A Variable Priorities MAC Protocol for Broadband Wireless Access with Improved Channel Utilization Among Stations

L. F. M. de Moraes, P. D. Maciel Jr.
High-Speed Networks Laboratory - RAVEL
Systems Engineering and Computer Science Program - COPPE/UFRJ
Tel/Fax: +55 21 2562-7844 - Rio de Janeiro - RJ, Brasil
{moraes,paulo}@ravel.ufrj.br

Abstract— **The Broadband Wireless Access (BWA) technology has been proposed to support different traffic classes with distinct quality of service (QoS) requirements in broadband metropolitan area networks. Therefore, as part of its specifications, such a system must properly address the combined requirements of wireless communications and multimedia applications. We propose in this paper an extension to the MAC protocol presented in [1, 2], by assigning variable priorities to the network stations sharing access to the communication channel. An analytical model to evaluate the performance of the proposed protocol is also developed and results obtained for the messages waiting times at different stations are presented. In addition, simulation data are used to compare the results obtained from the analytical model presented previously. It is concluded that the proposed variable priority MAC protocol is able to improve channel utilization and provide throughput and queueing delay fairness among the stations in the network.**

Index Terms— **Broadband Wireless Access, IEEE 802.16, Performance Evaluation.**

I. INTRODUCTION

The Broadband Wireless Access (BWA) is a technology proposed to offer wireless access to network stations in a broadband metropolitan area environment. These networks are designed to operate at high data rates and to deal with several applications, resulting in different types of traffic profiles and demands. Therefore, the system is required to work with various types of real-time and non-real-time service classes, with different traffic characteristics and quality of service (QoS) guarantees.

In [1, 2] , a new MAC scheme for BWA, incorporating a scheduling mechanism based on message and/or station priorities, was proposed as an alternative protocol to the IEEE 802.16 Standard [3]. However, only a fixed priority approach to service differentiation among terminals was used.

In this paper, we propose an extention to the MAC protocol described in [1, 2], with a variable priority scheme among stations in the network. Moreover, we developed a simulation model to compare the analytical results presented in [1, 2] and the simulation results exposed here.

The remainder of this paper is organized as follows. Section II gives a short description of IEEE 802.16 standard. The MAC protocol and the variable priority approach are described in Section III. An analytical model, which allows to obtain the average message waiting time for different priority classes, is provided in Section IV. Section V presents some numerical results and the paper is concluded in Section VI with a few discussions concerning the proposal.

II. IEEE 802.16 STANDARD

We give next a brief description of the MAC layer of IEEE 802.16 and the QoS architecture present in the standard. We noted that the MAC layer of 802.16 only describes support for an QoS architecture, without specifying explicitly how a particular solution should be exactly implemented.

A. MAC Layer

In the basic architecture there are one Base Station (BS), and one or more Subscriber Stations (SSs). Transmissions are assumed to take place through two independent channels: a Downlink Channel (DL) from the BS to the SSs, and an Uplink Channel (UL) from the SSs to the BS. Hence, there is no contention associated with the the DL channel, while the UL channel must be shared by the SSs through the use of some multiple access control protocol.

During the DL, only the BS transmits in broadcast to all the SSs. The BS determines the number of slots to be allocated for each SS in the UL, and broadcast this information in an UL-MAP message at the beginning of each frame. The stations transmit their data in predefined time slots as indicated in the UL-MAP. A scheduling module for the UL is necessary to be kept in the BS in order to determine the transmission opportunities using the bandwidth requests (BW-Request) sent by the SSs. Figure 1 illustrates the structure of the MAC frame.

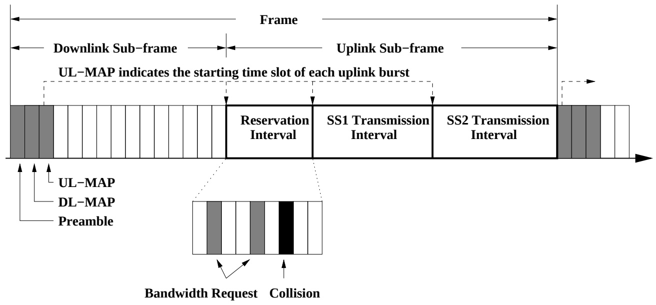


Fig. 1. MAC frame structure.

