FIFO by sets ALOHA (FS-ALOHA): a collision resolution algorithm for the contention channel in wireless ATM systems

D. Vázquez-Cortizo [a,*], J. García [b,1], C. Blondia [a,2], B. Van Houdt [a,3]

[a] *University of Antwerp, Department of Mathematics and Computer Science, Performance Analysis of Telecommunication Systems Research Group, Universiteitsplein, 1, B-2610 Antwerp, Belgium*
[b] *Polytechnic University of Catalonia, Computer Architecture Department, Jordi Girona, 1–3, E-08034 Barcelona, Spain*

Abstract

This paper presents a Collision Resolution Algorithm which significantly improves the performance in terms of throughput and delay of the uplink contention channel existing in Wireless ATM MAC protocols such as MASCARA [9], where slotted ALOHA is employed. FS-ALOHA groups the requests arrived at the mobile terminals during a frame length and serves these groups on a FIFO basis using slotted ALOHA. An exact analytical model has been developed for FS-ALOHA with Poisson input model. From a simulation with finite population input process we show that FS-ALOHA largely outperforms slotted ALOHA in the delay quantiles, allowing a better meeting of the Quality of Service guarantees. © 1999 Published by Elsevier Science B.V. All rights reserved.

Keywords: Contention resolution; Transmission sets; QBD processes

1. Introduction

During the last years, the increasing interest in wireless indoor access systems has motivated the standardization process of DECT [11] (the European standard for digital cordless communications) and both IEEE 802.11 [1] and HIPERLAN I [2] wireless LAN standards. More recently, the combination with the trends towards high capacity, service integration and Quality of Service (QoS) provisioning in fixed networks (with the ATM standardization on the one hand and IETF activities on the other) has resulted in a good number of proposals for ATM-compatible wireless MAC protocols.

Within the framework of the European research programmes RACE II and ACTS schemes such as DSA++ [13] (MBS project) or MASCARA [9] (WAND project) have been developed. Other interesting

---

* Corresponding author. E-mail: dcortizo@uia.ua.ac.be
[1] E-mail: jorge @ac.upc.es
[2] E-mail: blondia@uia.ua.ac.be
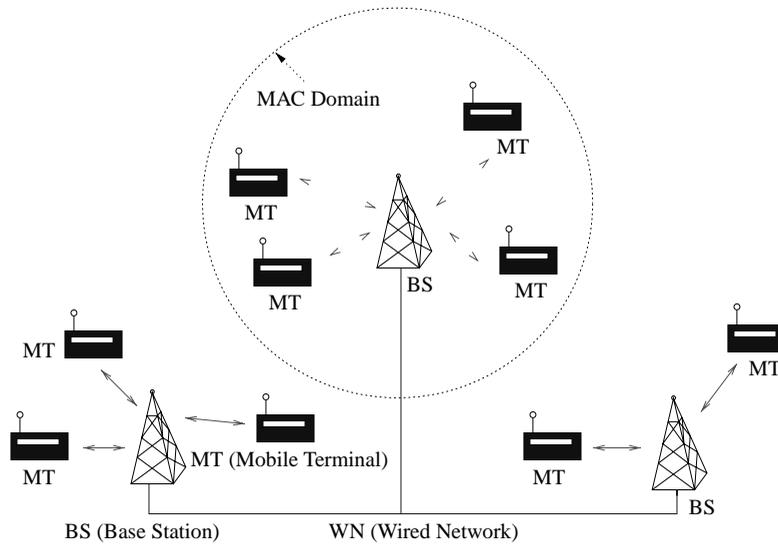[3] E-mail: vanhoudt@uia.ua.ac.be

Fig. 1. Architecture of the system.

protocols are DQRUMA [8], GRAPO [6] and those proposed in [4] and [14]. Among the features shared by the majority of these protocols the following can be highlighted:

- Centralized architecture, with a Base Station (BS) providing seamless connection with the fixed network and ruling the access to the broadcast radio channel (see Fig. 1).
- Two logically distinct communication channels, Uplink from the Mobile Terminals (MTs) to the Base Station and Downlink from Base Station to Mobile Terminals.
- Dynamic Time Division Multiple Access with either Frequency Division Duplexing (FDD) or Time Division Duplexing (TDD) to separate Uplink from Downlink.
- The systems are slotted and sometimes slots are subdivided into minislots.
- Existence of a frame structure (with a header emitted by the BS) which allows the switching of the MTs into power saving modes that reduce the drainage of the batteries.
- The Mobile Terminals inform the Base Station about their bandwidth needs using requests which are piggybacked in upstream data packets.
- MTs that cannot use the piggybacking mechanism to send their requests (e.g. MTs that have remained silent for a long time, with a sudden increase in their transmission needs or first entering the cell after handover from the adjacent cell) use a contention period existing in the Uplink.
- The BS informs the MTs about the slots they are granted in the Uplink channel by means of permits sent in the Downlink frame header.

In an integrated services environment the Contention Resolution Algorithm used in the contention period may have a major impact on the QoS provisioning and overall system performance. The use that a given connection makes of the contention period will strongly depend on the traffic characteristics: a CBR connection might never use the contention period, while the bursty nature of VBR traffic makes it prone to a continuous oscillation between periods where the bandwidth needs are simply piggybacked and periods where the contention period will be used. In these situations, a delay-efficient Contention Resolution Algorithm is needed in order to comply with the QoS contract, as the access delay will

be a multiple of the frame (Uplink frame in FDD systems) duration (1.7 ms in [3] and 2.0 ms in [15]). The schemes used in most of the protocols mentioned above are either slotted ALOHA [9] or splitting algorithms [13]. This paper presents and evaluates the performance (via analytical models and simulation) of FS-ALOHA, a new Collision Resolution Algorithm to be implemented during the contention period of the Uplink channel which allows an efficient transmission of the requests for capacity from the MTs to the Base Station. FS-ALOHA can be regarded as a Group Random Access Protocol as:

- the requests are grouped in Transmission Sets (TSs) so that just one group attempts transmission at a time (i.e. a subset of the total number of requests);
- the requests belonging to a certain group use a Random Access Algorithm (slotted ALOHA) to gain access to the medium and therefore get the permits from the scheduler at the base station.

FS-ALOHA combines the simplicity of slotted ALOHA with the efficiency obtained by grouping the requests arrived at the MTs during frame length intervals. Similar ideas were applied in [7] to define a MAC protocol for an ad-hoc wireless network. As we will show, FS-ALOHA largely outperforms slotted ALOHA in the delay quantiles without adding significant complexity (it could be even discussed if any at all). As for the throughput, FS-ALOHA shows a much more stable behavior when the network is congested, avoiding the well known throughput collapse experienced by slotted ALOHA. Only when the number of slots associated to the contention channel is very large, the maximum attainable throughput can be somewhat lower than the corresponding results obtained with Slotted ALOHA (provided an optimal retransmission probability is used).

To summarize, the advantages of the proposed mechanism are the following:

- Simplicity, as its associated computational complexity is comparable to that of slotted ALOHA.
- Good delay response, with delay percentiles that largely improve those obtainable with slotted ALOHA.
- Stability, as even during highly congested periods the throughput does not significantly decrease from the maximum.

The system we consider for FS-ALOHA is similar to all those related to the Wireless ATM protocols described above, as the only needed assumptions are the following:

- Centralized, slotted, dynamic TDMA system with either Frequency Division Duplexing or Time Division Duplexing.
- The frame length [4] is assumed to be fixed and equal to L slots. Although this constraint is not needed for the operation of the protocol, it greatly simplifies the analysis of the algorithm as different frame lengths would imply different grouping interval lengths.
- A frame is subdivided into a contentionless period where the mobiles transmit data packets according to the received permits together with piggybacked requests for future permits and a contention period where the mobiles send the transmission requests that can not be piggybacked.

The remainder of this paper is structured as follows. Section 2 describes the FS-ALOHA algorithm. Section 3 is devoted to the analytical performance evaluation of FS-ALOHA. Section 4 compares the performance of FS-ALOHA and slotted ALOHA schemes under different loads and sizes of the contention channel. Finally, Section 5 presents the conclusions of the paper.

---

[4] With FDD systems we only consider the Uplink frames.

2. FIFO-by-Sets ALOHA algorithm

In this section the operation of FS-ALOHA is described in detail. Basically, the potential users of the contention channel are Mobile Terminals (MTs) that can not communicate their bandwidth requirements to the Base Station through the piggybacking of the request in a previously scheduled data packet (transmitted in the contentionless period).

From the L slots of the fixed length frame C slots ($C \leq L$) will be used for contention resolution. Assuming k requests can be transmitted during a slot (i.e. a slot contains k minislots) then a maximum number of $T = k \times C$ requests can be transmitted. As the scope of FS-ALOHA only involves the contention period in the Uplink, it is the time needed to send a request what really matters, and not the time needed to send a data packet in the contentionless period. Hence, from here on with the word slot we will refer to the time needed to transmit a request in the contention period.

In slotted ALOHA systems, an MT with a pending request will randomly choose one out of the T possible slots to send its request in the hope that no other MT will choose the same slot. FS-ALOHA, on the contrary, divides the T slots of the contention channel into two disjoint sets of S and N slots such that $S + N = T$ (see Fig. 2). The operation of FS-ALOHA is as follows:

- Those requests newly arrived in the system during the last frame (at different MTs) will attempt transmission choosing randomly one out of the S slots.
- If a request arrives without collision at the BS, it is used to schedule a future cell transmission in the Uplink.
- If some of these attempts are unsuccessful they are grouped into a Transmission Set, which joins the queue of TSs pending to be served.
- The other N slots will be used to serve the queue of backlogged Transmission Sets on a FIFO basis. A Transmission Set will remain in the head of the queue until all its associated requests have been successfully sent to the Base Station.

  Hence the parameters of the scheme are just two:
- number of signalling slots per frame $S \geq 1$ used to transmit the newly arrived requests; this parameter determines the TS generation rate;
- number of slots $N \geq 2$ allocated to the service of the backlogged Transmission Sets in the distributed queue.

For any FS-ALOHA system with a given T there is always a trade-off between S, which indirectly determines the arrival rate at the Transmission Set distributed queue, and N, which affects the statistical distribution of the service time required by the Transmission Sets in the queue.

Consecutive sequence numbers are assigned to the Transmission Sets (see Fig. 3). In every frame i a new TS $N_G(i)$ may be generated, if not all the requests arrived during the previous frame are successfully transmitted in the S contention slots. At the same time, another TS numbered $N_S(i)$ is
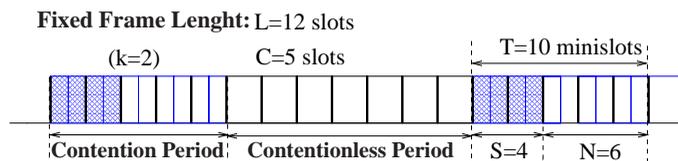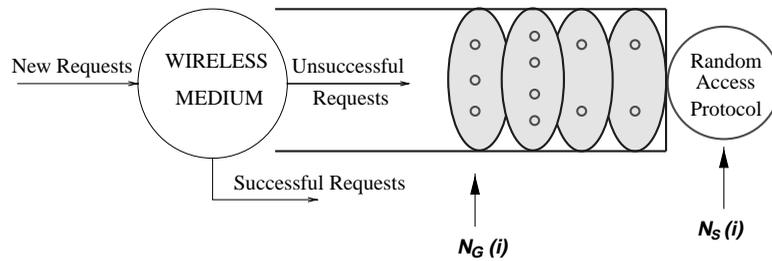

Fig. 2. Frame structure in the Uplink channel.

