Wireless Networks 7, 55–63, 2001

A Pseudo-Bayesian ALOHA Algorithm with Mixed Priorities *

JEAN-FRANÇOIS FRIGON ** and VICTOR C.M. LEUNG
Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, BC, Canada, V6T 1Z4

Abstract. In reservation-based multiple access protocols, before obtaining a contention-free access to the channel, a mobile terminal must wait for its request packet to be successfully sent to the base station. A pseudo-Bayesian ALOHA algorithm with multiple priorities is proposed in this paper to reduce the waiting time of delay sensitive request packets in a multimedia environment. Packets are transmitted in each slot according to a transmission probability based on the channel history and a priority parameter assigned to their respective priority class. An adaptation of the slotted protocol to the framed environment proposed for wireless ATM is also described. Simulation results are presented to show that the protocol offers a significant delay improvement for high priority packets with both Poisson and self-similar traffic while low priority packets only experience a slight performance degradation.

Keywords: multiple access methods

1. Introduction

In reservation-based medium access control (MAC) protocols, a mobile terminal sends a request to the base station to obtain a contention-free access to the wireless channel. However, in most of these systems, the request packet is sent in contention with request packets from other connections. Thus, the most critical factor affecting the delay-related quality of service (QoS) in a reservation MAC protocol is the contention phase that a connection goes through before it is allowed contention-free packet transmissions.

For certain classes of traffic the contention phase is the limiting factor in providing a delay-QoS guarantee, while other traffic classes are less sensitive to the introduced delay. The first packet of a voice talkspurt, a request for new bit-rate in a real-time variable bit rate connection, or a handoff request for a real-time connection are examples of control packets sensitive to contention delay. On the other hand, control packets for new data messages or request to establish a new connection are less time sensitive. It is, thus, necessary to find a contention access protocol that gives priority to delay sensitive request packets in order to improve the performance of multimedia wireless MAC protocols such as those proposed to support wireless ATM.

Several contention protocols with mixed priorities have been presented in the literature. Some protocols avoid collisions between high and low priority packets [3,7,10,11]. They postpone low priority packet transmissions until they detect, through channel feedback, that there is no more high priority packet in the system. Other protocols allow collisions between high and low priority packets [9,12,13,17] and service priority is given by the collision resolution protocol. However, these priority protocols work on a slot-by-slot basis and require an immediate feedback after each contention slot. This prevents the application of these schemes in reservation protocols employing a frame structure similar to the one illustrated by figure 1, as proposed for some wireless ATM systems [1,5,6,15]. In this structure, feedback for the uplink transmissions is available only at the beginning or end of each frame. A straightforward modification that can be implemented is to divide the terminals into N different groups, one accessing each uplink control slot. Then, feedback information sent in the downlink control slots can be used to update each independent group. Even if this modification will maintain the maximum global throughput, the separation between groups is not desirable since it will cause longer delays (for the same reason that N servers of capacity C/N with distinct queues produce a longer delay than a single server of capacity C).

Thus, there is a need to find a new random access protocol with mixed priorities that can be specifically adapted to the frame structure of many reservation protocols. Since these protocols are centrally controlled, information about the actual state of the multiple access algorithm can be centrally maintained, allowing an easy adaptation by the algorithm to the sensed channel state. On the other hand, a deterministic algorithm cannot be used since we cannot make any assumption about the active user population size. Finally, we consider that in reservation protocols the overall throughput is not determined by the random access throughput, but the QoS is highly dependent on the delay encountered by request packets. Hence, our design criteria put more emphasis on the access delay than on the throughput. Thus, we want to

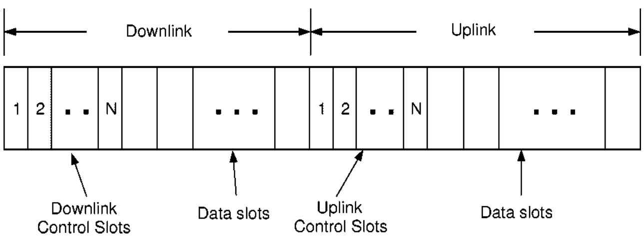


Figure 1. Multimedia wireless MAC protocol frame structure.

* This work was supported by the Natural Sciences and Engineering Research Council of Canada through a Postgraduate Scholarship and Grant OGP0044286.

** Now with the Department of Electrical Engineering, University of California, Los Angeles, CA, USA.

have low and fairly constant delays for the high priority traffic over a relatively wide range of total traffic rates, without introducing excessive delays for low-priority traffic.

To develop this new protocol, we have chosen to explore the possibilities offered by the slotted ALOHA protocol. The CSMA and CSMA/CD protocols [2] could also have been considered to implement the desired random access protocol with priorities. However, the ALOHA protocol has the advantage of simplicity of operation for the mobile terminals and is not affected by the hidden terminal problem. Furthermore, if each mobile is allowed only one control packet transmission in each frame as is usually the case for reservation protocols, the CSMA and CSMA/CD protocols cannot take advantage of their sensing capabilities to increase the throughput.

The paper is organized as follows. In the next section, we propose a slotted ALOHA protocol with mixed priorities. A modification to adapt the protocol to the framed transmission environment is then introduced in section 3. Finally, simulation results are presented in section 4.

2. Pseudo-Bayesian slotted ALOHA with priorities

It is known that the basic slotted ALOHA algorithm, which allows a node to transmit new packets in the next slot when it receives them, and to retransmit backlogged packets with a fixed probability q_r in subsequent slots, is unstable for any value of the arrival rate. Thus, to implement a stabilized slotted ALOHA with priority classes, we derive an algorithm similar to the pseudo-Bayesian ALOHA stabilization algorithm presented in [2,16].

