FAFS: A New MAC Protocol for Wireless ATM

Sandeep Jain
Dept. of Computer Sc. & Engg.
Indian Institute of Technology, Kanpur (India)
e-mail: `sandeep@cse.iitk.ac.in`

Vinod Sharma
Indian Institute of Science, Bangalore (India)
e-mail: `vinod@ece.iisc.ernet.in`

Dheeraj Sanghi
Dept. of Computer Sc. & Engg.
Indian Institute of Technology, Kanpur (India)
e-mail: `dheeraj@cse.iitk.ac.in`

ABSTRACT

Wireless ATM (WATM) is being envisaged as the next generation wireless solution for meeting the demands of the users. Besides supporting high data transfer rates, it also provides QoS (Quality-of-Service) guarantees to the connections. Since the wireless channel is error-prone (Bit Error Rate $> 10^{-3}$) and bandwidth is scarce, it becomes an important design issue to efficiently allocate bandwidth among users. MAC (Medium Access Control) layer arbitrates the channel access to the wireless medium and is also responsible for bandwidth allocation to different users. In this paper we present a new MAC protocol, FAFS (Fair Access Fair Scheduling), which provides a new contention scheme along with a novel scheduling algorithm that allocates data slots to the sources based on their packet queues, negotiated traffic requirements, signal intensity and the last time when the channel was granted to the connection (for best effort service only). Our results show that the delay probability of the real time packets is decreased using the techniques mentioned in the paper. We also provide improved service to the best-effort sources by avoiding large queues at the mobile terminal.

I. INTRODUCTION

Future high speed networks will support different QoS requirements for the integrated services they will provide.To provide Personal Communication Services certain parts of the network may involve wireless links. In case of ATM networks it will be advantageous to support the ATM protocol also. However as against wired medium the wireless channel is much more error prone and the available bandwidth is scarce. This can result in substantial delays and packet loss. Therefore, careful planning is required to support the ATM protocol efficiently which was originally designed for very high capacity and reliable optical fibers. Also, since the delays and packet loss in the wireless medium will form a substantial part of the end-to-end performance, it is critical to satisfy these bounds in the wireless medium to satisfy the QoS requirements of the users. This has resulted in a spurt of activity to support QoS requirements in the wireless medium (see [9] for a survey).

In this paper we provide one more solution to the QoS support in the wireless channel and consider several of the above issues not addressed in many previous solutions. The problems we address are relevant in providing QoS in wireless medium which support other network technologies also (e.g. Internet).

Important decisions made by the MAC protocol include granting channel access to the mobile terminals (MTs) and scheduling the active connections in such a manner that each user gets the promised QoS. Other services include encryption support, good adaptability, conserving power at the MTs etc. We use the queue information at the MTs and the leaky bucket concept to decide the number of data slots a source is supposed to get.

FAFS is a protocol for the Infrastructure Wireless networks where there is a single base station (access point to the backbone network) and many mobile terminals. In other words, it is based on a client-server paradigm where the server (BS) is used to arbitrate the channel among different clients (MTs).

Ours is a TDMA (Time Division Multiple Access) kind of protocol in which only one carrier frequency is used for transmission. This has been chosen to allocate bandwidth for uplink (UL) and downlink (DL) connections according to their requirements. Giving one carrier frequency for each may result in inefficient utilization of bandwidth.

The MAC protocols proposed for Wireless ATM include DQRUMA [5], PRMA/DA [6], DSA++ [8], MASCARA [7], RQMA [4]. Other good related references are [2] and [3]. All these protocols use a single contention period for channel access. FAFS is unique in the sense that it provides two different contention slots (one for CBR (Constant Bit Rate) and VBR (Variable Bit Rate), other for ABR (Available Bit Rate) and UBR (Unavailable Bit Rate)) and thus providing QoS at the contention level itself. The MAC protocols differ in the way they allocate data slots to sources. Some protocols do not consider priority among sources at all [5] while other use parameters like leaky bucket [7], negotiated traffic descriptor [6] or number of waiting ATM packets [8] to allocate data slots. Our protocol uses a combination of these parameters along with the fading information to allocate data slots to sources. This helps in improving the throughput for real time sources and reduces the overall delay in transmission. The protocol also uses a novel feature to avoid denial-of-service for the best effort class of services. A UBR source transmits a parameter to the Base Station signifying its unsuccessful contention frequency. The BS reserves some slots for the UBR source if the value of the parameter exceeds some fixed threshold. FAFS is adaptive to changing network traffic. It has the provision for dynamically controlling the contention period length which limits the number of MTs that can access the channel at a particular time.