Fig. 3. Construction of the distributed FIFO queue.

being served making use of the N contention slots, so the number of TSs in the distributed queue is $N_G(i) - N_S(i)$.

Depending on the system implementing FS-ALOHA, requests can be issued on a per-VC or per-MT basis. In both cases, and assuming that new requests can be sent before the previous ones have been acknowledged by the Base Station, FS-ALOHA does not guarantee the correct request sequencing at the central scheduler. However, it is important to note that this characteristic does not preclude the correct sequencing of ATM cells guaranteed in any ATM system, as the central scheduler will send a permit in the Downlink and the corresponding Mobile Terminal will then send its packet(s) in the scheduled slot(s) in the Uplink, logically preserving the sequencing of the information.

2.1. Generation of the transmission sets

Let us assume that the system has already been initialized and that the Transmission Set distributed queue is not empty. All the mobiles with a new request for the contention channel generated during the last frame i will randomly choose one of the S contention slots to send their request packet. For each of the S slots the Base Station observes one of the following situations:
- there has been an error in the channel or there has been more than one request attempt (i.e. an errored CRC or collision is detected);
- there has been just one request attempt (there has been activity in the channel and the CRC is correct);
- there has been no request attempt (silence in the signalling slot).

Based on these observations the Base Station takes the corresponding actions regarding the generating TS $N_G(i)$:
- If neither collisions nor errors were detected in any of the S slots the TS $N_G(i)$ is not generated.
- Otherwise a new Transmission Set is built and stored in the distributed TS FIFO queue (see Fig. 3).
- Passes the successful requests from the MTs to the central scheduler located at the BS.
- Notifies the MTs whether or not a new TS was generated and which requests were successfully received. This can be done by including in the frame header a field of length S bits (S-bit header field). Each bit in this field is associated to one of the S contention slots, and contains a 1 if a successful request was received in the slot or a 0 if no request, collision or error were detected.

By keeping track of the S-bit header field (see also Section 2.3) the MTs can keep an image of the state of the TS distributed queue knowing at every moment how many TSs there are in it. When this TS FIFO queue is empty, either right after the initialization of the network or after a period of low activity, $N_G(i) = N_S(i)$ and there are no backlogged Transmission Sets to be served using the N contention slots. In this situation, the arrived packets will randomly choose one out of the total number $T = S + N$

of slots in the contention channel, instead of just using the S slots. Therefore, when the system is empty (i.e. no TSs are backlogged) the behavior of the system is identical to slotted ALOHA. The detection of a collision or errored request in one of these T slots will cause the generation of a TS and the beginning of a busy period: in the next frame only the first S slots will be used for new requests and the other N to serve the Transmission Set in the queue.

2.2. Service of the transmission sets

During the busy periods of the system (i.e. when the TS distributed queue is not empty) N out of the $T = S + N$ slots of the contention period will be used to serve the groups formed by unsuccessful requests. This will be done on a FIFO basis. The N slots will be used by the Transmission Set at the head of the queue, frame after frame, until all its associated requests are successfully transmitted (actually only during a limited number of frames to avoid Head of the Line blocking; see Section 2.3).

Assuming that the MTs have built correctly the distributed TS queue, all the mobiles with requests belonging to the TS in service will choose one out of the N slots to send their request, using slotted ALOHA algorithm. After receiving the information contained in the contention period the base station does the following:

- Hands over the successfully received requests to the central scheduler, in charge for generating the corresponding permits. The bandwidth allocation algorithm needed to assign the permits to the various MTs is not the subject of the paper.
- Notifies to the MTs the results by including in the downlink frame header a field of length N bits (N-bit header field). In the same way as the S-bit header field, each bit in this field is associated to one of the N contention slots, and is set to 1 if a successful request was received in the slot or to 0 if no request, a collision or an error were detected.

If there is a collision (at least two mobiles made the attempt in the same slot) or some request is errored in any of the N slots the service of the Transmission Set will continue in the next frame. Otherwise, the TS is removed from the queue and the service of the next TS will be initiated in the following frame. If the queue becomes empty an idle period starts, where all the slots (T) available in the contention channel are used by the new request arrivals.

Let us recall that during all this process the mobiles are blind, just knowing about their own sent requests and without any information on collisions. The necessary feedback is provided by the N-bit header field transmitted by the Base Station in the Downlink channel (more details are given in the next section).

2.3. Robustness and implementation issues

In this section we will discuss two implementation aspects regarding robustness, namely how to ensure that the Mobile Terminals construct the right image of the distributed queue of Transmission Sets generated at the Base Station and how to avoid Head of the Line blocking (HOL).

The main mechanism addressing the former issue has already been described in the two preceding sections: the Base Station includes in the frame header a field of $S + N$ bits, each of them associated to one of the slots of the contention period. Each bit is set to 1 if a successful request was conveyed in the associated slot or to 0 otherwise. The MTs know which contention slot they have used to send their request so by listening to the $S + N$ bit header field they know exactly whether their request was successful or not.

However, the preceding mechanism can only ensure the correct distributed queue at an MT if the mobile has been connected since the initialization of the network and no errors have occurred in the generation, transmission or processing of this header field. In order to have a robust mechanism for the construction of the distributed FIFO queue of Transmission Sets, at least two options are available:

- Inclusion in the frame header (i) of two new fields, one containing the sequence number of the TS in generation $N_G(i)$ and the other the sequence number of the TS in service $N_S(i)$.
- Inclusion in the frame header of just one new field containing the number of Transmission Sets in the distributed queue.

In addition to one of the two mechanisms above, the Mobile Terminals could also include in their requests the sequence number of the Transmission Set they belong to, leaving up to the Base Station the checking of the correctness of these sequence numbers and the notification to the corresponding mobile when necessary.

Another problem associated to the wireless channel is capture (see [12]), which may cause that one of the uplink requests involved in a collision is nevertheless successfully received by the Base Station. If just one bit per request slot in the contention period is used to indicate either success or collision, the MTs whose requests have been captured will erroneously assume that their requests were successfully received by the BS. A solution for this problem is to indicate the MAC address of the Mobile Terminal whose request has been received, thus replacing the $(S + N)$-bit field in the Downlink frame header by a field containing $(S + N)$ Mobile Terminal MAC addresses.

A different problem arises when, due to the special characteristics of the wireless medium, the propagation link between an MT and the BS enters a state of high BER (Bit Error Rate) which can last for several milliseconds. In this case a Head of Line blocking may appear: in the temporarily bad propagation channel many retransmissions of packets will occur causing long delays in the TSs services. Due to the FCFS discipline used in the TS service, the overall network performance will fall. To avoid this performance degradation, each packet will be retransmitted only a limited number of times during the service of a TS. When a packet reaches the maximum it will be removed from the TS in service and reserved for service in the next TS. If this situation remains for several TSs, the packet will be dropped.

3. Analytical performance evaluation of FS-ALOHA

3.1. Analytical model

In this section an exact analytical model is developed, allowing the computation of the delay density function associated to the request packets under the following assumptions:

- The grouping of requests in Transmission Sets is based on a time period corresponding to the frame length L.
- Slotted ALOHA is used as the request contention protocol.
- A Poissonian request arrival process with rate λ arrivals per frame (let us recall that these are transmission attempts in the contention channel, and not the total offered load in the network).
- The number of slots used for contention is fixed and such that $1 < L \times k$.
- Within these T slots, S are used for service of the new arrivals and N for service of the Transmission Sets in the queue.
- If there are no Transmission Sets in the queue the total T slots are used by the new arrivals.

- The impairment in performance caused by the Bit Error Rate (BER) of the wireless link has not been analyzed, as we have assumed that there are no errors in the channel.

The state of the system can be modelled by the pair (G, q), where G is the number of backlogged non-empty Transmission Sets and q is the number of requests pending for transmission in the Transmission Set being served (note that always $q \geq 2$).

The transitions in the system take place at the end of every frame according to the following rules:

- q may remain the same or move down to $\{q - 1, \ldots, \max(0, q - N)\}$, depending on the number of successful requests sent in the N contention slots.
- G builds up to $(G + 1)$ if q does not go down to zero and a new non-empty Transmission Set is effectively generated.
- G remains the same if either q does not go down to zero and no Transmission Set is generated or q does go down to zero but a new TS is placed in the distributed queue.
- G decreases, moving down to $(G - 1)$, if q goes down to zero and no Transmission Set is generated.

We can therefore conclude that the evolution of the system can be modelled by a discrete-time Markov chain, as the state (G_n, q_n) at step n only depends on the number of new requests arrived during the last frame, the number of successful new requests sent in the S contention slots using slotted ALOHA, the previous state (G_{n-1}, q_{n-1}) and the number of successfully transmitted requests (up to $\min(q_{n-1}, N)$) using slotted ALOHA in the N contention slots.

3.2. Transition probabilities and QBD-Markov chain

The transitions among states depend on the number of arrived requests and the number of requests that manage to leave the system (i.e. successfully transmitted to the base station). As the request arrival process is assumed to be Poisson, the probability of n new request arrivals is simply

$$P_a(n, \lambda) = \mathrm{e}^{-\lambda} \frac{\lambda^n}{n!}. \tag{1}$$

The requests allowed to leave the system, $q \geq 2$ if the state is (G, q), are only those belonging to the Transmission Set in service. As the contention protocol is slotted ALOHA they will all attempt transmission in a randomly chosen slot out of the N available. Aside from errors in the channel (which are not included in the model), errored transmissions are caused by collisions in the request slots. The probability of $n \leq \min(N, q)$ successful transmitted requests is given by

$$P_d(n, q, N) = \begin{cases} \dfrac{1}{N^q} \displaystyle\prod_{i=0}^{i=q-1} (N - i), & n = q, \\[2ex] 0, & n = q - 1, \\[2ex] 0, & n = N, < q \\[2ex] \dfrac{1}{N^q} \displaystyle\prod_{i=0}^{i=n-1} (N - i) \binom{q}{n} \displaystyle\sum_{j=1}^{\min(\lfloor \frac{q-n}{2} \rfloor, N-n)} \left[\binom{N - n}{j} \right. \\[2ex] \left. \times \displaystyle\prod_{m=1}^{j-1} \sum_{i_m=2}^{(2m+q-n-2j-\sum_{k=1}^{m-1} i_k)} \binom{q - n - \sum_{k=1}^{m-1} i_k}{i_m} \right], & \text{otherwise.} \end{cases} \tag{2}$$

being $\binom{m}{n} = \frac{m!}{n!(m-n)!}$. It must be noted, however, that in the practical implementation of the model the closed formula shown in Eq. (2) has been substituted by an iterative method. This method builds a sequence of matrices for increasing values of n but its computational complexity at each step is independent of the value of n, as opposed to what happens with Eq. (2). The former function $P_d(\cdot)$ will also be used to determine the probability of n successful new requests transmitted in the S contention slots provided that there have been l new arrivals, which is given by $P_d(n, l, S)$, and therefore to determine the probability of generation of a new Transmission Set:

$$\Pr(\text{TSgen}) = \begin{cases} \displaystyle\sum_{n=0}^{\infty} P_a(n, \lambda)\,(1 - P_d(n, n, S))\,, & \text{nonempty TS queue,} \\ \displaystyle\sum_{n=0}^{\infty} P_a(n, \lambda)\,(1 - P_d(n, n, T))\,, & \text{empty TS queue.} \end{cases} \tag{3}$$

The maximum number of request arrivals q per TS (i.e. per frame) is bounded by the maximum number of mobile terminals $q_{\max}$ allowed by the network. Moreover, in order to obtain a finite transition matrix we must truncate the maximum value of q to a certain value q_m. We can choose this q_m to be the same as the maximum number of stations $q_m = q_{\max}$ although this choice could be too conservative. If a tighter bound is used, the calculation of the solution vector will be simplified (the order of the transition matrix is multiple of q_m) and the impact on the accuracy of the results will be negligible provided that the bound is large enough.