Let new packets be regarded as backlogged immediately after their arrivals at the respective mobile terminals. They will attempt transmission in subsequent slots until success with a probability determined by their priority class and the estimated backlogged state of the system. Consider dividing the traffic sources into p priority classes. The cumulative input arrival process consists of p independent Poisson processes with intensities $\lambda_1, \ldots, \lambda_p$. Let $n_1, \ldots, n_p$ and $q_1, \ldots, q_p$, respectively, be the number of backlogged packets and the transmission probability of each traffic class. Then, the channel traffic generated by class i is $G_i(n_i) = n_i q_i$ and the total attempt rate is $G(n_1, \ldots, n_p) = \sum_{i=1}^{p} n_i q_i$. The probability that a packet of the ith traffic class is successfully transmitted in a slot is then given by

$$P_{\text{succ}}^{i} \approx G_i(n_i)\mathrm{e}^{-G(n_1,\ldots,n_p)}, \tag{1}$$

and the probability that a packet from any class is successfully transmitted is

$$P_{\text{succ}} \approx G(n_1, \ldots, n_p)\mathrm{e}^{-G(n_1,\ldots,n_p)}. \tag{2}$$

We see that if $G(n_1, \ldots, n_p)$ is maintained at the optimal value of 1, the system can achieve its maximum throughput of $1/\mathrm{e}$. The throughput of priority class i is then $G_i(n_i)/\mathrm{e}$. Thus, it is possible to adjust the throughput of each class to a specific value, by adjusting its fraction of the total traffic. Let $\gamma_1, \ldots, \gamma_p$ be the priority parameters of the respective traffic classes and $G_i(n_i) = n_i q_i = \gamma_i$. From here on we will assume without any loss of generality that the traffic classes are ordered in decreasing priority (i.e., traffic class 1 has the highest priority and traffic class p the lowest priority). If we impose the constraint $\sum_{i=1}^{p} \gamma_i = 1$, we obtain the desired maximum throughput $1/\mathrm{e}$ and each traffic class has a throughput γ_i/e.

Now, assume that at the beginning of a slot the number of backlogged packets n_i of priority class i ($1 \leqslant i \leqslant p$) is statistically independent of other priority classes and given by a Poisson distribution with parameter $\hat{n}_i \geqslant \gamma_i$. Furthermore, each packet of class i is independently transmitted in the slot with probability $q_i = \gamma_i/\hat{n}_i$.

Let n_i' denote the number of backlogged packets of class i at the end of a slot, excluding new arrivals. Using Bayes' rule:

$$\begin{aligned} &p(n_1, \ldots, n_p \mid \text{slot event}) \\ &\quad = \frac{p(\text{slot event} \mid n_1, \ldots, n_p)\, p(n_1, \ldots, n_p)}{p(\text{slot event})}, \end{aligned} \tag{3}$$

and since the marginal distribution of n_i given the slot event is given by

$$p(n_i \mid \text{slot event}) = \sum_{\substack{j=1 \\ j\neq i}}^{p} \sum_{n_j=0}^{\infty} p(n_1, \ldots, n_p \mid \text{slot event}), \tag{4}$$

we can find the joint and marginal *a posteriori* distributions of the n_i's for all the slot events (idle, success by a packet from priority class j, or collision). Furthermore, $n_i' = n_i$ when we have either an idle or collision slot. For the case where a packet from priority class j ($1 \leqslant j \leqslant p$) was successfully transmitted, $n_i' = n_i$ for $i \neq j$ and $n_j' = n_j - 1$.

Therefore, we can show that the n_i''s joint and marginal probability distributions, given that the slot was either idle or occupied by the successful transmission of a packet from priority class j, are respectively given by

$$p\big(n_1', \ldots, n_p' \,\big|\, \text{idle or succ}_j\big) = \mathrm{e} \prod_{i=1}^{p} \frac{(\hat{n}_i - \gamma_i)^{n_i'} \mathrm{e}^{-\hat{n}_i}}{n_i'!} \tag{5}$$

and

$$p\big(n_i' \,\big|\, \text{idle or succ}_j\big) = \frac{(\hat{n}_i - \gamma_i)^{n_i'}}{n_i'!} \mathrm{e}^{-(\hat{n}_i - \gamma_i)}. \tag{6}$$

Furthermore, the arrival process is Poisson and independent of the contentions over the channel. Thus, the number of backlogged packets of priority class i, including new arrivals, after either an idle or successful slot, is also independent and Poisson distributed with parameter $\hat{n}_i - \gamma_i + \lambda_i$.

For the case where a collision has occurred in the slot, the n_i''s joint and marginal probability distributions are:

$$p(n_1', \ldots, n_p' \mid \text{coll}) = \frac{\mathrm{e}}{\mathrm{e}-2} \prod_{i=1}^{p} \frac{\mathrm{e}^{-\hat{n}_i}}{n_i'!} \left(\prod_{i=1}^{p} \hat{n}_i^{n_i'} - \prod_{i=1}^{p} (\hat{n}_i - \gamma_i)^{n_i'} - \sum_{i=1}^{p} \left[n_i' \gamma_i (\hat{n}_i - \gamma_i)^{n_i'-1} \prod_{\substack{j=1 \\ j \neq i}}^{p} (\hat{n}_j - \gamma_j)^{n_j'} \right] \right), \quad (7)$$

$$p(n_i' \mid \text{coll}) = \frac{\mathrm{e}}{\mathrm{e}-2} \frac{\mathrm{e}^{-\hat{n}_i}}{n_i'!} \Big[\hat{n}_i^{n_i'} - \mathrm{e}^{\gamma_i - 1} \big[(2-\gamma_i)(\hat{n}_i - \gamma_i)^{n_i'} + n_i' \gamma_i (\hat{n}_i - \gamma_i)^{n_i'-1} \big] \Big]. \quad (8)$$

We can clearly see that after a collision slot, the numbers of backlogged packets in the p priority classes are neither independent nor Poisson distributed. However, we can compute the mean and variance of the obtained marginal distribution and compare them with the mean and variance of a random variable X that is Poisson distributed with parameter $\hat{n}_i + \gamma_i/(\mathrm{e}-2)$. We then find:

$$\begin{aligned} E[n_i' \mid \text{coll}] &= \hat{n}_i + \frac{\gamma_i}{\mathrm{e}-2}, \\ E[x] &= \hat{n}_i + \frac{\gamma_i}{\mathrm{e}-2}, \\ \mathrm{Var}[n_i' \mid \text{coll}] &= \hat{n}_i + \frac{\gamma_i}{\mathrm{e}-2} - \left(\frac{\gamma_i}{\mathrm{e}-2}\right)^2, \\ \mathrm{Var}[x] &= \hat{n}_i + \frac{\gamma_i}{\mathrm{e}-2}. \end{aligned} \quad (9)$$

Thus, we see that even if the distribution of n_i', given that there was a collision, is not quite Poisson, its mean and variance are similar to the Poisson distribution with parameter $\hat{n}_i + \gamma_i/(\mathrm{e}-2)$. Furthermore, we know that $\gamma_i \leqslant 1$ and that $\hat{n}_i$ is likely to be large when a collision occurs (and stabilization will be more necessary when the number of backlogged packets increase). Under these conditions, we see that the variance becomes almost equal to the Poisson distribution. The similarity of the two distributions can also be observed through plotting of their respective probability density function. For these reasons, the distribution of the number of backlogged packets for priority class i, after a collision slot, including new arrivals, is reasonably approximated by a Poisson distribution with parameter $\hat{n}_i + \gamma_i/(\mathrm{e}-2) + \lambda_i$.