B. QoS Architecture

The IEEE 802.16 supports many traffic types (data, voice, video) with different QoS requirements. The standard defines four types of data flows, each one associated with distinct applications and QoS requirements [3]:

1) *Unsolicited Grant Service* (UGS): for applications with constant bandwidth allocation requirements.
2) *Real-Time Polling Service* (rtPS): for applications with specific bandwith requirements and maximum acceptable delay.
3) *Non-Real-Time Polling Service* (nrtPS): for applications with a minimum bandwidth allocation requirements, that are intolerant to delay.
4) *Best Effort Service* (BE): for applications without bandwidth allocation requirements, that receive the remaining bandwidth after the allocation to the three previous types of services.

Figure 2 shows the QoS architecture present in 802.16. The UL packet scheduling (UPS) module controls all the packet transmissions in the UL. As the protocol is connection-oriented, the application should establish a connection between the BS and the associated service flow (UGS, rtPS, nrtPS or BE). The BS identifies the connections by assigning a unique Connection ID (CID) to each one. The 802.16 defines the signaling process for the establishment of a connection (Connection-Request and Connection-Response) between SS and BS, but does not specify the rules for admission control.

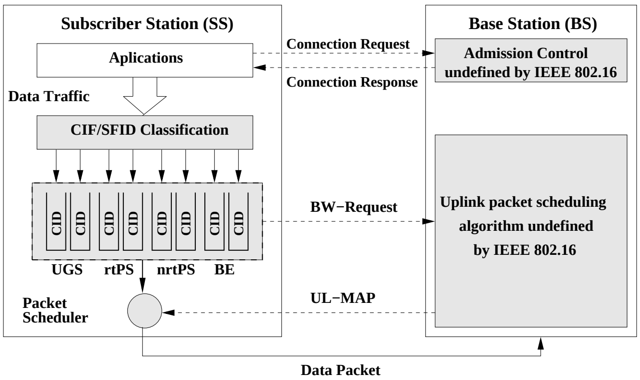


Fig. 2. QoS architecture of IEEE 802.16.

In summary, the IEEE 802.16 specifies: the signaling mechanism for information exchange between the BS and the SSs, as the connection configuration, BW-Request and UL-MAP; and the scheduling of UL for UGS traffic. The standard does not define: the scheduling of UL for rtPS, nrtPS and BE services; admission control and traffic policing.

III. Related Work

In [1, 2], the authors proposed a new MAC protocol for IEEE 802.16 that uses an access scheme called RPAC (Reservation-Priority Access Control), described in [4]. This access scheme incorporate a traffic scheduling mechanism based on messages and/or stations priorities, and a reservation period governed by a TDMA discipline with one time slot allocated per station in the network. The difference between the protocols described in [1] and [4] is that, in the former, the access scheme was reformulated and adapted to 802.16 MAC structure. The MAC frame structure for the proposed protocol in [1] is illustrated in Figure 3.

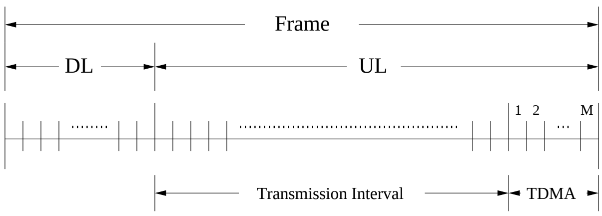


Fig. 3. MAC frame structure for the proposed protocol (from [1]).

The authors assumed that, in the UL sub-frame, the channel is allocated to the stations following the sequence $1, 2, 3, ..., M$. Messages arriving at each station, belong to one of the different P classes. Two versions of the protocol were considered, according to the priority rules used to determine the order in which messages should be transmitted during the transmission period:

Version I: for any $p, q \in \{1, ..., P\}$ such that $p < q$, all the class-p messages are transmitted before any messages of class q, independent of which station it belongs to. For messages belonging to the same class but in distinct stations, the order of the transmissions is according to the order in which the stations access the channel (first station 1 and last station M). For messages in the same station with the same class of priorities, the transmissions occur by order of arrival. The behavior of the channel according to this version is illustrated in Figure 4.

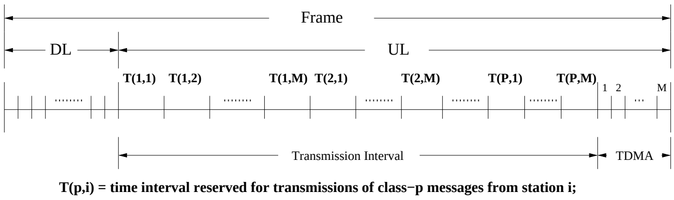


Fig. 4. Version I of proposed protocol.