With the former expressions the transition matrix $\boldsymbol{P}$ can be written as:

$$\boldsymbol{P} = \begin{pmatrix} \boldsymbol{B}_{00} & \boldsymbol{B}_{01} & 0 & 0 & 0 & \dots \\ \boldsymbol{B}_{10} & \boldsymbol{A}_1 & \boldsymbol{A}_0 & 0 & 0 & \dots \\ 0 & \boldsymbol{A}_2 & \boldsymbol{A}_1 & \boldsymbol{A}_0 & 0 & \dots \\ 0 & 0 & \boldsymbol{A}_2 & \boldsymbol{A}_1 & \boldsymbol{A}_0 & \\ \vdots & \vdots & & \ddots & \ddots & \ddots \end{pmatrix}. \tag{4}$$

The matrix $\boldsymbol{A}_0$ contains the transition probabilities from states (G, q) to states $(G+1, q')$ with $G \geq 1$, the matrix $\boldsymbol{A}_1$ contains the transition probabilities from states (G, q) to states (G, q') with $G \geq 0$ and the matrix $\boldsymbol{A}_2$ contains the transition probabilities from states (G, q) to states $(G-1, q')$ with $G \geq 1$. In all cases q and q' are such that $1 \leq q, q' \leq q_m$. The expressions for $\boldsymbol{B}_{00}, \boldsymbol{B}_{01}, \boldsymbol{B}_{10}, \boldsymbol{A}_0, \boldsymbol{A}_1$ and $\boldsymbol{A}_2$, can be found in Appendix A.

From the structure of $\boldsymbol{P}$ we see that this Markov chain is a generalized Quasi-Birth-and-Death (QBD) process [10]. The calculation of the solution vector $\boldsymbol{v}/\boldsymbol{v} \cdot \boldsymbol{P} = \boldsymbol{v}$ can be accomplished by applying any of the algorithms existing in the literature [16].

3.3. Calculation of the request delay density function

In this section a method for the calculation of the request delay density function is developed using the stationary state probabilities

$$\Pi(G, q) = \begin{cases} \boldsymbol{v}(0), & G = 0, \ q = 0, \\ \boldsymbol{v}(G * (q_m - 1) + q - 1), & G \geq 0, \ 2 \leq q \leq q_m. \end{cases} \tag{5}$$

The request delay $f_D(N, L, \lambda)$ is defined as the time elapsed from the generation of an arbitrary request at the mobile terminal to the reception of that request at the base station. As we have assumed a Poissonian arrival process we can make use of the PASTA property (Poisson Arrivals See Time Averages) to conclude that the probability of an arbitrary request finding the system in state (G, q) is given by Eq. (5).

The probability of n requests in the Transmission Set associated to our arbitrary request is given by

$$P(n, \lambda) = \begin{cases} 0, & n = 0, \\ P_a(n - 1, \lambda), & 1 \leq n \leq q_m. \end{cases} \tag{6}$$

with $P_a(\cdot)$ previously defined in Eq. (1). If the arbitrary request (or observer) is successfully transmitted in its first immediate attempt the associated delay will be zero. This probability, however, depends on the system state as when the system is empty T slots are used instead of S for immediate transmission. If the observer transmission attempt is unsuccessful then the delay will be in general the sum of three independent delays each a multiple of the frame length, which turns the delay density function into a discrete function by just considering the number of frames the request is delayed

$$\begin{aligned} f_D(N, \lambda)[n] = {}& \big[(1 - \Pi(0, 0)) p_{0\,\text{delay}|\text{Non Empty}} + \Pi(0, 0) p_{0\,\text{delay}|\text{Empty}} \big] \delta[n] \\ & + \Pi(0, 0) \cdot f_{CE}(\lambda, S, N)[n] \\ & + \sum_{G, q \neq (0,0)} \Pi(G, q) \cdot f_A(q, N)[n] * f_B(G, \lambda, S, N)[n] * f_C(\lambda, S, N)[n] \end{aligned} \tag{7}$$

and

$$\begin{aligned} p_{0\,\text{delay}|\text{Non Empty}} &= \sum_{j=1}^{q_m} P_a(j - 1, \lambda) \sum_{l=1}^{j} \frac{l}{j} P_d(l, j, S), \\ p_{0\,\text{delay}|\text{Empty}} &= \sum_{j=1}^{q_m} P_a(j - 1, \lambda) \sum_{l=1}^{j} \frac{l}{j} P_d(l, j, T). \end{aligned} \tag{8}$$

Indeed, the request will have to wait a random time t_A (with density function $f_A[n]$) until the q requests in the Transmission Set in service are successfully transmitted, plus a random time t_B (with density function $f_B[n]$) required for the G backlogged, non-empty (with at least two requests) TSs to be served and finally will have to contend with the other requests present in its own TS for a time t_C (or t_{CE} if the system was empty when the observer arrived) until it finally gets its way through to the base station. The details for the calculation of the density function $f_A(q, N)[n]$, $f_B(G, \lambda, S, N)[n]$, $f_C(\lambda, S, N)[n]$ and $f_{CE}(\lambda, S, N)[n]$ are conveyed in Appendix B.

3.4. Throughput of the protocol

In this section the maximum request (Poisson) input traffic load is determined as a function of the number of contention slots S and the number of slots devoted to the service of the Transmission Sets N.

Note that the overall function of the protocol can be viewed as the operation of a GI/GI/1 (distributed) FIFO queue. Customers are non-empty TSs, and the service of the queue corresponds to the service of a non-empty TS by means of slotted ALOHA. Being λ the average number of Poisson requests arriving at the system per frame (L slots), the maximum request traffic load can be obtained by finding the maximum value of λ verifying the inequality

$$\Lambda_{\text{TS}}(\lambda, N) \geq \mu_{\text{TS}}(\lambda, N, S), \tag{9}$$

where $\Lambda_{\text{TS}}(\lambda, N)$ is the average non-empty Transmission Set interarrival time measured in number of frames and $\mu_{\text{TS}}(\lambda, N, S)$ is the average service time of the non-empty Transmission Sets measured in frames as well.

Let us first address the calculation of $\Lambda_{\text{TS}}(\lambda, N)$. As we are aiming at calculating the maximum throughput we assume that the Transmission Set queue will never be empty, and therefore there will always be S slots (and not T) to process the new requests. We may then consider that the probability of getting a non-empty TS follows a Bernouilli distribution with parameter p_{TSgen} as defined in Eq. (A.4), and the average time elapsed between non-empty TSs is given by $\Lambda_{\text{TS}}(\lambda) = 1/p_{\text{TSgen}}$.

Using the expression for $f_{\text{TS}}(\lambda, N)$ in Eq. (B.5), the average service time of the non-empty Transmission Sets is simply $\mu_{\text{TS}}(\lambda, N, S) = E\left[f_{\text{TS}}(\lambda, N)[n]\right]$. The stability of the mechanism is determined by the speed at which TSs are generated (S dependent) and the time needed to process them (N dependent). Once the value $\lambda_{\max}(S, N)$ is obtained, the throughput of FS-ALOHA with parameters (S, N) is given by

$$\text{Throughput}(S, N) = \frac{\lambda_{\max}(S, N)}{S + N}, \tag{10}$$

as $(S + N)$ is the total number of slots consumed by the protocol and therefore is also the maximum number of requests which could ideally be transmitted using the same amount of resources.

3.5. Numerical results

In Figs. 4 and 5 we have depicted the delay percentiles obtained with FS-ALOHA and infinite population (Poisson) input process for different values of T (3,5,10) and different combinations for the parameters S and N. Aside from the obvious conclusion that the higher the load the worse the delay percentiles, it can be observed that when the load of the system is low then the best choices are those with a small value of S and correspondingly high value of N: as the load is low, it will be very unlikely to be unsuccessful in the first trial, the system will be empty most of the time and the first trial uses the whole T slots (i.e. independently of the values chosen for N and S). Even if the system is not empty, the probability of colliding in the S slots is low due to the (expected) few arrivals. Thus, having a high value for S will not really improve the performance of the system. However, whenever the queue of Transmission Sets builds up and the N slots are used for the service of the TSs, then the impact of a high value for N is important: if a TS is backlogged then it is likely to contain a large number of requests, and its service time will be shortened if a high number N of slots is used for its service.

On the contrary, when the load of the system is fairly high a good behavior of FS-ALOHA requires that both S and N are large enough: high values of S allow a good number of requests to be transmitted without any delay and reduce the TS generation rate, while large values of N allow a fast service of the backlogged Transmission Sets. Therefore, when the load of the network is high a compromise between
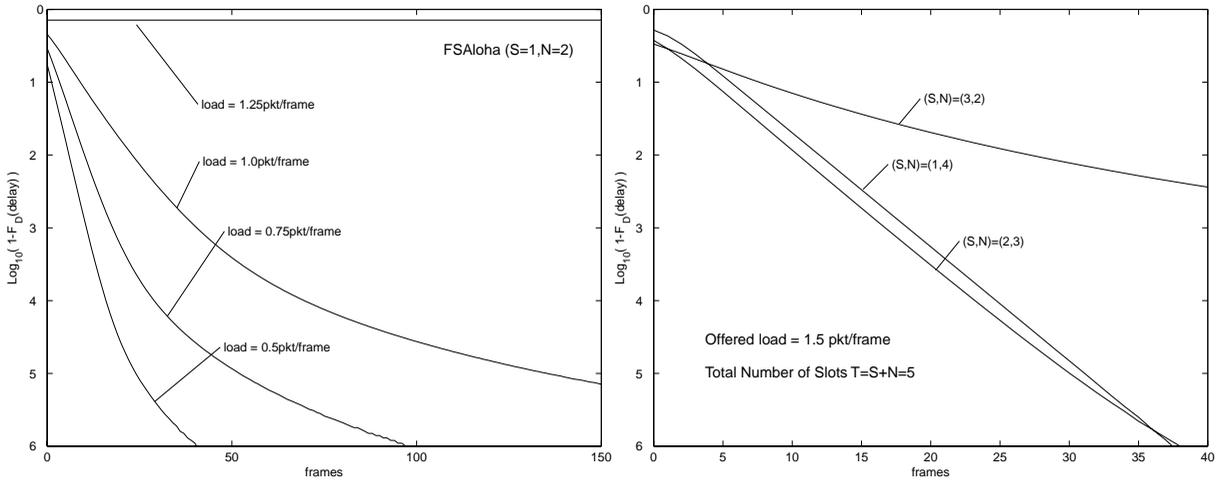
Fig. 4. Delay percentiles for $T = 3$ and $T = 5$ (Poisson model).