It can be shown that the correlation between the numbers of backlogged packets of two different priority classes i and j ($i \neq j$) is given by

$$\mathrm{Corr}[n_i', n_j' \mid \text{coll}] = \frac{-\gamma_i \gamma_j (\mathrm{e}-2)^{-2}}{\left[\left(\hat{n}_i + \frac{\gamma_i}{\mathrm{e}-2} - \left(\frac{\gamma_i}{\mathrm{e}-2}\right)^2\right) \left(\hat{n}_j + \frac{\gamma_j}{\mathrm{e}-2} - \left(\frac{\gamma_j}{\mathrm{e}-2}\right)^2\right) \right]^{1/2}}. \quad (10)$$

Since $\gamma \leqslant 1$ and $\hat{n}$ is likely to be large when there is a collision, the correlation can be considered negligible. Furthermore, the arrivals for each traffic class are independent. Thus, we can reasonably assume that the n_i''s are independent. Therefore, our initial assumptions on the independence and Poisson distribution of the number of backlogged packets of each traffic class i are satisfied for all three possible slot events, i.e., idle, success, or collision.

2.1. Algorithm

Based on the above results we can derive an algorithm to implement a multiple access protocol with mixed priorities. As before, consider p different priority classes with independent Poisson arrival processes of intensities $\lambda_1, \ldots, \lambda_p$. A lower index corresponds to a higher priority class. Let γ_i be the priority parameter specified for traffic priority class i. To maintain the priority order, we must have $\gamma_1 \geqslant \gamma_2 \geqslant \cdots \geqslant \gamma_{p-1} \geqslant \gamma_p$ and the parameters must satisfy the relation $\sum_{i=1}^{p} \gamma_i = 1$.

The algorithm operates by maintaining for each priority class i an estimate $\hat{n}_i^t$ of the number of backlogged packets n_i^t at the beginning of each slot t. For each priority class i, an effective priority parameter $\hat{\gamma}_i^t$ is also computed (this is needed to avoid the condition that $\gamma_i \geqslant \hat{n}_i^t$). A new arrival during slot t is immediately regarded as backlogged and it will attempt transmission in each subsequent slot after its arrival until success. The transmission probability for each priority class is derived below. While a mobile terminal is backlogged, no other arrival is allowed until the terminal has successfully transmitted its packet.

At the beginning of each slot t, $\hat{n}_i^t$ is updated from $\hat{n}_i^{t-1}$, $\hat{\gamma}_i^{t-1}$ and the slot event feedback for slot $t-1$ according to the rule:

$$\hat{n}_i^t = \begin{cases} \max\left(\lambda_i, \hat{n}_i^{t-1} + \lambda_i - \hat{\gamma}_i^{t-1}\right), & \\ \quad \text{for idle or success,} & \\ \hat{n}_i^{t-1} + \lambda_i + \dfrac{\hat{\gamma}_i^{t-1}}{\mathrm{e}-2}, & \text{for collision.} \end{cases} \quad (11)$$

The priority parameter γ_i is assigned a fixed value when the system is initialized. However, the transmission probability q_i^t of priority class i for slot t, given by $\gamma_i/\hat{n}_i^t$, cannot be greater than one. Therefore, if $\gamma_i > \hat{n}_i^t$ for some priority class i, we set $q_i^t = 1$, and thus, the "effective" value of the priority parameter γ_i ($\gamma_i = \hat{n}_i^t q_i^t = \hat{n}_i^t$) for the throughput equation is no longer equal to its initial optimal value. Furthermore, since $\hat{n}_i^t < \gamma_i$, we have $\sum_{i=1}^{p} \gamma_i < 1$ and the total throughput is lower than its optimal value of $1/\mathrm{e}$. Thus, to maintain the optimal throughput, we should increase the traffic of the remaining priority classes by assigning the difference between the fixed γ_i and $\hat{n}_i^t$ to the other priority classes. We propose a prorating algorithm to dynamically compute, at the beginning of each slot t, for each priority class i, the effective priority parameter $\hat{\gamma}_i^t$ based on the fixed priority parameters γ_i and the estimated numbers of backlogged packets $\hat{n}_i^t$.

The parameters $\hat{\gamma}_i^t$ are set such that $\sum_{i=1}^{p} \hat{\gamma}_i^t = 1$. For each priority class i, the prorating algorithm initially sets its effective priority parameter $\hat{\gamma}_i^t$ to γ_i if $\gamma_i \leqslant \hat{n}_i^t$, or to $\hat{n}_i^t$ otherwise. If $\sum_{i=1}^{p} \hat{\gamma}_i^t < 1$, to maintain the optimum throughput, the "leftover" (i.e., $1-\sum_{i=1}^{p} \hat{\gamma}_i^t$) is added to the effective priority parameter of the highest priority class (i.e., $\hat{\gamma}_1^t$) in order to increase its transmission probability and, therefore, decrease its waiting time. Then, if $\hat{\gamma}_1^t \leqslant \hat{n}_1^t$, the prorating algorithm is stopped; otherwise, $\hat{\gamma}_1^t$ is set to $\hat{n}_1^t$ and the same procedure is repeated for each priority class i ($i > 1$) in order of decreasing priority. After this procedure, if $\sum_{i=1}^{p} \hat{\gamma}_i^t < 1$, the "leftover" is assigned to each priority class in proportion to its packet arrival rate. The prorating algorithm can be summarized as follows:

for each priority class i

$$\hat{\gamma}_i^t = \min\left(\hat{n}_i^t, 1 - \sum_{j=1}^{i-1} \hat{\gamma}_j^t - \sum_{j=i+1}^{p} \min\left(\hat{n}_j^t, \gamma_j\right)\right),$$

$$L = 1 - \sum_{i=1}^{p} \hat{\gamma}_i^t,$$

for each priority class i

$$\hat{\gamma}_i^t = \hat{\gamma}_i^t + \frac{\lambda_i}{\sum_{j=1}^{p} \lambda_j} L.$$