We use minislots for channel access which helps to decrease wastage of bandwidth in case of collision. Power constraints at the MT are also taken into consideration wherein the MT notifies to BS when it starts sleeping (i.e. enter into sleep mode - no reception no transmission mode). This allows BS not to consider that particular MT while assigning data slots in subsequent frames.

Fixed sized frames have been used for information transfer between MT and BS . Thus an upper bound can be obtained on the delays of packets that are to be sent from the BS to the MT and vice-versa.

The rest of the paper is organized as follows. Section II describes in detail the schemes, procedures and algorithms specific to FAFS. Section III summarizes the simulation results. Finally Section IV gives the conclusion.

II. FAFS

We consider a fixed length frame for transmission between BS and MT. This frame is shared by both UL and the DL connections. Four classes of service are considered (as in the case of wired ATM) i.e CBR, VBR, ABR and UBR - considered in order of decreasing priority.

The MAC frame is divided into six fields: synchronization, contention (type 1 and type 2), acknowledgment and data. Field boundaries are variable in nature. The MAC frame is shown in Figure 1.

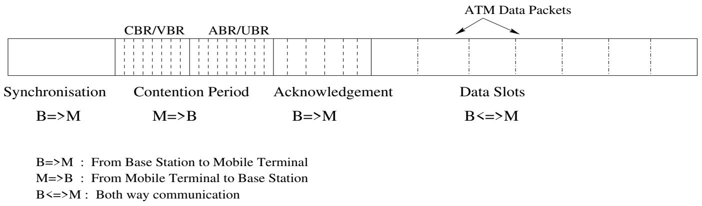

Fig. 1. MAC frame

Synchronization field is used to transmit information about the field boundaries to the MTs and also to synchronize the MTs with the BS. Contention field is an important part of the frame as the MTs request for the channel in this slot. The Acknowledgement period is used by the BS to transmit information about the allocated data slots to the respective MTs. The data slot is used for actual data transfer between the BS and the MTs (both UL and DL).

The number of slots in the contention field poses an upper (soft) limit on the number of active connections that can be set up with the BS and consequently on the number of the MTs to which it can provide service. Hence the length of contention field is a critical parameter for design of the MAC frame.

In order to provide QoS we use two different contention slots. First slot is used by the CBR/VBR sources and the second slot is used by UBR/ABR sources. This is done so that the higher priority sources (CBR/VBR) do not have to contend with the low priority services (UBR/ABR) for getting access to the channel. This helps in ensuring delay bounds for the CBR, VBR traffic. A contention field is made up of several *minislots*. Within their contention field, the sources request for data as follows. If a contention field has n minislots, any user requesting in that field picks one of the minislots with probability $1/n$ for sending its channel access requests. If two or more users send their requests in the same minislot, collision occurs. Since we use minislots, the resulting wastage due to collision is not significant. The QoS negotiation for the CBR, ABR and VBR sources takes place at the time of connection establishment. For this a new call first requests for a slot in the contention slot. Once its channel request has been successfully received at the BS, the relevant information needs to be transmitted by the MT in the data slots. For this, the BS allocates a UL slot in the data portion of the MAC frame to the new source (CBR, VBR and ABR) to allow it to send the necessary information about the QoS desired (traffic descriptor). Since UBR sources come under the "best-effort transmission" class, no QoS is guaranteed to them and consequently no QoS negotiation with BS takes place. The mini-slot for UBR/ABR contains a k-bit access id (to identify the MT) and a 8-bit field (to define the connection state) (Figure 2).

	Counter for Channel Access	Queue Condition	Sleep Mode	Type of Connection	Terminate Connection
No. of Bits	3	2	1	1	1

Fig. 2. Connection State for ABR/UBR sources

First three bits are used to specify the channel denial frequency. An active source might not get any data slots if there is a collision during contention or higher priority sources are alloted the channel. If the UBR source sends a request in the contention slot and does not get back an acknowledgment it increases the 3-bit counter (the first three bits) by 1. This counter gives an indication to the BS about the number of times service-denial has taken place to the UBR source. The larger the value of this field, more is the time the source has been waiting to get access to the channel. If the counter reaches its maximum value it is not incremented further. It is reset to 0 only when it gets data slots for transmission. This information is used to differentiate among the various active UBR sources which have approximately the same queue length but different waiting time.