S and N is needed (see Fig. 5 with input rate $\lambda = 3$ packets per frame, where the best results are obtained for $S = 5$ and $N = 5$).

If no adaptive algorithm is to be implemented in order to obtain the optimum values for S and N ($S + N = T$ remaining constant), then the above results suggest that a good compromise for low and high input loads is sharing evenly the available number of slots T between S and N (if not possible S should be chosen lower than N).

In Fig. 6 both the maximum input load and the maximum throughput as defined in Eq. (10) are depicted for different values of S and N. It can be observed that the throughput decreases with the number of slots $T = S + N$ used by FS-ALOHA. When the number of available contention slots T is increased, the maximum input load does not increase proportionally and therefore the throughput decreases. The best throughput results for FS-ALOHA are obtained when T is small. At the same time,
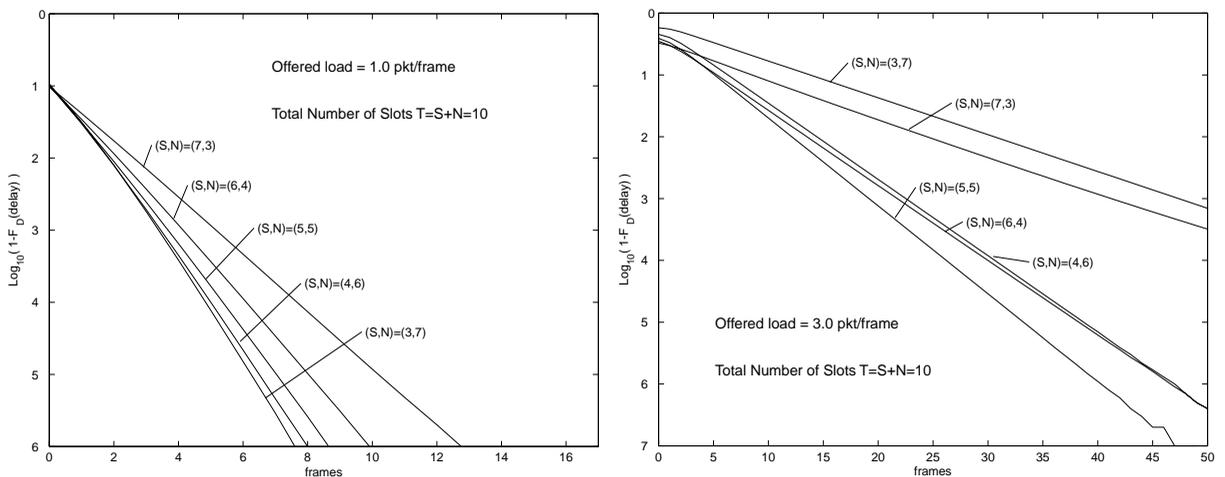
Fig. 5. Delay percentiles for $T = 10$ (Poisson model).
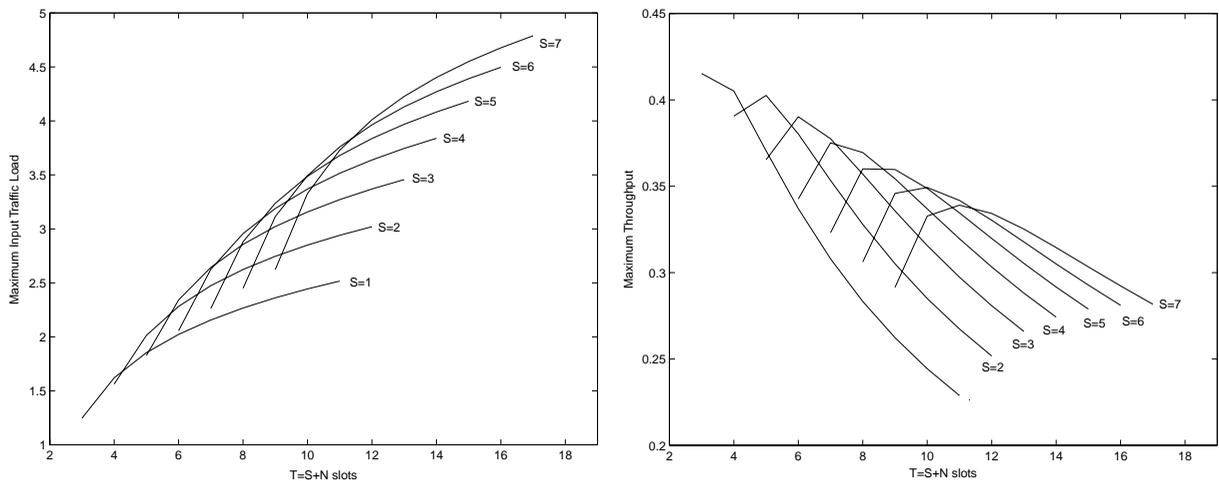
Fig. 6. Throughput and maximum input loads.

we can conclude that for a given T the best throughput result is obtained when S is equal or slightly smaller (one unit) than N.

4. Comparison with slotted ALOHA

In this section we compare FS-ALOHA with slotted ALOHA, a scheme that has been widely implemented in the uplink contention channel of wireless ATM systems due to its simplicity and the presumed low utilization of the contention period. In the analysis of the protocol we have only included scenarios where the size of the contention period is fixed. Although both slotted ALOHA and FS-ALOHA could definitely perform better if the parameter T was varied according to the offered request load, the aim in this paper is to evaluate the algorithms isolating them from the impact of added adaptive schemes. Moreover, when the network is heavily loaded it may be impossible to increase the size of the contention period and it is most interesting to see what happens in these situations. Unfortunately, the assumption of a constant size contention period precludes an easy comparison with splitting algorithms, which typically use a variable number of contention slots per frame according to the stage of the algorithm they are into (e.g. binary trees multiply by 2 the needed number of slots at each step of their group resolution algorithm).

Focusing on slotted ALOHA, now the problem is the inherent instability of these schemes (see [5]) if the available number of slots T is constant. The solution adopted is to assume a finite population input process, which allows the construction of an exact analytical model with Markov chains for slotted ALOHA. An analytical model for FS-ALOHA with finite population seems to be untractable, and therefore a simulator has been built for FS-ALOHA with finite population input process. The assumptions on the input model for both slotted ALOHA and FS-ALOHA are the following:
- There is a finite number M of MTs in the network.
- A global nominal input rate per frame λ is defined, equal to the actual input rate if the number of MTs is made infinite.
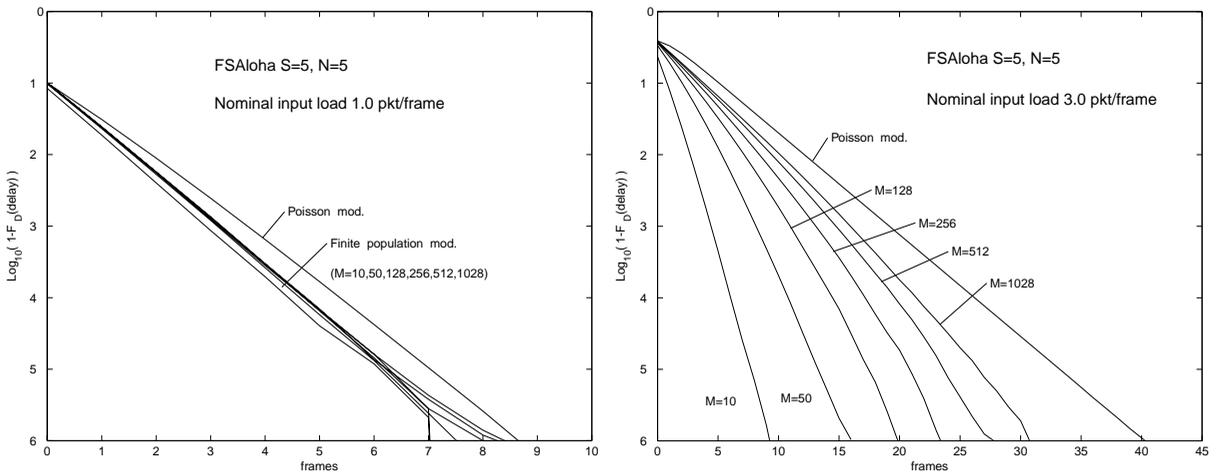
Fig. 7. FS-ALOHA performance under infinite Poisson and finite population input processes.

- An MT with a request pending to be transmitted (backlogged) will not generate a new request until having successfully transmitted the backlogged request.
- MTs with no backlogged requests will generate a new request in each frame with a probability $p_g = 1 - \mathrm{e}^{-\lambda/M}$.

In order to show the influence of the finiteness of the population, Fig. 7 depicts the performance of FS-ALOHA for both the Poisson (hence infinite population) and the finite population (with different number of Mobile Terminals) input processes, with equal nominal input loads. The real input load for the finite population model is actually lower due to the finite number of terminals, and it can be seen that the Poisson model is a worst case approximation. When the offered input load is high, the approximation is rather pessimistic if the number of terminals in the network is not too large.

Fig. 8 shows the behavior of the system when the offered load (3.0 requests per frame) is higher than the maximum input load guaranteeing stability for the Poisson model (as $T = 6$, FS-ALOHA throughput is less than 0.5), calculated in the previous section. In this case the infinite population model leads to an unstable system so no curve for the delay percentiles can be provided. From the curves shown in the figure it can be concluded that having a finite population leads to a stable system with bounded delays.

As for the slotted ALOHA system itself, the main assumptions are the following:

- The MTs will attempt transmission (choosing randomly one out of the T slots) with probability 1 when a new request is generated.
- When a new request is not successfully transmitted it is backlogged, and the corresponding MT will attempt transmission in the subsequent frames with probability p_{tx}.

The performance analysis of FS-ALOHA and slotted ALOHA will start with the case $T = 3$ (the minimum required for FS-ALOHA) and then continue with increasing values of T, namely 5, 6, 10 and 20. Both throughput and delay percentiles are depicted.

In all the scenarios studied for slotted ALOHA a trade-off has been found between throughput and delay through the choice of the retransmission probability p_{tx}: there is always a band of reasonable values within which a larger p_{tx} implies better delay percentiles but worse throughput (both the
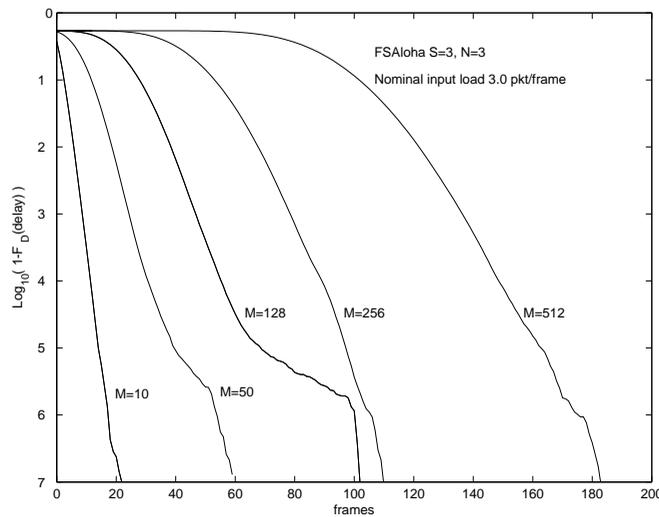
Fig. 8. Overloaded finite population model.

maximum attainable throughput and the abruptness of the throughput fall when the offered traffic surpasses the optimum). This behavior, although general, depends on the offered nominal load λ, the number of stations M and the size of the contention window T. Specifically, the larger the number of stations the worse the performance of slotted ALOHA as we approach the case of infinite population where slotted ALOHA is always unstable, independently of the offered load. This can be considered a first advantage of FS-ALOHA, which is stable even with infinite (Poisson) population as long as the offered load is kept below a maximum that can be analytically calculated (Section 3.4).