Version II: for any $i, j \in \{1, ..., M\}$ such that $i < j$, all the messages in station i are transmitted before any message in station j, independently of it's priority class. In any terminal, the messages are transmitted in accordance with their priority classes and in order of arrival, in the case of belonging to the same class; that is, at each station, the priority discipline HOL (Head-Of-the-Line) is applied with the highest priority assigned to class 1 and the lowest assigned to class P. The behavior of the channel according to this version is illustrated in Figure 5.

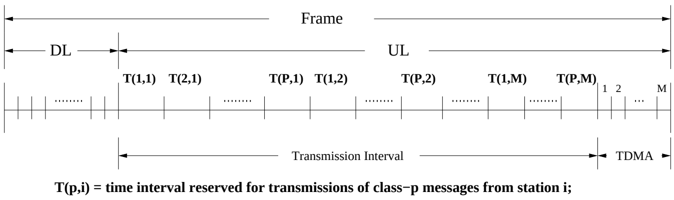


Fig. 5. Version II of proposed protocol.

The Version II can be seen as an exhaustive system, where all available packets of a station are transmitted during its corresponding data interval, independent of the class. However, as shown in [1], there is a greater degree of "unfairness" in the access to the media amongst network terminals, under the assumption of fixed terminal-priorities.

In this paper, we'll present an extention to Version II, by assigning variable priorities to the network stations sharing access to the communication channel. With this, we hope to improve the channel utilization, since the difference depends only on the traffic classes and all stations have the same average message waiting-time in the queue.

Our extention is based in the work described in [5], where the author presents a generic performance evaluation of any access scheme that can be seen as an alternating sequence of transmission and scheduling intervals, as illustrated by Figure 6. During the scheduling time, the stations that have messages to transmit are allocated in the next transmition interval.

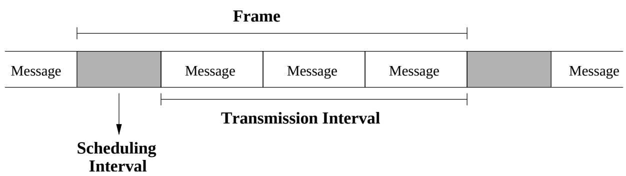


Fig. 6. Access scheme used by Bux [5].

With Fixed priorities the order of service among the stations is fixed, i.e., service is first given to all packets in the higher priority station and so on. On the other hand, with the Variable scheme, priorities of the stations are systematically varied according to certain rules to avoid or, at least, reduce the inherent "unfairness" of the fixed-priority discipline.

Two different types of variable priorities were considered in [5]: cyclic priorities and complementary priorities. However, in the analysis presented in [5], it was considered only one type of message (one traffic class), and the packet length is equal for all stations. Here, we extended the two scheduling schemes based on variable priorities of [5] for several traffic classes, and variable packet length, as can be seen in the next section.

IV. Analytic Model

To facilitate the understanding, we present first the expressions for Version I and II with fixed priorities scheme described in [1]. In the sequence, we outline the analysis involving variable priorities for Version II, which is the focus of this work. Due to space limitations, the reader should look at the above mentioned reference for additional and more specific details in the derivation of analytical models.

As defined in [1], the system considered has one BS and M ($M \geq 1$) client stations (SSs), each of which has an infinite buffer-size and is already associated with the base station. The transmission channel is assumed to be error-free, with a transmission rate equal to C bit/s. In general, the messages generated at each station are composed of a random number of fixed units of data, called packets, each of which contains μ^{-1} bits. The transmission time of each packet is made equal to a time slot (τ). Therefore, $\tau = (\mu C)^{-1}$.

Messages arriving at each station, belong to one of the different P classes and we assume that class-1 messages have the highest priority and class-P messages the lowest. Moreover, the arrival of messages is characterized by a Poisson point process, such that λ_i^p (messages per slot) is the average arrival rate of class-p messages to station i.

A. Fixed Priorities

1) Version I: As the protocol defined in [1] uses the reservation scheme of [4], the results from the later paper were used to obtain the expression for the steady-state average waiting-time of class-p messages at station i, in the former one:

$$\overline{W}_i^p = M + E[DL] + [\frac{(1+\rho_i^p)}{2} + \sum_{j=1}^{p-1}\sum_{g=1}^{M}\rho_g^j + \sum_{j=1}^{i-1}\rho_j^p]\frac{E[L^2]}{E[L]} - \frac{1}{2}, \tag{1}$$

where $\rho_i^p = \lambda_i^p b_i^p$ is the traffic in the terminal i due to the messages of class p, and $E[DL] = \lim_{n\to\infty} E[L_n^{DL}]$. This expression is still given as a function of $E[L]$ and $E[L^2]$, respectively, the first and second steady-state moments of the cycle length:

$$E[L] = \frac{M + E[DL]}{1-\rho};\ \rho = \sum_{p=1}^{P}\sum_{i=1}^{M}\rho_i^p < 1; \tag{2}$$

and

$$\begin{aligned} E[L^2] &= \frac{1}{1-\sum_{p=1}^{P}\sum_{i=1}^{M}(\rho_i^p)^2}\Big\{M^2 + E[DL]^2 + 2ME[DL] \\ &+ \Big[2\rho(M+E[DL]) + \sum_{p=1}^{P}\sum_{i=1}^{M}\lambda_i^p b_{2,i}^p\Big]E[L] \\ &+ \Big[\sum_{p=1}^{P}\sum_{i=1}^{M}\sum_{\substack{j=1\\ j\neq i}}^{M}\rho_i^p\rho_j^p + \sum_{p=1}^{P}\sum_{\substack{q=1\\ q\neq p}}^{P}\sum_{i=1}^{M}\sum_{k=1}^{M}\rho_i^p\rho_k^q\Big]E^2[L]\Big\}. \end{aligned} \tag{3}$$

2) Version II: The analysis for Version II follows in a direct manner noting that, the messages are transmitted in the same order as in Version I, with the classes of the messages exchanged for the numbers of the stations and vice-versa (see Section III). Therefore, the expression in the Version II is analogous to that in Version I, changing only the M by the P and the i for the p:

$$\overline{W}_i^p = M + E[DL] + [\frac{(1+\rho_i^p)}{2} + \sum_{j=1}^{i-1}\sum_{k=1}^{P}\rho_j^k + \sum_{j=1}^{p-1}\rho_j^p]\frac{E[L^2]}{E[L]} - \frac{1}{2}, \tag{4}$$

with $E[L]$ and $E[L^2]$ given by equations (2) and (3), respectively.

For fixed priorities, however, the Version II of proposed protocol provides a lower degree of "fairness" in the channel access amongst stations in the network, as will be seen in the next section. To reduce this problem, we consider two variable priorities strategies to periodically alter priorities among stations: cyclic and complementary priorities. These schemes were based on scheduling mechanism proposed in [5], including priorities for several traffic classes and assuming packets with variable length.

B. Variable Priorities

1) Version II with Cyclic Priorities: This discipline defines that, a station with priority p ($p \in 1, 2, ..., M$) in one frame, assumes next-lower priority ($p+1$) in the next. The station with lowest priority M obtains highest priority 1 in the next frame.

We denote $\overline{W}_i^p(x)$ as the average waiting-time in the queue of class-p messages at station i, given that this station has priority x. Moreover, we assume that $F_i^p(x)$ is the probability of a class-p message in the station i has been transmitted during the frame were station i has priority x. So, we have:

$$\overline{W}_i^p = \sum_{i=1}^{M}\overline{W}_i^p(x)\cdot F_i^p(x). \tag{5}$$

We note that, $F_i^p(x)$ is the probability that a message arrives during the frame in that the station i has priority $i-1$ (or priority M, in the case of $i = 1$). Since, the frames are i.i.d. (independent of the priority discipline used) [5] and assuming Poisson arrivals, we have:

$$F_i^p(1) = F_i^p(2) = ... = F_i^p(M) = \frac{1}{M}. \tag{6}$$

Using the result obtained from equation (4), follows $\overline{W}_i^p(x)$ through equation (7):

$$\begin{aligned} \overline{W}_i^p(x) &= M + E[DL] + (1+\rho_i^p)\frac{E[L^2]}{2E[L]} + \\ &+ \Big\{\sum_{j=1}^{x-1}\sum_{k=1}^{P}\rho_{1+(i-x+j-1)mod(M)}^k + \sum_{j=1}^{p-1}\rho_i^j\Big\}\frac{E[L^2]}{E[L]} - \frac{1}{2}. \end{aligned} \tag{7}$$

Substituting equations (6) and (7) in equation (5), we have the steady-state average waiting-time of the class-p messages at station i given by equation (8):

$$\begin{aligned} \overline{W}_i^p &= M + E[DL] + (1+\rho_i^p)\frac{E[L^2]}{2E[L]} + \Big\{\sum_{j=1}^{p-1}\rho_i^j + \\ &+ \frac{1}{M}\sum_{j=2}^{M}(M-j+1)\sum_{k=1}^{P}\rho_{1+(i-j)mod(M)}^k\Big\}\frac{E[L^2]}{E[L]} - \frac{1}{2}. \end{aligned} \tag{8}$$