Next two bits are used for specifying the queue condition at the MT. There can be 4 possible states using 2 bits. These four states can be used to model the queue condition at MT viz. normal queue, building up of queue, worsening situation and abnormally large queue at MT. Next bit being on indicates that the MT wants to enter the sleep mode. An active connection may want to enter "sleep mode" if it detects power shortage at its end.

Next bit is used to specify the type of connection i.e. UBR/ABR. The last bit is used by MTs to terminate the connection (for CBR, VBR and ABR only).

For the CBR/VBR sources the mini-slot is same as that for UBR/ABR sources except that the initial 3-bit counter is absent. This is because an active CBR source will be allocated a fixed number of slots in each frame hence there is no point in using the information about the time it last accessed the service. For the VBR source the queue length and the tokens in the leaky bucket give adequate information about the source traffic. If an active VBR source has not accessed the channel for an appreciable amount of time, its bucket will be full of tokens. Once a CBR source request has been successfully received at the BS and the connection is established, it does not need to contend for the channel again. When it wants to terminate the connection, it sends an end connection message in the contention slot. ABR sources also behave in the similar fashion. On the other hand, the VBR and UBR sources have to contend for the channel each time they want to send data UL.

We consider dynamic contention period boundaries. This helps FAFS to accommodate the changes in the network traffic. We use the following algorithm to change the length of the contention period. Initially both contention periods have equal length. This number can be based on the measurements of connect frequency of the MTs to the BS for a particular region in a given interval of time. The length of the contention period is decreased if the percentage utilization of the slots is below 50% and in the rest of slots BS receives no signal. On the other hand if BS receives a corrupted signal in rest of the minislots (due to collision of contention requests), the BS increases the contention period length. The increase/decrease takes place only if the situation continues for k-successive frames. The change in the length of the contention period is gradual (10-20%) so that the protocol adapts to the new situation slowly. This is desired as the network topology for wireless networks changes quite rapidly.

A. *Connection Admission Control*

A Connection Admission Control (CAC) policy is needed to determine if a new connection can be given channel access without violating the QoS requirements of the existing connections. Various models have been proposed for the admission control for ATM networks [1] [10].

For a new CBR source, the bandwidth requested is given by the Peak Cell Rate (PCR) parameter of the traffic descriptor. Using this value and the Cell Delay Variation Tolerance (CDVT), scheduling is performed for the CBR source. If this results in successful inclusion of the new source, then it is granted the channel otherwise it is denied channel access.

A VBR source uses the following parameters in the traffic descriptor: PCR, CDVT, Sustainable Cell Rate (SCR) and Maximum Burst Size (MBS). Given these parameters and a maximum delay of D sec and the Burst Time (BT) the minimum bandwidth requirement for the VBR source can be found by using the function: min(max((1/PCR+CDVT-1)/(D(1/PCR-1)+CDVT), PCR) , max((1/SCR+CDVT+BT-1)/(D(1/SCR-1)+CDVT+BT),SCR)). For a new ABR source, the Minimum Cell Rate (MCR) parameter is used to determine its bandwidth requirements. Again a temporary scheduling is performed and if this is successful, the source is granted the channel otherwise the call is blocked. For UBR sources there is no call admission as there is no bandwidth guarantee for this category of service.

B. *Scheduling Algorithm*

The scheduling algorithm decides the number of data slots for an active source and the position of such slots in the MAC frame. For optimal scheduling, it is necessary to satisfy the QoS of a large number of users and minimize the cell loss due to ATM cell deadline violation. We use two basic parameters for determining this - the queue information at the MT and the leaky bucket at the BS. As packets are generated at the MT, queues start building up. The queue information is passed along with the channel request in the contention frame to the BS. The BS also maintains a leaky bucket for every connection except the UBR source and CBR source. The leaky bucket is used for policing as well as for deciding the number of data slots to be allocated to a particular source. Tokens are added to the bucket at a constant rate (equal to the SCR for the VBR source and MCR for the ABR source). The size of the bucket is governed by the maximum burst size of the connection. Every time a slot is alloted to the source, a token is removed from the bucket. So at any time, the number of tokens in the bucket signifies the bandwidth that is available to the source as per the initial QoS negotiations. The information about queues at the MTs is maintained as a variable at BS.