Figs. 9 and 10 show the results for the case $T = 3$, where for FS-ALOHA the only possible choice is ($S = 1, N = 2$). In the curve depicted for slotted ALOHA we assume $p_{tx} = 0.1$, as higher
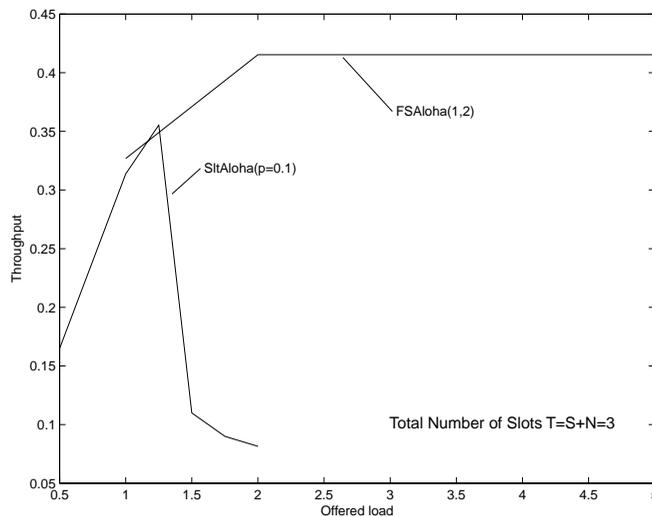
Fig. 9. Throughput for $T = 3$.
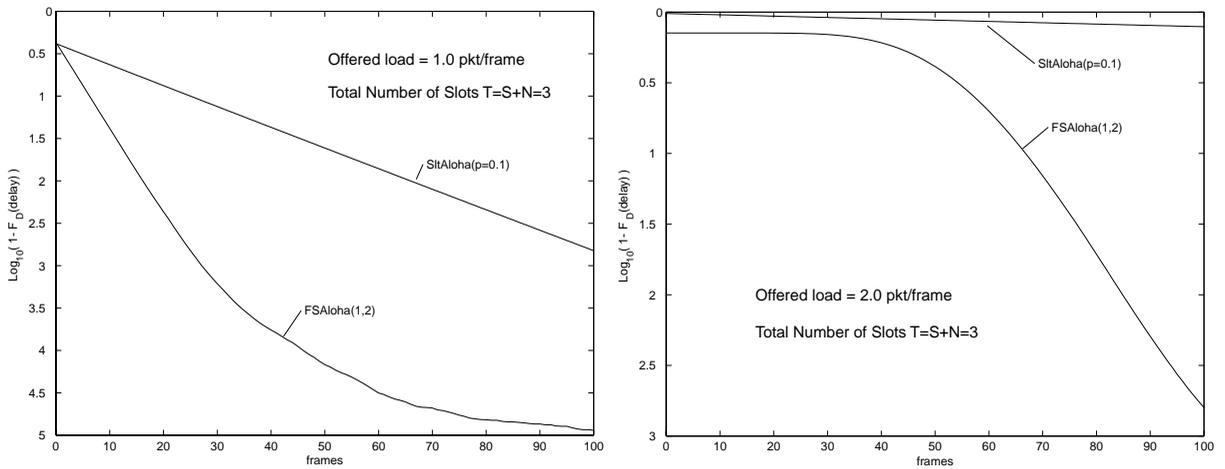
Fig. 10. Delay percentiles for $T = 3$, $\lambda = \{1.0, 2.0\}$.

retransmission probabilities are of no interest (with $p_{tx} = 0.2$ and a nominal load of $\lambda = 1.0$ the throughput is 1.379×10^{-3}). FS-ALOHA largely outperforms slotted ALOHA not only in the delay percentiles (Fig. 10) but also in the maximum throughput, higher than 0.4, and the excellent throughput behavior when the system is overloaded (Fig. 9).

The same conclusions can be drawn from the cases $T = 5$ (Figs. 11–13) and $T = 6$ (Figs. 14–17). In all cases, the best delay results with slotted ALOHA (always much worse anyway than those of FS-ALOHA) are achieved at the expense of low throughput and even a collapse in case of overload. On the contrary, the best delay results for FS-ALOHA are obtained with the same combination of the pair (S, N) ((2,3), (2,4) and (3,3)), almost independently of the traffic load. Given a fixed T there is no need
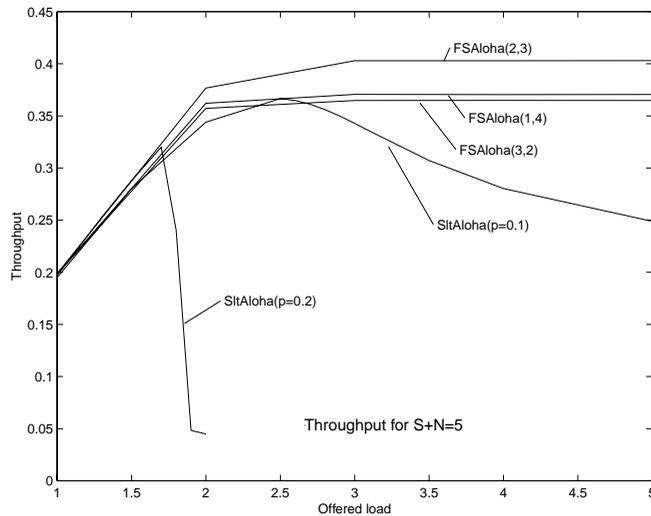
Fig. 11. Throughput for $T = 5$.
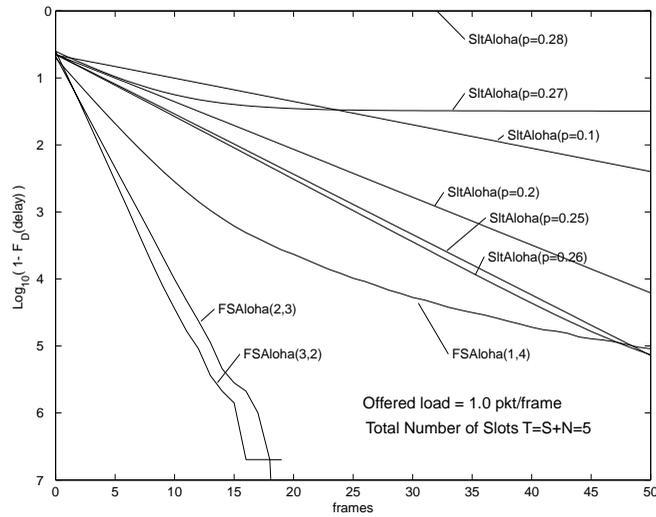
Fig. 12. Delay percentiles for $T = 5$, $\lambda = 1.0$.

therefore to develop an adaptive algorithm to vary (S, N) according to the offered load. Moreover, these optimal options for the delay turn out to provide the best throughput results. In all cases, the throughput attained with FS-ALOHA is very stable under overload conditions and does not decrease significantly from the maximum in situations where slotted ALOHA throughput is nearly zero.

Figs. 18–20 show the performance of FS-ALOHA when the number of available slots T is 10. Assuming that there are T slots available for contention with slotted ALOHA, the throughput when there are l requests is given by $(l/T)[1 - (1/T)]^{l-1}$ and it is maximized for $l = T$ (see [9]). If it was possible for the MAC protocol (using slotted ALOHA) to ensure that at every frame the mobiles send exactly T trials then the throughput of slotted ALOHA would be as depicted with the dotted line in Fig. 18. However, we consider that this is a rather unrealistic assumption and thus a quite unreachable bound.
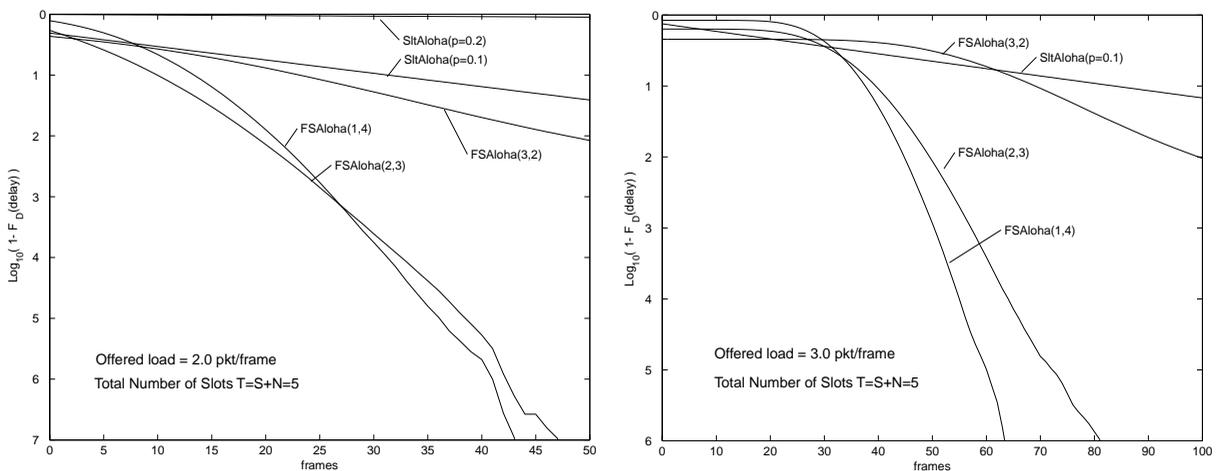

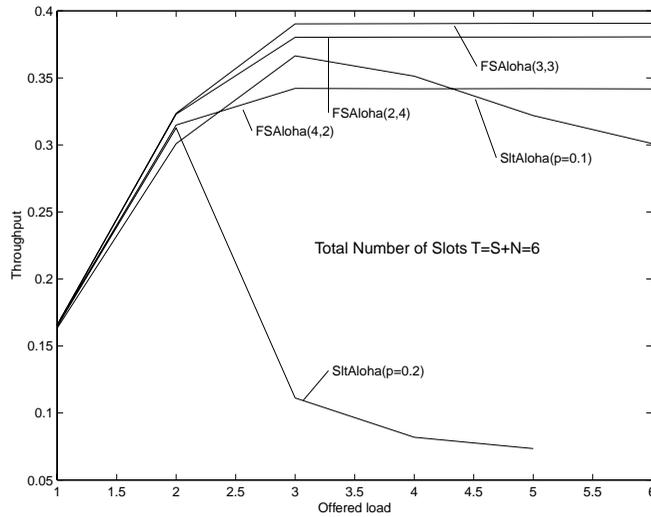Fig. 13. Delay percentiles for $T = 5$, $\lambda = \{2.0, 3.0\}$.