Having determined $\hat{n}_i^t$ and $\hat{\gamma}_i^t$ at the beginning of slot t, each backlogged packet in each priority class i is independently transmitted in slot t according to the transmission probability q_i^t, which is calculated as follows:

$$q_i^t = \min\left(1, \frac{\hat{\gamma}_i^t}{\hat{n}_i^t}\right). \tag{12}$$

3. Framed pseudo-Bayesian ALOHA with priorities

Many reservation protocols employ a frame structure similar to the one presented in figure 1. Therefore, control packets are not sent on a slot basis but on a frame basis. To adapt the slotted algorithm for such reservation protocols' control traffic we propose a strategy whereby a new packet, having arrived at a mobile terminal, waits until the next frame before attempting its first transmission (i.e., a gated system). Starting from the next frame, the terminal will independently attempt to transmit the packet in each frame t with probability q_r^t, in a slot chosen randomly and independently from frame to frame.

Suppose that there are K slots in a frame and the *a priori* distribution of the total number of backlogged packets n^t at the beginning of frame t is Poisson with parameter $\hat{n}^t$. If each backlogged packet independently chooses a slot in the frame for transmission, then the distribution of the number of backlogged packets n_k^t that have chosen a given slot k ($k = 1, \ldots, K$) is Poisson with parameter $\hat{n}^t/K$ and independent of other slots in the frame. We can, therefore, apply the pseudo-Bayesian algorithm presented in the previous section (for the case where there is only one priority class) independently for each slot to update the estimated number of backlogged packets $\hat{n}_k^{t+1}$, based on the feedback on the outcome of each slot. The exact time of the feedback is not important, as long as it is received before the next frame. Thus, n_k^{t+1}, the number of backlogged packets in each slot k after either an idle, success or collision slot, is an independent Poisson random variable with parameter $\hat{n}_k^{t+1}$. Hence, n^{t+1}, the total number of backlogged packets at the beginning of frame $t+1$, is also Poisson with mean $\sum_{k=1}^{K} \hat{n}_k^{t+1}$, which satisfies our initial assumption.

This result can be easily extended to the case where we have p priority traffic classes contending for the slots in the frame. If each of the $\hat{n}_i^t$ backlogged packets of priority class i ($1 \leqslant i \leqslant p$) chooses independently one of the K slots in the frame for transmission with a uniform probability, then at the beginning of frame t, the number of backlogged packets of priority class i for each slot k ($1 \leqslant k \leqslant K$) is independently Poisson distributed with parameter $\hat{n}_i^t/K$. We can then independently apply for each slot k the pseudo-Bayesian rule given by equation (11) to compute the updated estimate of the number of backlogged packets after the slot. Furthermore, the number of backlogged packets of each priority class i in each slot k after either an idle, success or collision slot, is an independent Poisson random variable. Therefore, $\hat{n}_i^{t+1}$, the updated estimate of the total number of backlogged packets of priority class i at the end of frame t, is given by the sum of the updated estimates of all k slots in frame t.

3.1. Algorithm

Using these results, we can derive from the pseudo-Bayesian priority algorithm presented in the previous section a multiple access protocol with mixed priorities for a K-slot frame. The arrival rate λ_i for each priority class i ($1 \leqslant i \leqslant p$) is given in number of packets per slot. The same definitions that were presented in section 2 for γ, $\hat{\gamma}$ and priority order are assumed.

The algorithm operates by maintaining for each priority class i an estimate $\hat{n}_i^t$ of the total number of backlogged packets n_i^t at the beginning of each frame t. A new arrival during frame t is immediately regarded as backlogged and it will attempt transmission in each subsequent frame after its arrival until success. Meanwhile, the respective mobile terminal is blocked from further new arrivals.

At the beginning of each frame t, for each priority class i, $\hat{n}_i^t$ is updated from $\hat{n}_i^{t-1}$, $\hat{\gamma}_i^{t-1}$ and the slot event feedback for frame $t-1$ (let n_{nc} be the number of idle or success slots and n_{c} the number of collision slots in frame $t-1$) according to the rule:

$$\hat{n}_i^t = K\lambda_i + n_{\mathrm{nc}} \max\left(0, \frac{\hat{n}_i^{t-1}}{K} - \hat{\gamma}_i^{t-1}\right) + n_{\mathrm{c}}\left(\frac{\hat{n}_i^{t-1}}{K} + \frac{\hat{\gamma}_i^{t-1}}{\mathrm{e}-2}\right). \tag{13}$$

To maintain the optimum throughput in each slot k in the frame, we apply the prorating algorithm presented in section 2, with the only difference that the estimated number of backlogged packet in a slot is given by $\hat{n}_i^t/K$. We, thus, find the effective priority parameter of each priority class using the following modified prorating algorithm:

for each priority class i

$$\hat{\gamma}_i^t = \min\left(\frac{\hat{n}_i^t}{K}, 1 - \sum_{j=1}^{i-1} \hat{\gamma}_j^t - \sum_{j=i+1}^{p} \min\left(\frac{\hat{n}_j^t}{K}, \gamma_j\right)\right),$$

$$L = 1 - \sum_{i=1}^{p} \hat{\gamma}_i^t,$$

for each priority class i

$$\hat{\gamma}_i^t = \hat{\gamma}_i^t + \frac{\lambda_i}{\sum_{j=1}^{p} \lambda_j} L.$$