The scheduling algorithm works as follows. The DL requests are considered first as the volume of DL traffic is generally more than that of UL. The DL and the UL transmission takes place as a single chunk in the data slot. This is to avoid frequent shift between UL and DL transmissions which requires a guard band. CBR and VBR are first allocated data slots based on the packet deadlines and other information. Then ABR sources are considered and given their minimum data slots. If there are more slots left, ABR additional bandwidth requests and UBR sources are considered (in that order). In case there are no slots left and there are UBR sources with maximum counter value we reserve a single slot to one of them (chosen randomly). The detailed algorithm is described below.

After contention for the channel has taken place, there are DL as well as UL requests for the channel. First DL connections are considered and then the UL ones. The deadlines of all the active real-time connections (CBR and VBR), awaiting transmission DL, are calculated first. Let these deadlines be $d_1, d_2, ...d_n$. These deadlines are then sorted in ascending order. The CBR source having the earliest deadline is chosen first for data slot allocation. Each CBR source is alloted a constant number of data slots per frame. This number is decided by the traffic descriptor of the source and is negotiated at the time the CBR source is assigned the channel. This is repeated for all active CBR sources. Then VBR sources are considered in an order defined by the function given below.

$$f(x, y) = 0.7x + 0.3y$$

where
x= Number of tokens for that source
y= Queue length at the source
Priority among VBR sources is thus a function of queue lengths and number of tokens. This gives a fair access to the channel. The number of data slots to be allocated to it is calculated using the function: *min(no. of tokens, queue length)*. Using such a function allows BS to allocate an *optimal* number of slots to the VBR source. The ABR sources have to be given a minimum number of slots per frame. If after above allocation, no data slots remain or the remaining data slots are insufficient to meet the minimum data slots requirements of ABR sources, earlier alloted slots are *deallocated*. This is done by deallocating slots from the largest deadline and selectively removing the slots that were assigned to the VBR sources (CBR sources are not considered for deallocation because of the service guarantees). This is continued until there are enough data slots to fulfill the minimum service requirement of the ABR users. This selective removal might result in gaps in the allocated

slots. This space is filled by left shifting the slots that are allocated later. This shifting does not violate the deadline requirements of the ATM cells as this shift leads to an earlier scheduling.

For the UL allocation, the CBR and ABR sources get the required slots each MAC frame (PCR and MCR respectively). These slots are given in every frame unless the source decides to terminate the connection. VBR sources also get slots according to their deadline requirements and their queue condition. If some data slots remain after such an allocation, the scheduling algorithm looks at the additional bandwidth requirements of the ABR connections and tries to allocate slots for them. If after such an allocation slots are exhausted and there are some UBR connections waiting with the maximum counter value, a random selection of a UBR source from this group is done. This source is allocated a single slot in the MAC frame from the ABR additional bandwidth slots. If there were no ABR additional bandwidth slots available, the UBR source is not allocated any data slot. The UBR sources are thus partially guaranteed some QoS. This selective transmission ensures that denial of service doesn't take place to UBR sources.

The fading information is also used in the scheduling algorithm. If the signal received from the MT at the BS is below a certain threshold , the BS assumes that the MT is going out of sight. These MTs are not considered by the BS for the DL scheduling. The DL connection with such MTs is resumed once the channel condition improves. This increases the probability of successful transmission assuming that the fading condition is momentary (due to mobility of the source).

In order to improve the throughput (and hence overall performance) of the wireless system one should also use forward error correction (FEC). Of course there is a question of which codes to use and how many bits to allocate. Various studies have addressed this issue [4] [3]. Any such appropriate solution can be used with our MAC algorithm to improve the performance.

III. SIMULATION

This section presents the simulation results of the scheduling algorithm. We have simulated CBR, VBR, ABR and UBR sources for UL as well as DL. The simulator has been written in C. The mean transmission rates for CBR, VBR, ABR and UBR are 64 kbps, 128 kbps, 128 kbps and 64 kbps. The peak transmission rate for a VBR source is 256 kbps. For CBR sources packets are generated at a constant rate given by its PCR. For simulating VBR and ABR sources, ON-OFF traffic source model has been used [10]. It has been modeled using a two-state Markov process (the states denoting the ON and the OFF states of the source). In ON state packets are generated at a constant rate τ while in the OFF state no packets are generated. The mean duration of the ON state is taken to be 250ms and that of the OFF state to be 100ms. Taking τ to be 3ms we get a transmission speed of approximately 128kbps. To get the maximum transmission rate of VBR sources the value of τ was halved randomly. UBR sources have been simulated using Poisson traffic arrival model with mean arrival time 200 ms. We measure performance of the protocol in terms of cell loss probability (i.e. the ratio of number of dropped ATM cells to the total number of cells generated) and the mean delay experienced by the packets while in the queues at the MTs. For an efficient MAC protocol the cell loss probability and the mean cell delay should be low.