Fig. 14. Throughput for $T = 6$.

Figs. 19 and 20 include the delay percentiles of slotted ALOHA for different retransmission probabilities. It can be concluded that slotted ALOHA provides poor performance with regards to delay percentiles. For an offered load of 2.0 requests per frame and assuming an Uplink frame length of 1.7125 ms (see [3]), the 10^{-4} delay percentile is 12 ms for FS-ALOHA whereas 30.82 ms for the best slotted ALOHA case (with $p_{tx} = 0.5$). Moreover, the best delay results for slotted ALOHA are achieved at the expense of having an unstable protocol, collapsing at the slightest network overload (see in Fig. 18 the throughput curve corresponding to the optimum delay slotted ALOHA retransmission probability p_{tx}).

In all the scenarios shown in Figs. 19 and 20 the best FS-ALOHA results are obtained with the combinations $(S, N) = \{(3, 7), (4, 6), (5, 5)\}$. In case of low loads, and for the same reasons explained
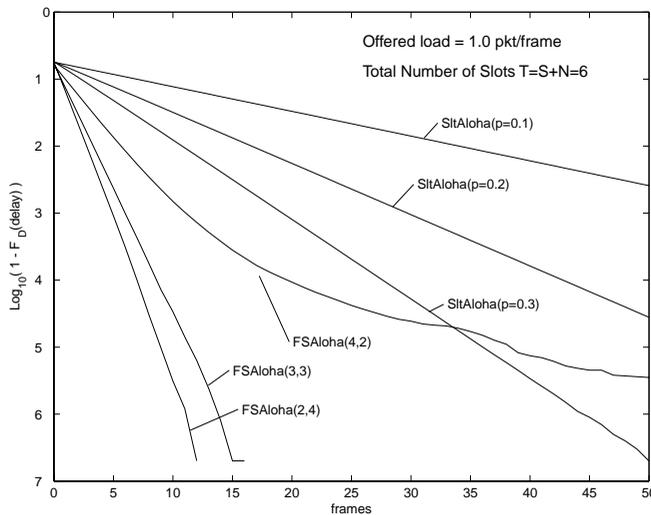

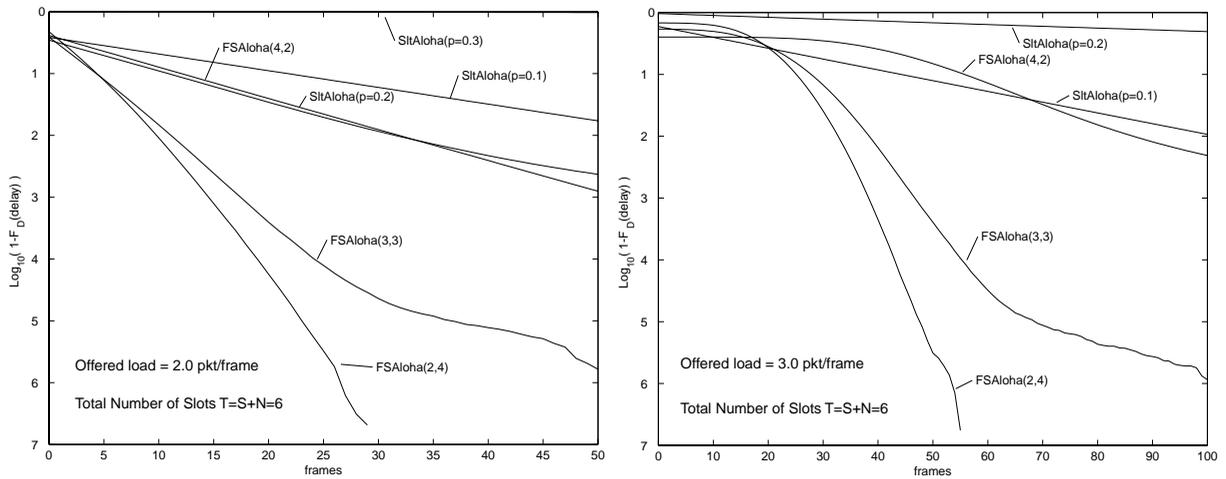Fig. 15. Delay percentiles for $T = 6$, $\lambda = 1.0$.

Fig. 16. Delay percentiles for $T = 6$, $\lambda = \{2.0, 3.0\}$.

in Section 3.3, the combination (3,7) is the best one. However, throughput considerations extracted from Fig. 18 might suggest the choice of (5,5) as a compromise between maximum throughput and best delay results: although for every (S, N) combination we get a very stable behavior of FS-ALOHA, with throughputs that do not decrease significantly despite strong overload conditions, there is indeed a difference among the maximum achievable throughputs of the combinations (3,7), (4,6) and (5,5). In Fig. 18 we can also see that slotted ALOHA with a very low retransmission probability $p_{tx} = 0.2$ slightly outperforms FS-ALOHA with respect to maximum throughput for a range of nominal offered loads between 4.0 and 6.0 requests per frame, even though its associated delay percentiles are very poor.

An extreme example showing the limitations of FS-ALOHA is presented in Figs. 21–23. This example is of little practical interest as if average values for the frame length lie between 80 and 100
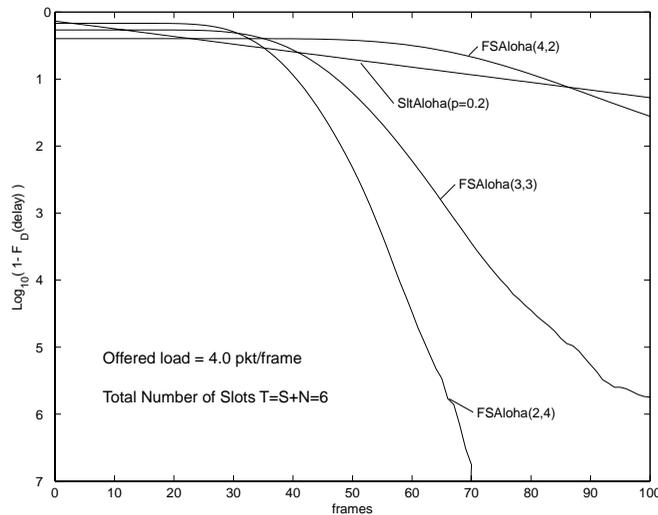

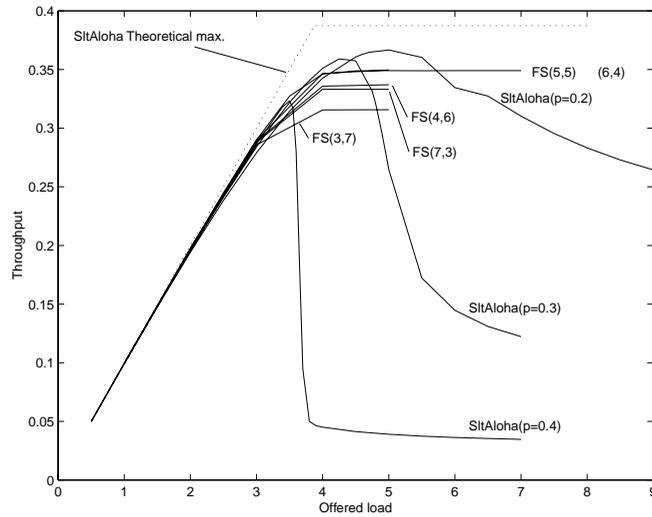Fig. 17. Delay percentiles for $T = 6$, $\lambda = 4.0$.

Fig. 18. Throughput for $T = 10$.

slots, having 20 slots devoted to the contention channel would imply that the network is very lightly loaded (few slots are reserved for the transmission of data traffic in the contentionless channel). In these situations of very low activity the throughput of the contention channel should not be a major concern. However, we can see that in this situation FS-ALOHA does not so clearly outperform slotted ALOHA. For low offered loads ($\lambda = 1.0, 3.0$) FS-ALOHA offers better delay percentiles than slotted ALOHA. The results of slotted ALOHA comparable with those of FS-ALOHA are obtained with very high retransmission probabilities ($p_{tx} = 0.9$), which in turn lead to an unstable behavior (collapse with overloads). FS-ALOHA also outperforms slotted ALOHA when the offered load is very high ($\lambda = 9.0$), as one of the main advantages of FS-ALOHA is its inherent stability in strong overload situations. Nevertheless, for a certain range of offered loads (e.g. $\lambda = 6.0$) slotted ALOHA (with $p_{tx} = 0.7, 0.8$)
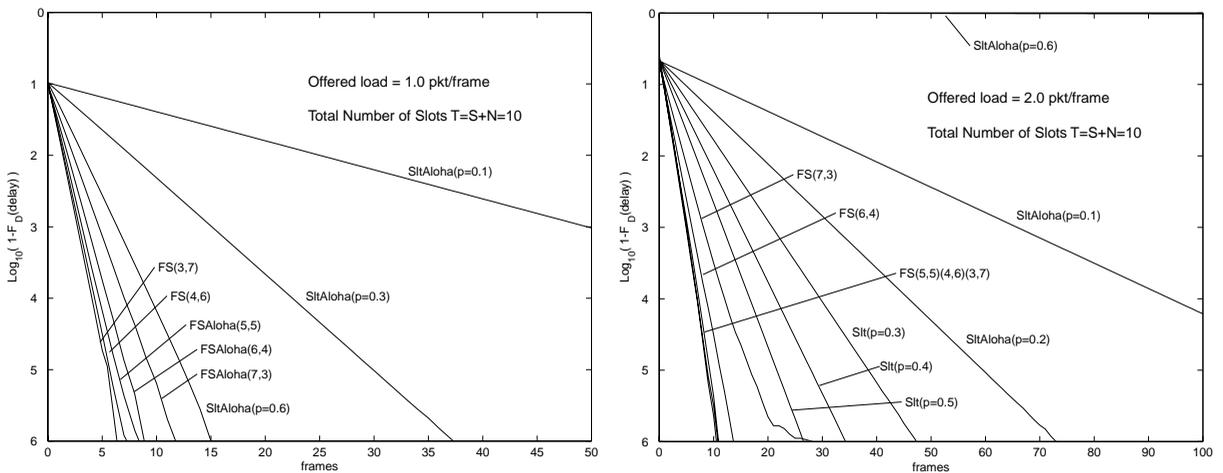

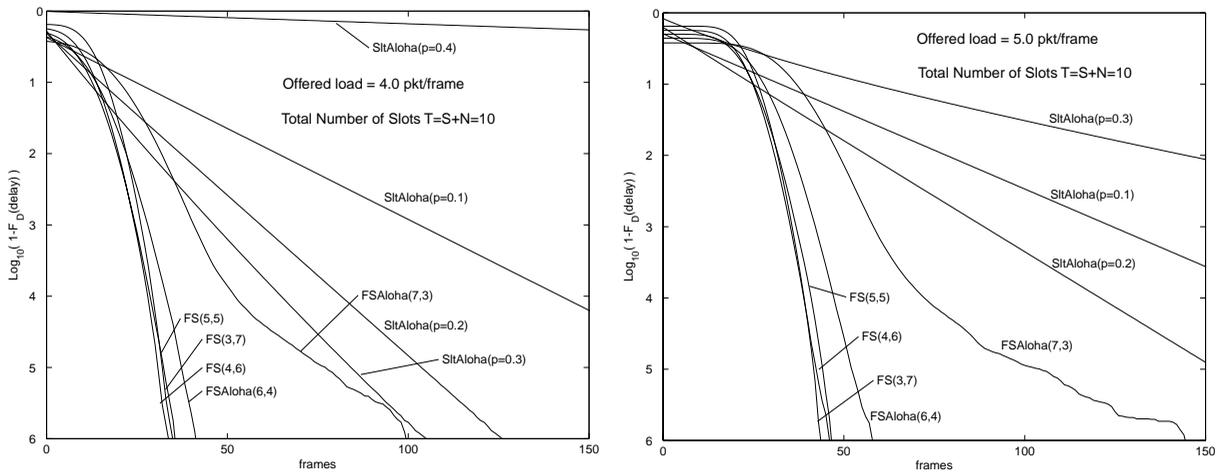Fig. 19. Delay percentiles for $T = 10$, $\lambda = \{1.0, 2.0\}$.