2) Version II with Complementary Priorities: This discipline defines that, a station i assumes priority x in one frame and priority $(M+1-x)$ in the next one. Then it returns to priority x, and so on. Using the same analysis applied in the cyclic priority, we have that, for complementary priorities, the average waiting-time is equal to all stations given by equation (9):

$$\overline{W}_i^p = M + E[DL] + (1+\rho_i^p)\frac{E[L^2]}{2E[L]} + \Big\{\sum_{k=1}^{p-1}\rho_i^k + \frac{1}{2}\sum_{\substack{j=1\\ j\neq i}}^{M}\sum_{q=1}^{P}\rho_j^q\Big\}\frac{E[L^2]}{E[L]} - \frac{1}{2}. \quad (9)$$

Let us address now the particular situation in which all messages, at any given terminal, belong to the same class and has a fixed length. This is equivalent to make $P = 1$ and $\rho_i = \lambda_i b$ in our general model. Under this assumptions, the results for $\overline{W}_i^p$ are independent of p and identical of those found in [5], illustrated by equations 10 and 11 to Cyclic and Complementary priorities, respectively. So, we extended the results derived in those paper to a more general scheduling scheme, involving priorities based on stations and traffic classes.

$$\overline{W}_i = E[DL] + \Big\{\frac{1}{M}\sum_{j=2}^{M}(M-j+1)\rho_{1+(i-j)mod(M)} + \frac{(1+\rho_i)}{2}\Big\}\frac{E[L^2]}{E[L]}; \quad (10)$$

$$\overline{W}_i = E[DL] + \Big\{\frac{1+\rho}{2}\Big\}\frac{E[L^2]}{E[L]}; \quad i = 1,2,...,M. \quad (11)$$

Through the analytical models presented here, some numeric results will be presented in the next section, to illustrate the performance of the proposed protocols.

V. Numeric Results

As described in [1], we considered two distinct scenarios to evaluate the level of differentiation obtained with the described protocols. In each scenario there is a differentiated probability between four types of traffic classes ($P = 4$), as shown in Table I. In Scenario I there is a greater probability for the classes of high priority. On the other hand, in the Scenario II the classes of low priority prevail over those of high priority. Thus, it is possible to compare which is the influence of a greater load of flows of lower priority over those of higher priority and vice-versa.

TABLE I
Traffic scenarios used.

Traffic Classes	Scenario I	Scenario II
Class 1	40%	10%
Class 2	30%	20%
Class 3	20%	30%
Class 4	10%	40%

We assume that the number of packets in each message of class p in the station i follows a geometric distribution with average $b_i^p = 5$ and $b_{2,i}^p = 45$, for each $p = 1,2,3,4$; and $i = 1,...,10$. Moreover, in each scenario there are 10 stations ($M = 10$) with a balanced traffic between these (that is $\lambda_i^p = \lambda^p/10$, where λ^p represents the rates of messages from class p). In this way, the average waiting-time in the queue for the station i, where λ_i represents the message rate in the station i, is given by equation:

$$\overline{W}_i = \sum_{p=1}^{4}\frac{\lambda_i^p}{\lambda_i}\overline{W}_i^p.$$

Due to space limitations, only the results from Version II are present in this paper. At first we illustrate the results presented in [1], obtained with fixed priorities scheme. Following, we expose the results obtained with the proposed schemes of variable priorities, which are the focus of this work. Finally, we compare the analytical model described in [1] with a simulation model proposed here.

A. Fixed Priorities

Figures 7(a) and (b) present the average waiting-time in the queue (in function of terminals) of Version II with fixed priorities in Scenario I and II, respectively. Three values of traffic intensity ($\rho = 0.3, 0.6$ *and* 0.9) were used to represent low, medium and heavy load.