Then each backlogged packet of each priority class i is independently transmitted in a randomly selected slot (each slot has a probability $1/K$ of being chosen) in frame t according to the transmission probability q_i^t. The transmission probability for each class i is selected such that the overall access rate in each slot of the frame is maintained at its optimal value. This is achieved using equation (12) to compute the transmission probabilities by replacing $\hat{n}_i^t$ by $\hat{n}_i^t/K$, the estimated number of backlogged packets in each slot of a frame of length K. Thus, the transmission probability in frame t for priority class i is calculated as follows:

$$q_i^t = \min\left(1, \frac{\hat{\gamma}_i^t}{\hat{n}_i^t} K\right). \tag{14}$$

4. Simulation results

In this section we present the simulation results for the slotted and framed pseudo-Bayesian priority algorithms. To evaluate the proposed algorithms, the waiting time statistics are compared with those obtained without access priority between the traffic classes ("Basic PB" protocol). In the simulations, the arrival rate values used by the backlog estimation algorithm are estimates computed from a moving time-average of successful transmissions for each traffic class over a window period of 500 slots. To ensure reliable steady-state statistics, we have run the simulation for a period of at least 10 million slots with each parameter set. The framed scheme is configured with ten slots per frame and the waiting time for this scheme is measured in number of frames, a time unit commonly used in a framed environment. We also assumed an infinite number of terminals such that no arrivals are discarded.

Figure 2 shows, for the framed system, the effect of the priority parameter on the average waiting time when the arrival rates are fixed for both priority classes. As expected from the throughput equations presented in section 2, we observe that a traffic class has a delay advantage when its arrival rate is smaller than its fair share of the total throughput (i.e., $\lambda_i \leqslant \gamma_i \sum_{j=1}^{p} \lambda_j$). An interesting phenomenon can also be observed that the average waiting time experienced by both traffic classes shows a decreasing trend as the priority parameter deviates by an increasing amount from the fair share point. Thus, the optimum operating point is at $\gamma_1 = 1$ when traffic class one is the high priority traffic.

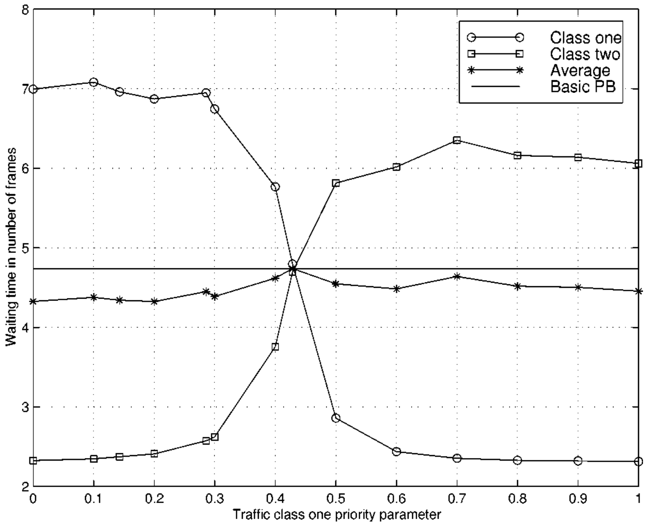


Figure 2. Average waiting time as a function of the priority parameter for the framed system. $\lambda_1 = 0.15$ and $\lambda_2 = 0.20$ packets/slot.

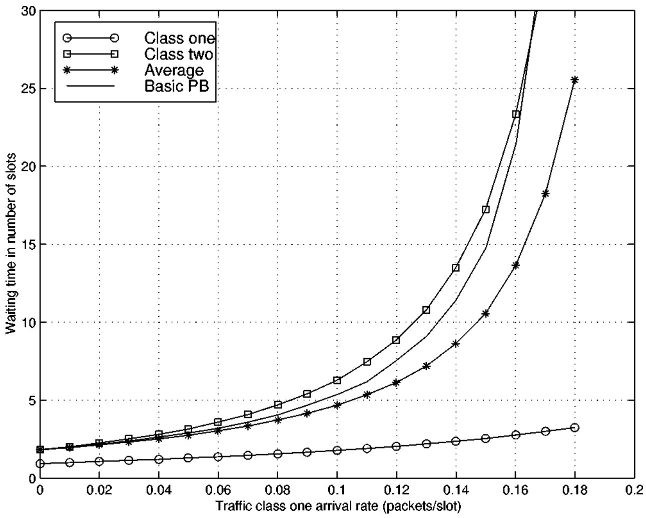


Figure 3. Average waiting time as a function of traffic class one arrival rate for the slotted system. $\lambda_2 = 0.18$ packets/slot and $\gamma_1 = 1$.

The average waiting time as a function of the arrival rate of the high priority traffic class is illustrated in figures 3 and 4. Since, as explained previously, the optimal operating point is at $\gamma_1 = 1$, this value is used to obtain all subsequent results presented in this section. It is apparent from the figures that the average waiting time of the high-priority traffic is relatively constant (in fact, it increases very slowly as the arrival rate of the high priority traffic increases) over a wide range of traffic conditions, even when the overall arrivals are near the maximum total arrival rate (around 0.36 packets/slot) that can be supported by a slotted ALOHA system. Also, the average waiting time of the two priority classes taken together is always lower than that of the reference basic PB protocol, and the low-priority traffic class suffers only a small degradation of its waiting time compared to

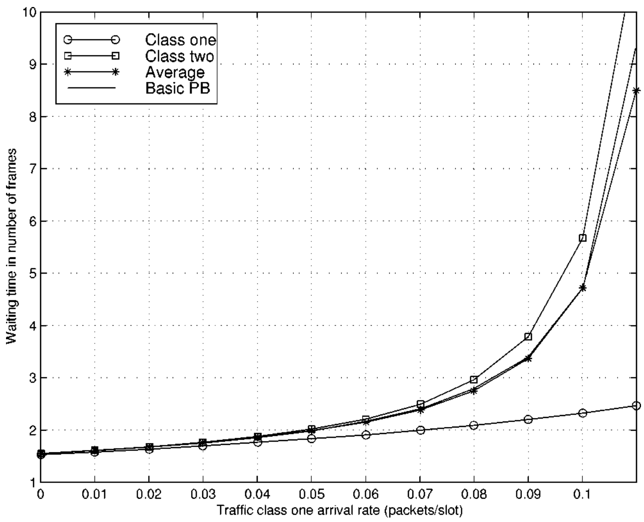


Figure 4. Average waiting time as a function of traffic class one arrival rate for the framed system. $\lambda_2 = 0.25$ packets/slot and $\gamma_1 = 1$.