Fig. 20. Delay percentiles for $T = 10$, $\lambda = \{4.0, 5.0\}$.

can show better delay percentiles than slotted ALOHA. The maximum throughput exhibited by FS-ALOHA (Fig. 21) is lower than the obtained with slotted ALOHA, although FS-ALOHA throughput is stable even during highly congested periods and it does not rely on the optimal assessment of a retransmission probability.

It should be noticed that in order to be close to (with low loads) or even outperform (with intermediate loads) FS-ALOHA, the retransmission probability has to be very carefully estimated, as a little deviation from the optimum leads either to poor delay percentiles if a low, conservative value for p_{tx} is chosen or to a network collapse, with very poor results for both throughput and delay (see Fig. 23), if the optimal p_{tx} is slightly overestimated.
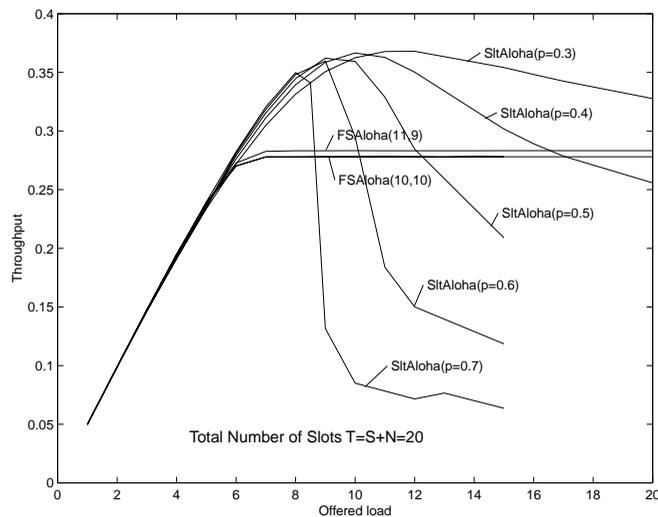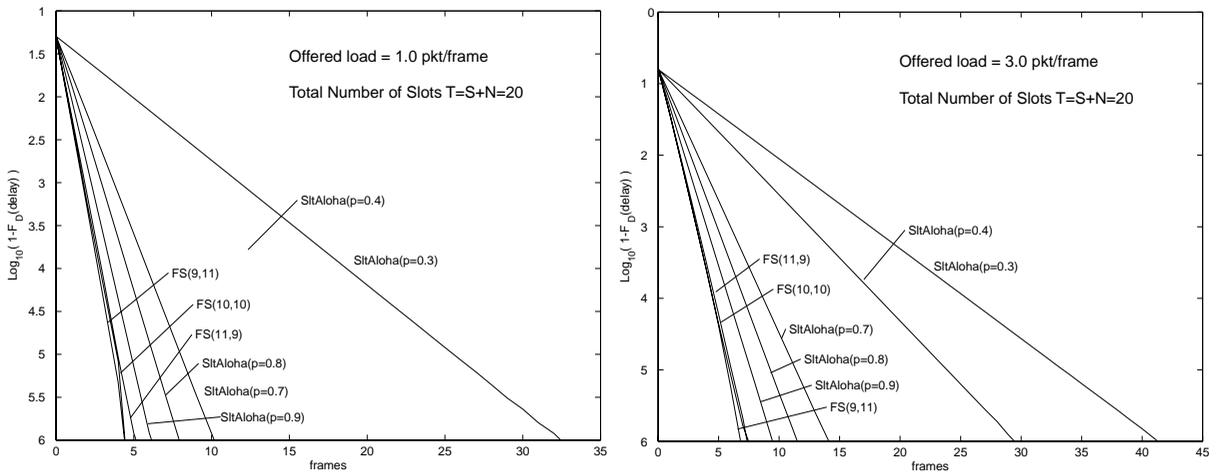

Fig. 21. Throughput for $T = 20$.

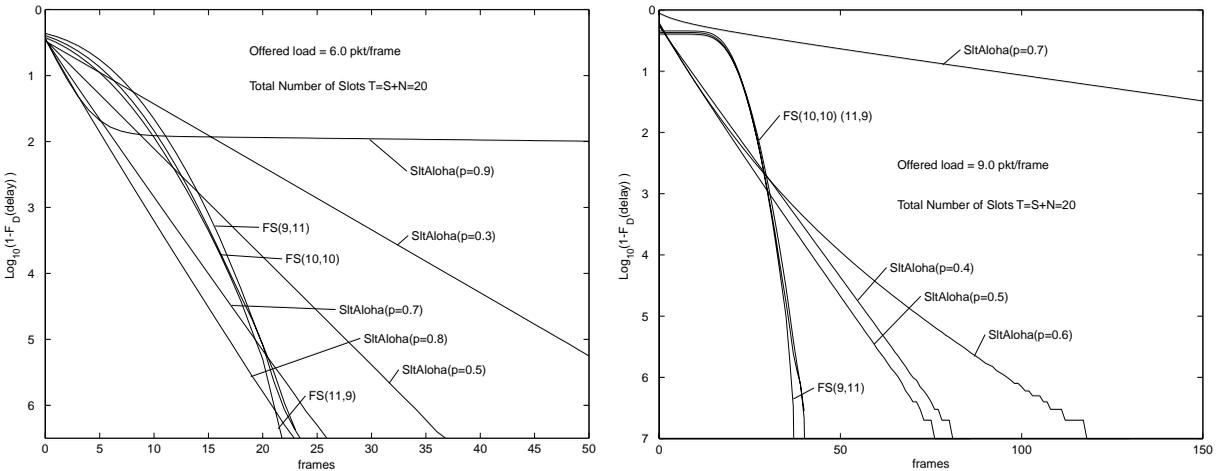Fig. 22. Delay percentiles for $T = 20$, $\lambda = \{1.0, 3.0\}$.

Fig. 23. Delay percentiles for $T = 20$, $\lambda = \{6.0, 9.0\}$.

5. Conclusions

In this paper we have described and analyzed FS-ALOHA, a new Contention Resolution Algorithm to be deployed in the uplink contention channel present in most of the Wireless ATM MAC protocols existing in the literature. The performance achieved in this contention channel may have an important effect on the overall system performance, especially for bursty traffic with delay constraints (e.g. rt-VBR).

An analytical model has been developed for Poisson input traffic and the performance of FS-ALOHA has been compared with slotted ALOHA (mechanism commonly used in the contention period of WATM systems), using different retransmission probabilities in subsequent attempts after initial failures (i.e. p-persistent slotted ALOHA). The complexity of FS-ALOHA is comparable to that of slotted

ALOHA, while showing much better delay percentiles, stability as the throughput does not drop significantly in overload conditions, and higher maximum throughput when the size of the contention period is not too large. Moreover, the maximum throughput achievable with FS-ALOHA is always attainable, as the performance does not rely on any load-dependent optimal estimation of parameters associated to the protocol (such as the retransmission probability in slotted ALOHA).

Acknowledgements

Appendix A

Using the expressions for $P_a(n, \lambda)$ and $P_d(n, q, N)$ given in Eqs. (1) and (2) we have

$$\boldsymbol{B}_{00} \in \boldsymbol{M}_{\mathbf{1} \times \mathbf{1}}, \quad \boldsymbol{B}_{00}(1, 1) = P_a(0, \lambda) + P_a(1, \lambda) + \sum_{j=2}^{q_m} P_a(j, \lambda) P_d(j, j, T), \tag{A.1}$$

$$\boldsymbol{B}_{01} \in \boldsymbol{M}_{\mathbf{1} \times \boldsymbol{q}_m - \mathbf{1}}, \quad \boldsymbol{B}_{01} = (b_{01}(1, q))_{1 \leq q \leq q_m - 1},$$
$$b_{01}(1, q) = \sum_{j=q+1}^{q_m} P_a(j, \lambda) P_d(j - q - 1, j, T), \tag{A.2}$$

$$\boldsymbol{B}_{10} \in \boldsymbol{M}_{\boldsymbol{q}_m - \mathbf{1} \times \mathbf{1}}, \quad \boldsymbol{B}_{10} = (b_{10}(q, 1))_{1 \leq q \leq q_m - 1},$$
$$b_{10}(q, 1) = (1 - p_{\text{TSgen}}) P_d(q + 1, q + 1, N), \tag{A.3}$$

where, according to Eq. (3)

$$p_{\text{TSgen}} = \Pr(\text{TSgen} \wedge \text{ nonempty TS queue}) = \sum_{n=0}^{\infty} P_a(n, \lambda) (1 - P_d(n, n, S)), \tag{A.4}$$

$$\boldsymbol{A}_0 \in \boldsymbol{M}_{\boldsymbol{q}_m - \mathbf{1} \times \boldsymbol{q}_m - \mathbf{1}}, \quad \boldsymbol{A}_0 = \begin{pmatrix} a_0(1, 1) & 0 & 0 & \dots & 0 \\ a_0(2, 1) & a_0(2, 2) & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & 0 & \vdots \\ a_0(q_m - 1, 1) & a_0(q_m - 1, 2) & \dots & a_0(q_m - 1, q_m - 1) \end{pmatrix}, \tag{A.5}$$
$$a_0(q, q') = p_{\text{TSgen}} P_d(q - q', q + 1, N), \quad q' \leq q,$$

$$\boldsymbol{A}_1 \in \boldsymbol{M}_{\boldsymbol{q}_m - \mathbf{1} \times \boldsymbol{q}_m - \mathbf{1}}, \quad \boldsymbol{A}_1 = (a_1(q, q'))_{1 \leq q \leq q_m - 1, 1 \leq q' \leq q_m - 1},$$
$$a_1(q, q') = P_{\text{arr}}(q' + 1, \lambda, S) P_d(q + 1, q + 1, N) + (1 - p_{\text{TSgen}}) P_d(q - q', q + 1, N), \tag{A.6}$$

$$A_2 \in M_{q_m-1 \times q_m-1}, \quad A_2 = \left(a_2(q, q')\right)_{1 \leq q \leq q_m-1, 1 \leq q' \leq q_m-1},$$

$$a_2(q, q') = P_{arr}(q' + 1, \lambda, S) P_d(q + 1, q + 1, N) \frac{1 - p_{TSgen}}{p_{TSgen}}, \tag{A.7}$$

and

$$P_{arr}(n, \lambda, S) = \sum_{j=n}^{q_m} P_a(j, \lambda) P_d(j - n, j, S) \tag{A.8}$$

is the probability of getting a new Transmission Set with n requests when the TS queue is not empty.