The channel bandwidth is assumed to be 10Mbps. The synchronization period is of 30 bits, the contention period length is 360 bits for real-time sources while for the non-real time sources it is 450 bits. The acknowledgement period is 1600 bits. Considering an average of 200 ATM data slots per MAC frame the total MAC frame time comes out be 10.3 msec (using bandwidth=10Mbps).

Fading has been implemented as a two-state Markov chain with a good state and a bad state (similar to the ON-OFF source model). Since the channel conditions may be different for different users we have used different transition matrices (giving probability of transition between the two states) for different users. For simulation purposes three different transition matrices have been considered which model three conditions of channel: good, bad and worse. At the time of inclusion of a source in the network, a transition matrix is picked at random which defines the channel condition for the connection till the connection lasts.

The simulation model does not take into account the handoffs that take place between adjacent cells. The effects of propagation delay are not considered in the simulation model. Assuming a cell radius of 500m the maximum back-to-back propagation time comes out to be 3.3 μsec which can be safely neglected considering the frame time to be 10.3 msec. The active mobile terminals always listen to the channel i.e. they do not enter "sleep mode". We have not considered this feature in the simulation for complexity reasons. Nevertheless this feature helps in conserving power at the MT and saving data slots which otherwise would have been idle. We have not calculated the call block probability but this is likely to be a function of call admission policy and the contention period length.

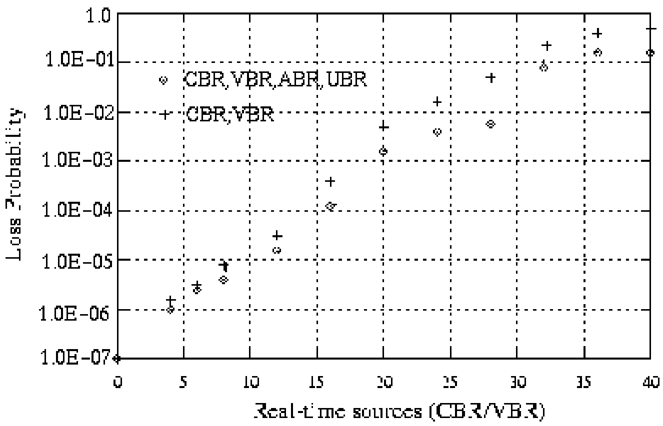

Fig. 3. Loss probability of the real-time sources

The call (connection request) arrival process has been modeled as Poisson process. In our simulations we have considered a maximum of 40 real-time sources (CBR/VBR) and 30 non-real time sources (ABR/UBR). The mean arrival rate for the sources is 10msec. The connection of the sources with the BS is assumed to last till the simulation ends. We observed that if ABR and UBR sources were added to the network having CBR and VBR sources there was only marginal increase in the loss probability (Figure 3). We are considering loss due to violation of deadlines only. This may be attributed to the admission control policy being used and the different contention slots for CBR/VBR and ABR/UBR. VBR sources have to contend for the channel every time they they need to send data to the BS. Similarly UBR sources also contend whenever there queue is not empty. By providing separate contention slots for these we decrease the probability of collision which has a considerable effect in improving the throughput of the system. Since the admission control algorithm is being used, the BS allows more connections only till the promised QoS of the existing connections is being satisfied.

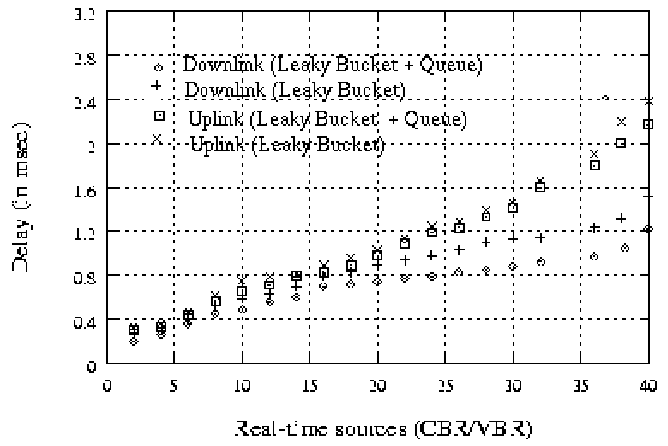

Fig. 4. Overall delay of ATM packets

The mean cell delay is depicted in Figure 4. As we see the DL packets experience lesser delay than the UL packets. This is because the scheduling algorithm first considers the DL data, as intuitively it would always be more than the UL data. We also see that by using both the leaky bucket and queuing information we have slight improvement in the delays experienced by the ATM packets. The queue information helps in considering those sources first that have pending queues at their end. This helps in decreasing the waiting time of the packets in the queue.