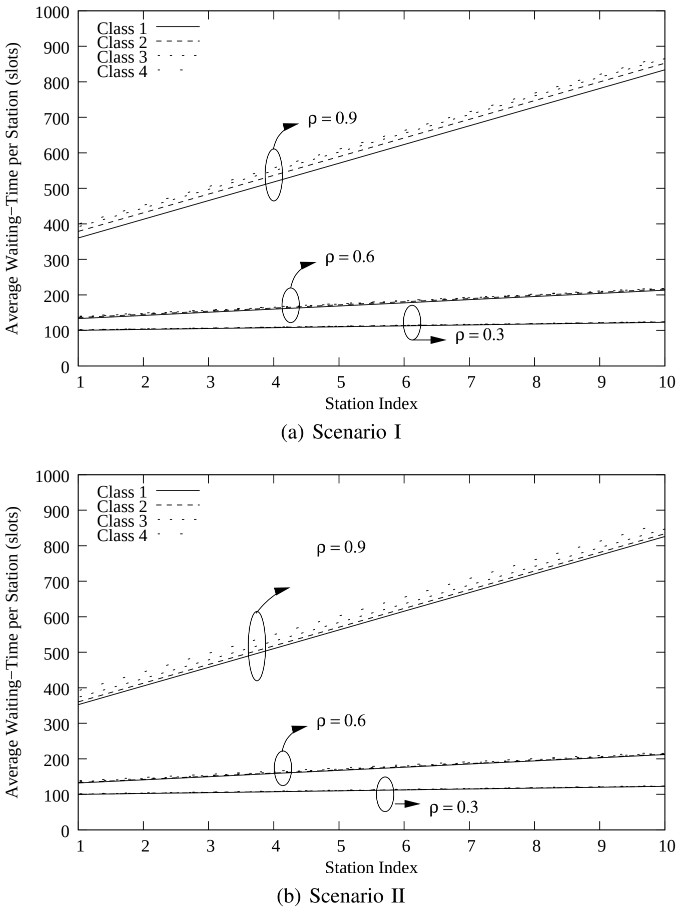


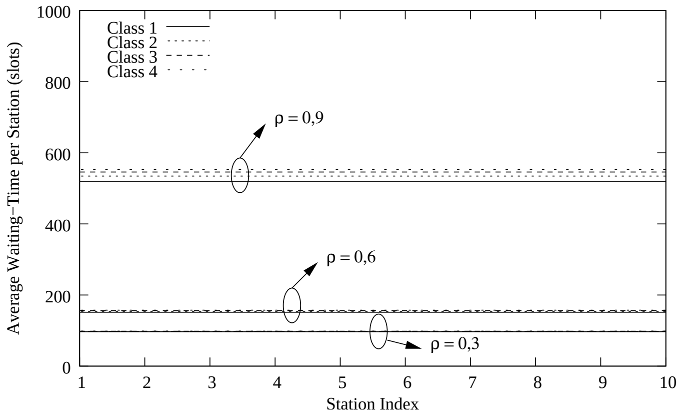
Fig. 7. $\overline{W}_i$ in Version II: Scenario I (a) and Scenario II (b) (from [1]).

From these figures, it can be observed that Version II promotes a differentiation in the channel access between the stations in such a way that, increases the average waiting-time in the queue for the messages of the station with a low priority. In respect to the change between the traffic scenarios, this version behaves in a similar way for the Scenario I and II. Furthermore, it can be perceived that there is a discrete differentiation among the traffic classes, guaranteeing a smaller waiting-time in the queue for the messages of greater priority, even in the Scenario II where there is a greater probability of the low-priority traffic.

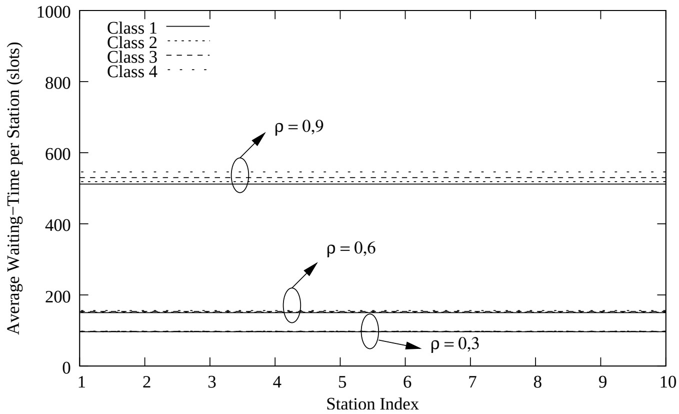
B. Variable Priorities

As can be seen by Figure 7, the Version II of the considered protocol promoves a higher degree of unfairness in channel access among stations. This is due to the fact that the second version utilizes a fixed priority discipline between the stations in the network. To solve this problem, the two variable priority schemes described in previous sections were used.

Figures 8 and 9 present the average waiting-time in the queue as a function of the station index for the cyclic and complementary disciplines, respectively. Figures 8(a) and 9(a) illustrate the values obtained under Scenario I and Figures 8(b) and 9(b) illustrate the values obtained under Scenario II. We can notice that, as expected, the behavior of the two disciplines differs from the one exposed in Figures 7(a) and (b) for the fixed priority scheme.



