't have to save a RACH burst for later retransmission, since a newly generated burst of newest dynamic parameters will be transmitted after a collision.
- Not only the access probability but in addition the position of subslots in which is transmitted can be varied.

Measurements have been executed using a simulator and varying amounts of subslots. Figure 4 contains the complementary distribution function of transmission delay for negative-exponential distributed inter arrival times.

3 Consideration of signal processing time

Up to now it was supposed that a reservation message in slot(n) would be followed by a reply in slot($n+1$). But in a real system the physical layer of a mobile station takes time for signal processing of the received burst, so that the contents of the burst is not available before a duration of τ_{Sig}. By introducing an offset between uplink and downlink of $\tau_{off} = \frac{1}{2} \cdot \lceil 2\tau_{Sig} \rceil$ the necessary time for signal processing is provided in mobile and base station. Realistic values for τ_{off} are in the range from 0.5 to 2.

Because of this offset the reaction speed of the DSA protocol is reduced. Priorities are calculated based on obsolete values. Therefore the dynamic parameters have to be corrected (*prediction*).

In figure 5 the influence of offsets determined by simulation on the distribution of delays is shown. Comparing the curves "Offset 2" and "Offset 2, Pre" clearly shows the gain caused by prediction.

4 Consideration of incorrect signalling caused by transmission errors

Fading or interference can cause transmission errors or losses of whole bursts. This can result in errors in the protocol execution which leads to decreasing channel efficiency and violation of QoS. The DSA protocol should be able to handle such situations.

Transmission errors can lead to following problems:

1.: If a mobile station doesn't receive a reservation message a certain uplink slot will not be used. The probability for the occurrence of this situation can be reduced by usage of *forward reservation*. The base station not only sends the reservation for the next, but for the next n slots. Thus each reservation message is repeated n times.

2.: If a base station doesn't receive a burst in a reserved slot it has to operate with old dynamic parameters and will continue to assign reserved slots to that mobile station. Problems occur, when the mobile station discards the ATM cell, because its maximum delay has been exceeded. In this case the mobile station switches to contention mode and will not use the reserved slot. To keep unused capacity as low as possible, the base station will reset the dynamic parameters after a fixed amount of time without an answer, forcing the mobile station to use the RACH. A timer in a mobile station will switch it into contention mode, if a fixed amount of time has passed without it receiving any reservation message.

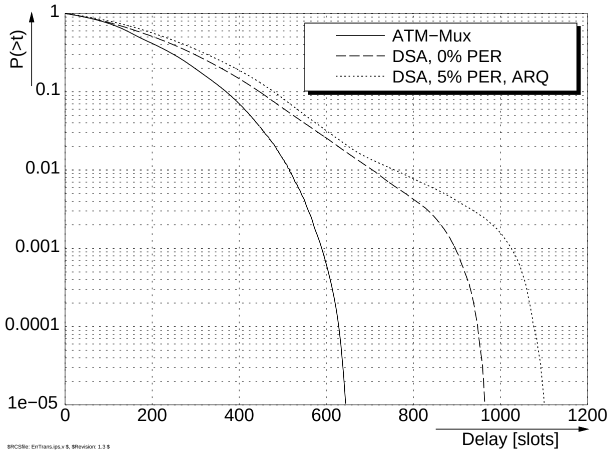

Figure 6: Complementary distribution function of delay when considering transmission errors and retransmissions (7 video sources, 76% load)

Simulation results in figure 6 show the influence of transmission errors and retransmission by a simplified ARQ protocol on transmission delay. Several video sources are used, each offering a data rate of 3.9Mbit/s. The resulting load is 76%. Maximum delay of this source is 1548 slots. To simplify things, the delay caused by transmission of acknowledgements of the ARQ protocol is set to zero. The delays increase in an acceptable manner regarding the load which due to retransmission increases to over 80%. For comparison, an ideal multiplexer with the same data rate (here 36Mbit/s) is given as lower bound of delays.

5 Conclusions

It has been described that the DSA protocol corresponds to the requirements on a MAC protocol for transparent ATM access. The simulation results show, that performance loss caused by transmission over the air interface still remain in the range of QoS requirements of most services. Especially the additional delays caused by contention can be taken care of in a sufficient manner.

Further studies will deal with the optimization of algorithms for collision resolution and priority calculation. Therefore traffic theoretical analyses are necessary.

A problem remains in the dynamic frequency assignment to base stations. Because of high bandwidth requirements mobile ATM networks will use pico cellular structures, which will make impossible static frequency assignments. Therefore the DSA protocol has to be extended by an algorithm for *Dynamic Frequency Selection*. This algorithm will be similar to the *Dynamic Channel Allocation* scheme of DECT–like systems.

6 Acknowledgements

The author would like to thank the members of the MBS project for their suggestions. Furthermore this work would not have been possible without the hard work and enthusiasm of the students working on their diploma thesis.

7 References

[1] D. Petras. *Performance Evaluation of Medium Access Control Schemes for Mobile Broadband Systems.* In *DMR VI*, pp. 255 – 261, Stockholm(S), June 1994.

[2] D. Petras. *Performance evaluation of the ASR-ARQ Protocol for MBS.* In *RACE Mobile Telecommunications Summit*, Cascais(P), November 1995. to be published.

[3] M. Progler. *MBS Air Interface Priciples.* In *RACE Mobile Telecommunications Summit*, Cascais(P), November 1995. to be published.