We also checked the loss probability with and without fading consideration. The results showed that the fading information helped to increase the total throughput of the system (around 10%) though it resulted in a slight increase in loss probability (around 2%). This can be explained by the conservative approach taken by the BS in attributing a weak signal to the loss of connection between the BS and the MT. Without fading information, BS would have allocated DL slots to the sources irrespective of their channel condition. This might have resulted in successful transmissions in some cases to the BS and thus deadlines of ATM cells for real-time connections would have been honored. But due to poor channel conditions this might result in unsuccessful transmission i.e. decreased throughput. By incorporating fading information, UBR and VBR sources gain most since they have no delay constraints on transmission (for them a *reliable* transmission is important rather than an *early* one).

IV. CONCLUSION

In this paper we described a MAC protocol for wireless ATM network. The unique feature of the protocol is the use of leaky bucket and queuing information to perform scheduling. It also uses a novel concept of having two separate contention fields - one for the real-time sources (CBR,VBR) and the other for the non-real time ones (ABR,UBR). This helps in improving the loss probability of ATM cells (due to violation of deadlines) by making the data slot requests of the VBR channel immune to the requests of ABR and UBR in contention frame. The use of minislots in the contention frame also helps in improving channel performance. The protocol also uses the fading information in scheduling which helps to increase the overall throughput of the system.

An admission control algorithm is also described which governs the entry of new connections to the network. Subsequently the scheduling algorithm is described which allocates slots to sources on basis of fading, leaky bucket, queue information and last channel access time. We thus take into account almost every factor that can affect the performance of the MAC protocol including the network topology. This makes our protocol different from earlier protocols proposed in this field. The simulations show that FAFS improves the channel throughput and decreases the loss probability by judiciously allocating slots to different connections.

REFERENCES

[1] Jaime Jangkok Bae and Tatsaya Suda,“ Survey of traffic control schemes and protocols in ATM networks,” *Proc. of IEEE*, 79(2):170–192, Feb 1992.

[2] C. S. Chang, K C. Chen, M. Y. You, J. F. Chang,“Guaranteed Quality of Service Wireless Access to ATM Networks,” *IEEE J. Selected Areas Comm.*, Vol 15, pp. 106–117, 1997.

[3] S. Choi and K. G. Shin,“A Cellular Wireless Local Area network with QoS Guarantees for Heterogeneous Traffic,” *INFOCOMM’97*, pp. 1030–1037.

[4] N. R. Figueira and J. Pasquale,“Remote-Queueing Multiple Access (RQMA): Providing Quality of Service for Wireless Communications,” *INFOCOMM’98*, pp. 307–314.

[5] Mark J. Karol, Zhao Liu and Kai Y. Eng,“An Efficient Demand-Assignment Multiple Access Protocol for wireless packet (ATM) networks,”*Proc. of International Conference on Communications*, pp. 1224–1231, June 1995.

[6] J. G. Kim and I. Widjaja,“ PRMA/DA: A new medium access protocol for wireless ATM,” *Proc. ICC’96*, Dallas, June 1996, 240-244.

[7] Nikos Passas et al.,“Quality-of-Service-Oriented Medium Access Control for Wireless ATM Networks,” *IEEE Comm. Mag.*, pp. 42–50, Nov 1997.

[8] Dietmar Petras et al.,“Joint Performance of DSA++ MAC Protocol and SR/D-ARQ Protocol for wireless ATM under realistic traffic and channel models”, *Proc. Of wmATM’98*, pp. 1–12, April 1998.

[9] J. Sanchez, R. Martinez,M. W. Marcellin,“A survey of MAC protocols proposed for wireless ATM,” *IEEE Network*, Vol 11, pp. 52–62, Nov/Dec 1997.

[10] D. Yates, J. Kurose, D. Towsley, and M. G. Hluchyj,“On Persession End-to-End Delay Distributions and The Call Admission Problem for Real-Time Applications with QoS Requirements,”*In Proc. of ACM SIGCOMM ’93*, pp. 227–236, March 1993.