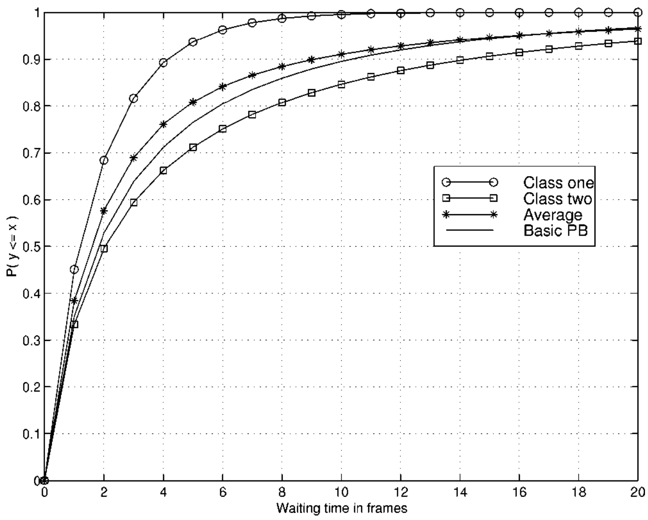


Figure 6. CDF of the waiting time for the framed system. $\lambda_1 = 0.15$ and $\lambda_2 = 0.20$ packets/slot. $\gamma_1 = 1$.

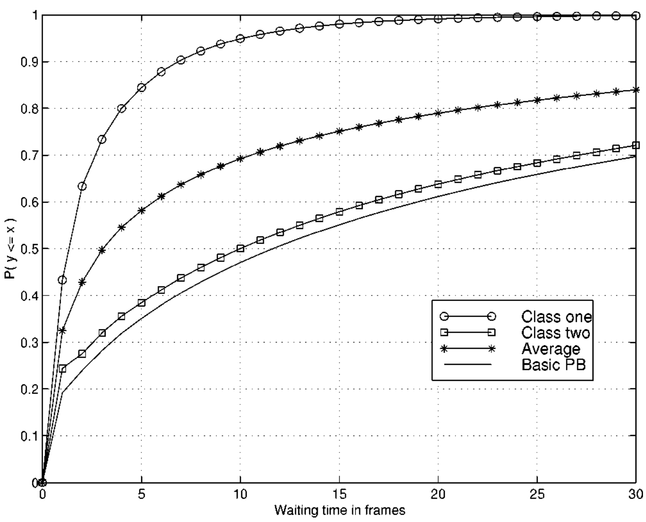


Figure 5. CDF of the waiting time for the slotted system. $\lambda_1 = 0.15$ and $\lambda_2 = 0.20$ packets/slot. $\gamma_1 = 1$.

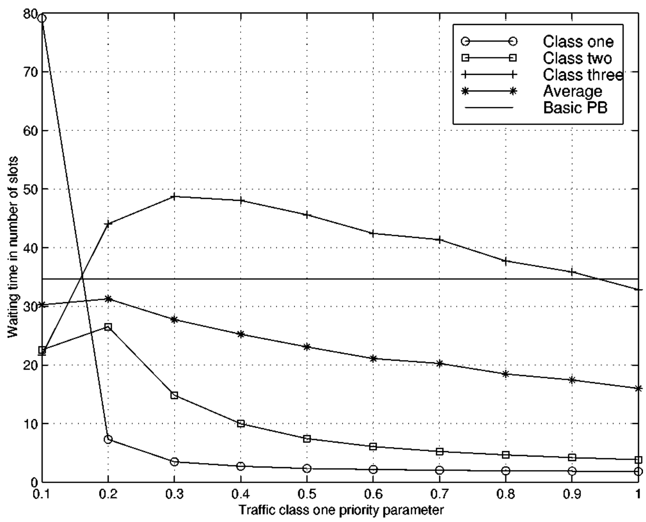


Figure 7. Average waiting time as a function of the traffic class one priority parameter for the slotted system. $\lambda_1 = 0.05$, $\lambda_2 = 0.15$ and $\lambda_3 = 0.15$ packets/slot.

the reference basic PB protocol. Similar results are observed for other traffic conditions [4].

Figures 5 and 6 show the *Cumulative Density Function* (CDF) of waiting time in the slotted and framed systems, respectively. The presented results are for a total arrival rate of 0.35 packets/slot which is almost the saturation point for the basic PB algorithm. For both systems, we can clearly observe the substantial improvements in terms of reduced waiting time for the high-priority traffic class. For example, for the slotted system (figure 5), 70% of the high priority packets are transmitted less than 3 slots after their arrivals, while for the reference algorithm, it takes 30 slots. Furthermore, 90% of the high-priority packets are transmitted less than 7 slots after their arrivals with the priority protocol while for the non-priority algorithm it is completely out of range. For the framed system, the improvement is less spectacular but is still quite interesting. For example, we improve from a situation where 90% of the packets were transmitted less than 10 frames after their arrivals in the reference system to a transmission under 4 frames for priority packets, which represents a decrease of the waiting time of 60%.

4.1. Multiple traffic classes

We have also performed simulations to determine how the algorithm reacts when more than two priority classes contend for the channel. The results presented in this section are for three priority classes but similar trends have been observed for higher number of classes. The prorating algorithm implements the following priority order: class one, class two and then class three. Thus, even if two classes have the same priority parameter, a priority order will be given by the prorating algorithm.

Figure 7 illustrates the impact of the class one priority parameter on the waiting time for the slotted system. The priority parameters of the two other classes were set to $\gamma = (1 - \gamma_1)/2$. Thus, no advantage is given to any of the low priority classes by the priority parameter. Only the prorating algorithm will implement the priority order among the two classes.

We observe that the algorithm is performing as expected with three priority classes. That is, when the priority parameter of traffic class one exceed its share of the total throughput, it has the lowest waiting time of all traffic classes. We also see that even if class two and three have the same priority parameter, class two traffic has a lower waiting time resulting from the priority order in the prorating algorithm. This confirms that the prorating algorithm is a key contributor to the performance of the priority system. Finally, we see that the optimum operating point is, as in the two priority classes case, at $\gamma_1 = 1$. At this point, the "average" waiting time is at its lowest value.