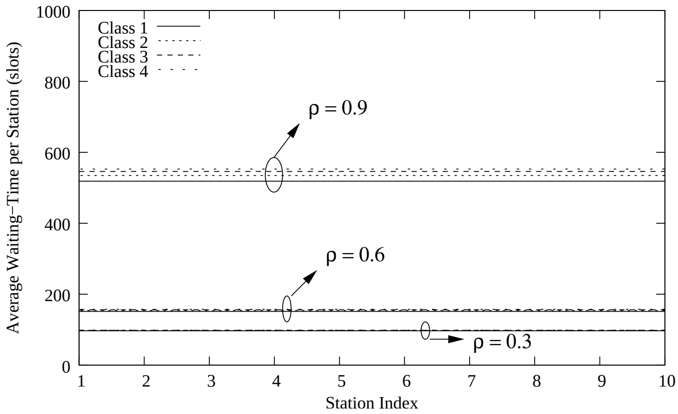
(a) Scenario I



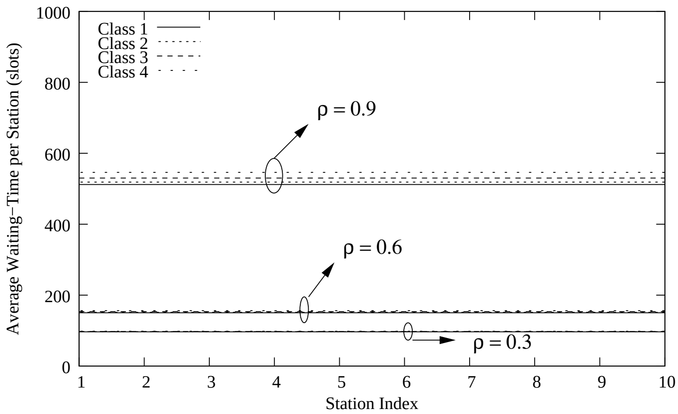
(b) Scenario II

Fig. 8. $\overline{W}_i$ with cyclic priorities: Scenario I (a) and Scenario II (b).

Both variable priority schemes eliminate the unfairness problem in the channel access among stations. Moreover, we observe that the two disciplines have similar behavior, as the traffic is balanced between the stations. According to [5], different results are expected for an unbalanced traffic. So, it is difficult to indicate a better scheme without a previous analysis involving unbalanced traffic. But it is clear that, using variable priorities with cyclic and complementary disciplines we obtained a greater degree of fairness in channel access between stations in the network in the Version II of the proposed protocol. With this, we have a complete set of solutions that can be applied in various practical scenarios of 802.16 networks.



(a) Scenario I



(b) Scenario II

Fig. 9. $\overline{W}_i$ with complementary priorities: Scenario I (a) and Scenario II (b).

C. Simulation

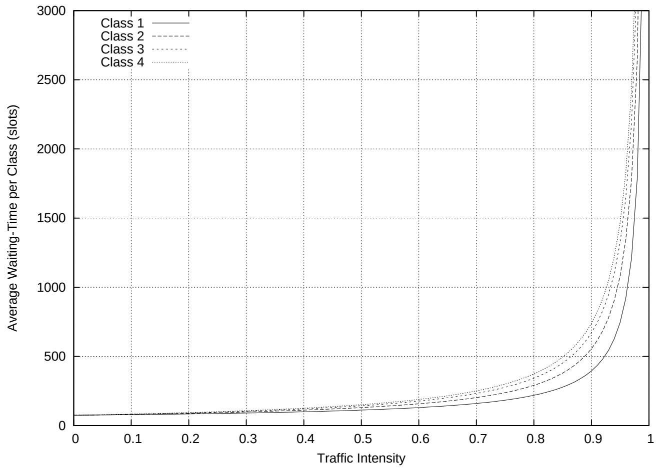
As mentioned in [2], the behavior of the UL sub-channel activity can be seen as a M/G/1 queue with vacation and priorities [6]. The UL sub-channel is the server of the system, where was assumed Poisson arrivals with class and/or station priorities. The service time depends of probability distribution used to represent message lengths, and the vacation intervals are due DL sub-channel. During this period, the system server (UL sub-channel) doesn't serve traffic classes.

It was developed a model for Version I of the considered protocol with four Poisson sources and one server, using a modeling and simulation tool called Tangram2 [7]. We attributed twenty values for traffic intensity: ρ = 0.05, 0.10, 0.15, ..., 0.95 and 0.99. And, for each value of traffic intensity, 50 simulations of 100,000 seconds were carried out, with a confidence interval of 95%. In the total, 1,000 simulations were carried out.