Appendix B

We will start with the calculation of the set of functions $f_A(q, N)[n]$, with $q \in \{0, 2, 3, \ldots, q_m\}$. If $q = 0$, which is only possible for the state $(0,0)$, then there is no delay associated and therefore

$$f_A(0, N)[n] = \delta[n]. \tag{B.1}$$

If $q \geq 2$ then the delay densities $f_A(q, N)[n]$ verify the following recursion:

$$f_A(q, N)[n] = \sum_{i=\max(0, q-N), i \neq 1}^{q-1} P_d(q - i, q, N) f_A(i, N)[n - 1] + P_d(0, q, N) f_A(q, N)[n - 1], \tag{B.2}$$

with $P_d(n, q, N)$ defined in Eq. (2). Solving the expression in Eq. (B.2) for $q = 2$ we obtain

$$f_A(2, N)[n] = P_d(2, 2, N) [P_d(0, 2, N)]^{n-1} u[n - 1]. \tag{B.3}$$

For $q > 2$, the functions $f_A(q, N)[n]$ can be calculated from the previous $f_A(q', N)[n]$, $q' < q$ and using Eq. (B.2),

$$f_A(q, N)[n] = \begin{cases} 0, & n \leq 0, \\ P_d(q, q, N), & n = 1, \\ \displaystyle\sum_{i=\max(2, q-N)}^{q-1} P_d(q - i, q, N) f_A(i, N)[n - 1] + P_d(0, q, N) f_A(q, N)[n - 1], & n \geq 2. \end{cases} \tag{B.4}$$

The set of functions $f_B(G, \lambda, N)[n]$ with $G \in \{0, 1, \ldots\}$ corresponds to the service time of $\{0, 1, \ldots\}$ Transmission Sets which are independent from each other and at least have two requests. Let $f_{TS}(\lambda, N)[n]$ be the delay density function associated to the service of one of these TSs

$$f_{TS}(\lambda, S, N)[n] = \frac{1}{p_{TSgen}} \sum_{q=2}^{q_m} P_{arr}(q, \lambda, S) f_A(q, N)[n]. \tag{B.5}$$

And thus

$$f_B(G, \lambda, S, N)[n] = \begin{cases} \delta[n], & G = 0, \\ \underbrace{f_{TS}(\lambda, N)[n] * \ldots * f_{TS}(\lambda, N)[n]_G}_{G}, & G \geq 1. \end{cases} \tag{B.6}$$

Finally, t_C corresponds to the delay experienced by the observer due to the other requests present in its Transmission Set, given that the system was not empty when the observer arrived. We will calculate the delay density function by conditioning on the number of requests conveyed in the Transmission Set. The same applies for t_{CE}, excepting that the system was empty (no TSs in the queue) when the observer arrived.

$$
\begin{aligned}
f_C(\lambda, S, N)[n] &= \sum_{q=2}^{q_m} P_{\text{obs}|\text{NonEmpty}}(q, \lambda, S) \cdot f_{C|q}(q, N)[n], \\
P_{\text{obs}|\text{NonEmpty}}(q, \lambda, S) &= \sum_{j=q}^{q_m} P_a(j-1, \lambda)\frac{n}{j} P_d(j-q, j, S),
\end{aligned}
\tag{B.7}
$$

$$
\begin{aligned}
f_{CE}(\lambda, S, N)[n] &= \sum_{q=2}^{q_m} P_{\text{obs}|\text{Empty}}(q, \lambda, S) \cdot f_{C|q}(q, N)[n], \\
P_{\text{obs}|\text{Empty}}(q, \lambda, S) &= \sum_{j=q}^{q_m} P_a(j-1, \lambda)\frac{n}{j} P_d(j-q, j, T).
\end{aligned}
\tag{B.8}
$$

But the set of functions $f_{C|q}(q, N)[n], 2 \le q \le q_m$, can be easily determined as they satisfy the following recursion:

$$
f_{C|q}(q, N)[n] = \left(\sum_{i=\max(0,q-N), i\neq 1}^{q-1} \frac{q-i}{q} P_d(q-i, q, N) \right) \delta[n-1]
$$

$$
+ \sum_{i=\max(2,q-N)}^{q-1} \frac{i}{q} P_d(q-i, q, N) f_C/i(i, N)[n-1] + P_d(0, q, N) f_C/q(q, N)[n-1], \quad \text{(B.9)}
$$

and if, again, we solve the former expression in Eq. (B.9) for $q = 2$ we obtain that

$$
f_C/2(2, N)[n] = f_A(2, N)[n] = P_d(2, 2, N) [P_d(0, 2, N)]^{n-1} u[n-1].
\tag{B.10}
$$

For $q > 2$, analogously as with $f_A(q, N)[n]$, the functions $f_C/q(q, N)[n]$ can be calculated from the previous $f_C/q'(q', N)[n]$, $q' < q$ and using the recursion in Eq. (B.9).

$$
f_{C|q}(q, N)[n] = \begin{cases}
0, & n \le 0, \\
\displaystyle\sum_{i=\max(0,q-N), i\neq 1}^{q-1} \frac{q-i}{q} P_d(q-i, q, N), & n = 1, \\
\displaystyle\sum_{i=\max(2,q-N)}^{q-1} \frac{i}{q} P_d(q-i, q, N) f_C/i(i, N)[n-1] \\
\qquad + P_d(0, q, N) f_C/q(q, N)[n-1], & n \ge 2.
\end{cases}
\tag{B.11}
$$

$$
\mu_{\text{TS}}(\lambda, N, S) = E\left[f_{\text{TS}}(\lambda, N)[n] \right].
\tag{B.12}
$$

with $f_{\text{TS}}(\lambda, N)$ as defined in Eq. (B.5). The stability of the mechanism is determined by the speed at which TSs are generated (S dependent) and the time needed to process them (N dependent).

References

[1] P802.11 Draft Standard for Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification, IEEE P802.11, D6.1, May 1997, AD165.

[2] Radio Equipment and Systems (RES); HIgh PErformance Radio Local Area Network (HIPERLAN) Type I; Functional Specification, European Telecommunication Standard Final Draft ETS 300 652.

[3] SAMBA: System for Advanced Mobile Broadband Applications, Specification of the Air Interface, CEC Deliverable Number: A0204/DB/F3F/DS/P004/a1.

[4] C. Apostolas, R. Tafazolli, B. Evans, Wireless ATM LAN, in: Proc. PIMRC'95, Toronto, ON, pp. 773–777.

[5] D. Bertsekas, R. Gallager, Data Networks, Prentice-Hall, Englewood Cliffs, NJ, 1987.

[6] H. Chou, C. Lee, K. Chen, Group randomly addressed polling with reservation for wireless integrated service network, in: Proc. PIMRC'95, Toronto, ON, pp. 618–622.

[7] D. Vázquez-Cortizo, J. García, A collision resolution algorithm for ad-hoc wireless LANs, in: IFIP BC'98, International Conference on Broadband Communications, Stuttgart, 1998, pp. 119–130.

[8] M.J. Karol, Z. Liu, K. Eng, Distributed-queueing request update multiple access (DQRUMA) for wireless packet (ATM) networks, in: International Conference on Communications (ICC'95) Conference Record, 1995, pp. 1224–1231.

[9] J. Mikkonen, J. Aldis, G. Awater, A. Lunn, D. Hutchison, The magic WAND — functional overview, IEEE J. Selected Areas Commun. 16 (6) (1998) 953–972.

[10] M.F. Neuts, Matrix-Geometric Solutions in Stochastic Models, An Algorithmic Approach, The John Hopkins University Press, Baltimore, MD, 1981.

[11] P. Olanders, DECT standardization — status and future activities, in: Proc. PIMRC-WCN'94, 1994, The Hague.

[12] K. Pahlavan, A.H. Levesque, Wireless Information Networks, Wiley Series in Telecommunication and Signal Processing, 1995.

[13] D. Petras, Medium access control protocol for transparent ATM access in MBS, in: RACE Mobile Telecommunication Summit, Cascais, Portugal, 1995.

[14] P. Smulders, C. Blondia, Application of the asynchronous transfer mode in indoor radio networks, Proc. PIMRC'94, The Hague, pp. 839–843.

[15] K. Sriram, Performance of MAC protocols for broadband HFC and wireless access networks, Adv. Performance Anal. 1 (1) (1998) 1–37.

[16] W.J. Stewart, Introduction to the Numerical Solution of Markov Chains, Princeton University Press, Princeton, NJ, 1994.

David Vázquez-Cortizo obtained his degree in Electrical Engineering from E.T.S.E.T., University of Vigo, Spain, in 1995. In 1995 he joined the Electronics and Systems Department at University of La Coruña, Spain, as an assistant teacher. He was a guest researcher in Helsinki University of Technology for six months in 1997–1998. In 1998 he joined the Performance Analysis of Telecommunication Systems group at University of Antwerp (UIA), Belgium. His research interests are focused on the design and performance evaluation of wireless MAC protocols for high-capacity multimedia networks.

Jorge García Vidal obtained his Telecommunication Engineering degree (1988) and Ph.D. degree in Telecommunication Engineering (1992) from Polytechnic University of Catalonia (UPC, award to the best Ph.D. thesis 1992). During 1992–93 he visited SIE Department of University of Arizona with a NATO fellowship. He is Associate Professor in the Computer Architecture Department of UPC. He has participated in several projects related with computer networks (RACE 1022, BAF, COST 242, COST 257). His research interest include wireless LANs, TCP/IP implementation issues and performance evaluation of computer systems.

Chris Blondia obtained his Master in Science and Ph.D. in Mathematics, both from the University of Ghent (Belgium) in 1977 and 1982, respectively. In 1983 he joined Philips Belgium, where he was a researcher between 1986 and 1991 in the Philips Research Laboratory Belgium in the group Computer and Communication Systems. Between August 1991 and end 1994 he was an Associate Professor at the Computer Science Department of the University of Nijmegen (The Netherlands). In 1995 he joined the Department of Mathematics and Computer Science of the University of Antwerp, where is currently a professor and head of the research group 'Performance Analysis of Telecommunication Systems'. He is lecturing Telecommunication and Performance Evaluation courses. His main research interests are related to mathematical models (stochastic modeling, queueing theory) for performance evaluation of communication systems and the impact of the performance on the architecture of these systems. In this context the areas of research are performance issues related to switching architectures, traffic control functions, access control protocols. He has published a substantial number of papers in international journals and conferences on these research areas. He has been member of many program committees of international conferences and is editor of the Journal of Network and Computer Applications. He is member of IFIP W.G. 6.3 on 'Performance of Computer Networks'.

Benny Van Houdt graduated from the University of Antwerp in 1997 in Mathematics and Computer Science. Currently he holds an assistant position at the University of Antwerp (UIA) and is an active member of the Performance Analysis of Telecommunication Systems (PATS) research group. His main research interest goes to the performance evaluation and stochastic modeling of wireless and mobile communication networks.