4.2. Self-similar traffic

The priority pseudo-Bayesian algorithms have been derived with the fundamental Poisson traffic assumption. However, it has been observed that real data traffic does not behave like a Poisson process; instead it exhibits long-range dependence and self-similar characteristics [8]. In order to validate the pseudo-Bayesian algorithms for utilization in the "real world", we must submit them to non-Poisson traffic to observe their robustness. Thus, we have run simulations of the pseudo-Bayesian algorithms with an asymptotically self-similar process for data packet arrivals.

It has been shown that multiplexing constant-rate data streams, that have a Poisson arrival process and a heavy-tailed distributed stream lifetime with infinite variance, results in an overall data traffic flow which packet arrival process is asymptotically self-similar [14]. It can be shown that the following probability density function satisfies the property of a heavy-tailed distribution with infinite variance:

$$p[X = x] = \frac{4}{x(x+1)(x+2)} \quad \text{for } x \geqslant 1. \qquad (15)$$

For comparison, the proposed algorithms are simulated with self-similar traffic sources which overall data flows have the same arrival rates as the respective Poisson sources. To generate a self-similar traffic flow with an arrival rate of λ_i (packets per slot) for priority class i, we generate multiple data streams of priority class i which arrival process is Poisson with intensity λ_i^c, such that $\lambda_i^c E[X] = \lambda_i$, and X is distributed according to equation (15). Then each data stream generates one packet of priority class i per slot (per frame for the framed algorithm) for a random period of time, i.e., the stream lifetime, drawn from the probability distribution given by equation (15). The data packets from the multiple data streams are combined into one data flow, which has an overall packet arrival rate of λ_i.

The basic (non-priority) and priority pseudo-Bayesian algorithms have been simulated with the same traffic scenarios. We have also computed separately the average waiting time of each traffic class in the non-priority algorithm to allow a better comparison of the results between the non-priority and priority algorithms. For each traffic scenario, we have also determined the average waiting time improvement for high priority (traffic class one) packets. By feeding self-similar traffic to the system based on Poisson traffic assumptions, it is not guaranteed that the system will be stable or that it will achieve maximum throughput. However, no unstable conditions were observed from the simulations with self-similar traffic when the total traffic was kept below the maximum throughput for Poisson traffic (i.e., $\sum_{i=1}^{p} \lambda_i \leqslant e^{-1}$). Although this does not guarantee stability, it demonstrates the robustness of the algorithm.

Table 1
Class one waiting time (frames) for framed pseudo-Bayesian algorithms with self-similar traffic (90% c.i.).

λ_1	λ_2	Poisson traffic	Non-priority algorithm	Priority algorithm	Improvement	
0.05	0.15	1.4	1.7	1.5	0.16	(0.15–0.16)
0.05	0.20	1.6	2.6	1.8	0.79	(0.77–0.81)
0.05	0.25	2.0	6.5	2.2	4.3	(4.0–4.6)
0.05	0.30	4.7	34	2.6	31	(29–33)
0.10	0.10	1.4	1.7	1.5	0.15	(0.15–0.16)
0.10	0.15	1.6	2.7	1.8	0.77	(0.75–0.79)
0.10	0.20	2.0	6.2	2.2	4.0	(3.8–4.2)
0.10	0.25	4.7	34	2.6	31	(30–33)
0.15	0.15	2.0	6.1	2.2	3.9	(3.7–4.1)
0.15	0.20	4.7	33	2.6	31	(29–32)

Table 1 presents the average waiting time results obtained for traffic class one with the framed system. Waiting time improvement results are given with their 90% confidence interval for 400 simulation runs. Results for the non-priority pseudo-Bayesian algorithm with Poisson traffic are also presented for comparison. From these results, it is evident that there is a degradation of the average waiting time with self-similar traffic, compared to the average waiting time obtained with Poisson traffic. However, this is expected since self-similar traffic is more bursty than Poisson traffic, and thus, some packets can experience very long delays. Nevertheless, the results show that the priority algorithm performs well compared to the non-priority algorithm. Furthermore, when the overall traffic is close to saturating the channel, the average waiting time of the high-priority traffic remains stable at a level similar to that under less congested traffic conditions. Similar results have been obtained for the slotted system. While it is not possible to present the results of all the simulations performed for a much wider number of traffic scenarios for both the slotted and framed systems (interested readers are referred to [4]), it is notable that in all cases considered, our proposed priority scheme always improves the average waiting time for the high priority traffic.

5. Conclusion

In this paper we have presented a new pseudo-Bayesian ALOHA algorithm with priorities. We have shown by our simulation results that our algorithm provides a significant delay improvement for high-priority packets with both Poisson and self-similar traffic characteristics, while low-priority packets only experience a slight performance degradation. The main advantages of the proposed scheme over previously known priority protocols are its simplicity and its adaptability to the frame structure widely used for wireless ATM. The proposed priority protocol can be used in any situation where multiple traffic streams with different quality of service requirements contend for the same multiple access channel.

Our framed priority protocol is well suited for application in reservation MAC protocols where a certain number of slots per frame are used for control traffic contention. The traffic in these slots, consisting of packets for reservation at the beginning of a voice talk spurt or data burst, requests for new connection admission, handoff requests, etc., can be well approximated as Poisson. Our priority protocol can be used to implement access priorities among these different control traffic types, since the contention delay is an important factor in the overall performance of these MAC protocols with respect to specific traffic classes. For example, in [4,5], we have proposed a wireless ATM MAC protocol where the control slots access is managed by the framed pseudo-Bayesian priority algorithm, and presented simulation results which show that the quality-of-service offered to voice connections is significantly improved and the overall throughput for the integrated voice and data system is enhanced. Thus, we believe that the proposed pseudo-Bayesian protocol with mixed priorities could be an important element of any efficient MAC protocol for multimedia wireless ATM.