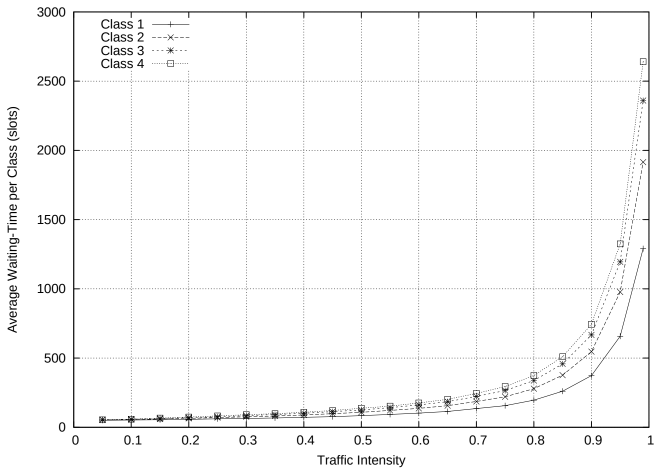
Figures 10(a) and (b) compare the analytical results with the simulated results. These graphs were generated with the same parameters described in the beginning of this section under Scenario I. To plot the curves, we used the average waiting-time for the class p, that is given by equation:

$$\overline{W}^p = \sum_{i=1}^{10} \frac{\lambda_i^p}{\lambda^p} \overline{W}_i^p .$$

By figures, we notice the similarity between the models and a more evident differentiation for a heavy traffic in the channel (with the raise of traffic intensity, the waiting time in the queue grows for



(a) Analytical Results



(b) Simulation Results

Fig. 10. $\overline{W}^p$ in Version I: analytical results (a) and simulation results (b).

all classes). With this, it can be perceived that the protocols manage to differentiate efficiently the traffic classes, guaranteeing a smaller waiting-time in the queue for the messages of greater priority.

We intend to use the analytical model described here to evaluate the behavior of the proposed protocols with traffic models to represent voice, video, and data traffic, more suitable than Poisson model.

VI. Conclusions

In this paper we proposed an extension of the MAC protocol described in [1, 2], in which a variable priority scheme is introduced to the Version II of the considered protocol. Based on [5], two priority disciplines were incorporated in the traffic scheduling mechanism: cyclic and complementary priorities. Moreover, we presented an analytical model for the average waiting-time of the messages in the queue under all priority disciplines. According to the results exposed in this work, we can conclude that the variable priority schemes provide an improvement of the channel utilization among stations, since the average message waiting-time in the queue is the same for all stations, and the difference depends only on the traffic classes.

As future work, we intend to include into the protocols an admission control mechanism, so that the overload of a specific type of traffic doesn't affect the response time of the others. Moreover, we are going to evaluate the behavior of the proposed protocols with other traffic models, more suitable to represent multimedia applications than Poisson model (such as Weibull and Pareto), and under unbalanced traffic conditions.

Acknowledgment

This work was supported by FINEP, through project grant number 1956/04.

References

[1] L. F. M. de Moraes and P. D. Maciel Jr., "Analysis and Evaluation of a New MAC Protocol for Broadband Wireless Access," in *IEEE WirelessCom 2005 - International Conference on Wireless Networks, Communications, and Mobile Computing*, Maui, Hawaii, June 2005.

[2] ——, "Proposta e Avaliação de um Novo Protocolo de Acesso ao Meio para Broadband Wireless Acess," in *XXII Brazilian Symposium of Telecommunications - SBrT'05*, Campinas, SP - Brasil, Sept. 2005.

[3] IEEE 802.16, *IEEE Standard for Local and Metropolitan Area Networks - Part 16: Air Interface for Fixed Broadband Wireless Access Systems*, IEEE Std. 802.16, Oct. 2004.

[4] L. F. M. de Moraes and A. N. L. Valverde, "Waiting-Time Analysis of a Reservation Access-Control Scheme with message-based priorities," *Third IFIP TC 6/wg 7.3 Inter. Conf. on Data Communication Systems and Their Performance*, June 1987.

[5] W. Bux, "Analysis of a Local-Area Bus System with Controlled Access," *International Conference on Performance of Data Communication Systems and their Applications*, pp. 11–22, Sept. 1981.

[6] L. Kleinrock, *Queueing Systems, Vol. 2: Computer Applications*. New York, NY: Addison-Wesley, 1976.

[7] Tangram-II - LAND/PESC/COPPE/UFRJ. [Online]. Available: http://www.land.ufrj.br/tools/tangram2/tangram2.html