Finally, it would be interesting to extend the analysis presented in section 2 to the CSMA and CSMA/CD protocols in order to obtain higher throughput. Furthermore, for the case where the control packets in a reservation protocol can be sent during the uplink control period in a slotted channel with carrier-sensing capability, the overall performance of the protocol could be significantly improved by using CSMA or CSMA/CD in combination with our pseudo-Bayesian priority algorithm.

References

[1] F. Bauchot et al., MASCARA: A MAC protocol for wireless ATM, in: *Proceedings ACTS Mobile Summit*, Granada, Spain (November 1996) pp. 17–22.

[2] D. Bertsekas and R. Gallager, *Data Networks*, 2nd edition (Prentice-Hall, Upper Saddle River, NJ, 1992).

[3] W.S. Chung and C.K. Un, Collision resolution algorithm for M-priority users, IEE Proceedings Communications 142 (June 1995) 151–157.

[4] J.F. Frigon, Dynamic reservation TDMA medium access control protocol for wireless ATM networks, Master's thesis, University of British Columbia (August 1998).

[5] J.F. Frigon, H.C.B. Chan and V.C.M. Leung, Data and voice integration in DR-TDMA for wireless ATM networks, in: *Proceedings ICC'99*, Vancouver, Canada (June 1999).

[6] O. Kubbar and H.T. Mouftah, Multiple access control protocols for wireless ATM: Problems definition and design objectives, IEEE Communications Magazine 35 (November 1997) 93–99.

[7] J.G. Lee and M.S. Corson, The performance of an "imbedded" Aloha protocol in wireless networks, in: *Proceedings of PIMRC'96*, Vol. 2, Taipei, Taiwan (October 1996) pp. 377–381.

[8] W.E. Leland, M.S. Taqqu, W. Willinger and D.V. Wilson, On the self-similar nature of Ethernet traffic (extended version), IEEE/ACM Transactions on Networking 2 (February 1994) 1–15.

[9] M. Liu and P. Papantoni-Kazakos, A random access algorithm for data networks carrying high priority traffic, in: *Proceedings of INFOCOM'90*, Vol. 3, San Francisco, CA (June 1990) pp. 1087–1094.

[10] D. Makrakis, R.S. Mander and L. Orozco-Barbosa, A spread slotted random access control protocol with multi-priority for personal and mobile communication networks carrying integrated traffic, in: *Proceedings of 1995 Canadian Conference on Electrical and Computer Engineering*, Vol. 1, Montréal, Canada (September 1995) pp. 132–136.

[11] P. Papantoni-Kazakos, Multiple-access algorithms for a system with mixed traffic: High and low priority, IEEE Transactions on Communications 40 (March 1992) 541–555.

[12] T. Papantoni-Kazakos, N.B. Likhanov and B.S. Tsybakov, A protocol for random multiple access of packets with mixed priorities in wireless networks, IEEE Journal on Selected Areas in Communications 13 (November 1995) 1324–1331.

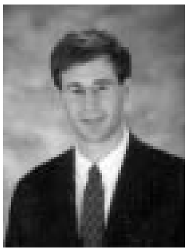
[13] M. Paterakis and Y. Gong, Performance analysis of random access multiuser algorithms for packets with different priorities, in: *Proceedings of INFOCOM'90*, Vol. 2, San Francisco, CA (June 1990) pp. 580–587.

[14] V. Paxson and S. Floyd, Wide area traffic: The failure of Poisson modeling, IEEE/ACM Transactions on Networking 3 (June 1995) 226–244.

[15] D. Raychaudhuri, L.J. French, R.J. Siracusa, S.K. Biswas, R. Yuan, P. Narasimhan and C.A. Johnston, WATMnet: A prototype wireless ATM system for multimedia personal communication, IEEE Journal on Selected Areas in Communications 15 (January 1997) 83–95.

[16] R.L. Rivest, Network control by Bayesian broadcast, IEEE Transactions on Information Theory 33 (May 1987) 323–328.

[17] I. Stavrakakis and D. Kazakos, A multiuser random-access communication system for users with different priorities, IEEE Transactions on Communications 39 (November 1991) 1538–1541.



Jean-François Frigon received the B.Eng. degree in electrical engineering from École Polytechnique de Montreal, Montreal, Canada, in 1996, and the M.A.Sc. degree in electrical and computer engineering from the University of British Columbia, Vancouver, Canada, in 1998. He is currently pursuing the Ph.D. degree in electrical engineering at the University of California, Los Angeles (UCLA). In 1998, he was awarded the Governor General's Gold Medal Award as the best graduate student of the University of Britich Columbia graduating class for the Master's degree. His Master and Ph.D. studies have been partially supported by a postgraduate scholarship from the National Sciences and Engineering Research Council of Canada. His research interests include wireless networks, MAC protocols and design of digital communication systems for high speed wireless communications.
E-mail: jeanf@ee.ucla.edu

Victor C.M. Leung received the B.A.Sc. (Hons.) degree in electrical engineering from the University of British Columbia (UBC) in 1977, and was awarded the APEBC Gold Medal as the head of the graduating class in the Faculty of Applied Science. He attended graduate school at UBC on a Natural Sciences and Engineering Research Council Postgraduate Scholarship and obtained the Ph.D. degree in electrical engineering in 1981. From 1981 to 1987, Dr. Leung was a senior Member of Technical Staff at MPR Teltech Ltd., specializing in the planning, design and analysis of satellite communication systems. He also held a part-time position as a Visiting Assistant Professor at Simon Fraser University in 1986 and 1987. In 1988, he was a Lecturer in the Department of Electronics at the Chinese University of Hong Kong. He joined the Department of Electrical and Computer Engineering at UBC in 1989, where he is a Professor and a member of the Centre for Integrated Computer Systems Research. In January 2000, he was appointed to the TELUS Mobility Research Chair in Advanced Telecommunications Engineering at UBC. He is also a project leader in the Canadian Institute for Telecommunications Research, a Network of Centres of Excellence funded by the Canadian Government. He has served on the Editorial Board of the IEEE Journal on Selected Areas in Communications – Wireless Communications Series since 1999. His research interests are in the areas of architectural and protocol design and performance analysis for computer and telecommunication networks, with applications in satellite, mobile, personal communications and high speed networks. Dr. Leung is a senior member of IEEE and a voting member of ACM.
E-mail: vleung@ece.ubc.ca