Optimal Retransmission Probability for S-ALOHA Under the Infinite Population Model

Mario E. Rivero-Angeles
UPIITA-IPN. Telematic Department
Mexico City, Mexico
mrivero@cinvestav.mx

Domínguez Lara-Rodriguez
CINVESTAV-IPN. Electrical Engineering Department.
Communications Section
Mexico City, Mexico
dlara@cinvestav.mx

Felipe A. Cruz-Pérez
CINVESTAV-IPN. Electrical Engineering Department.
Communications Section
Mexico City, Mexico
facruz@cinvestav.mx

Abstract — In this paper, a mathematical analysis method to simultaneously evaluate throughput and access delay considering an infinite population model is considered. Most of the previous related research has been done considering both finite population and saturation conditions where all the nodes in the system have always a packet ready to be transmitted and the transmission queue of each station is assumed to be always nonempty. This assumption is a good approximation for a Local Area Network working at full capacity; however, in a cellular system the assumptions of finite population and that every node in a cell has always a packet to transmit is not very realistic. Here, analytical results considering a S-ALOHA random access protocol with a Poisson arrival process – more suitable for the traffic model in a cellular system – for the users in the cells is presented. Using the Geometrical Backoff (GB) strategy, two approaches to find the optimum retransmission probabilities are developed; in the first one, the number of backlogged packets is required while a simpler and efficient alternative method requires only the knowledge of the new packet arrival rate.

Keywords— S-ALOHA, Geometric Backoff, throughput analysis, optimum retransmission probability.

I. INTRODUCTION

ALOHA random multiple access are widely used in data and cellular networks as the protocol for the random access phase because it is the most simple protocol, since it does not require any information about the channel conditions, and implementation is straightforward [1]-[3]. The users that arrive to the system trying to reserve a data channel simply wait for the beginning of the next time slot and transmit their packet. If more than one user attempts to transmit a packet then a collision will occur with probability 1 (capture effect is not considered in this paper). If only one packet attempts transmission in a time slot then it is received without any errors and the base station replies to the mobile with a positive acknowledge (ACK). For high offered loads, the amount of collisions increases drastically due to the lack of a central access coordination entity and the random nature of the arrivals. This means that when traffic load is high, the probability of collision is also high and none of the users in the system is able to transmit the registration packet successfully to the base station, hence very few users can actually have access to the data channels, rendering a very low average data rate for the system. In these conditions, the random access protocol presents an undesirable operation behavior [4], [5]. Traffic load is due to new packets arriving to the system and backlogged packets on retransmission. Hence, if the new packet arrival rate is low enough, it is possible to maintain stability by setting an appropriate value of the retransmission probability. The value of the retransmission probability is set by the Backoff protocol, which determine the rules that every node has to follow whenever a collision happens [6]-[10]. Basically, whenever a packet suffers a collision, it has to wait for a random time, whose distribution is specified by the backoff protocol, before it is transmitted again.

In this paper, a throughput analysis for a Geometrical Backoff (GB) protocol in order to achieve stability is presented. In our model, an infinite population model where new packet arrivals follow a Poisson process is considered. This new packet arrival rate remains constant, hence it does not depend on the number of backlogged packets. The backlogged packets retransmit according to the GB policy generating a transmitted packets arrival rate which completely depends on the retransmission probability. This probability is set accordingly in order to maximize throughput. In our previous work [11], [12] a conventional infinite population model was considered where the exact value of the traffic load due to the backlogged packets was not known; hence the retransmission probability did not have a direct impact on the total arrival rate. The access delay also depends on the value of the retransmission probability, hence a delay analysis is also performed. Although there are many other backoff algorithms, most of them are proposed for the IEEE 802.11 WLAN standard [5], [7], [13]-[18]. They are not considered because the focus in this paper is on S-ALOHA random access protocol and they consider a finite population model where the total arrival rate depends on the number of backlogged stations and the no buffering condition (the condition without queue) is used.

II. S-ALOHA SYSTEM MODEL

In the literature it is common to study the problem of stability from two different perspectives: throughput [6], [7], [25] (where an infinite population model is considered) and access delay point of view [5], [19], [20] (where a finite population model is considered, usually in saturation conditions). From the access delay perspective a random
access protocol is stable if the average access time (from the moment the packet is generated until it is successfully received) is not infinite. From the throughput perspective the system is stable if the throughput does not collapse to zero when the total traffic load increases [4]. In this paper, the throughput based definition of stability is considered.

The finite population model usually considers the no buffering assumption [7], [19], [20], where traffic load due to new packets is a function of the number of users in the backlogged state. Additionally, saturation conditions are considered where each node in the network has always a backlogged packet and the new packet arrival rate is zero. This model is more suitable for Local Area Networks environment but in a cellular system where nodes have a certain level of movement entering and leaving the cell it is not adequate a new packet arrival rate of zero. Additionally, it is not feasible to consider that all nodes always have a packet ready to be transmitted since it is known that even at the peak traffic loads hours there is a utilization of 5% to 15%. Then the considered model corresponds to an infinite population model with a finite number of packets in the backlog state, as expected in a real cellular system.

The infinite population model considered in this paper is suitable to model practical cellular systems. It has been observed that during some time intervals, traffic arrivals in data networks tends towards Poisson as the traffic loads increases [22]. The assumption of Poisson arrivals for all the packets (new plus retransmitted) is unrealistic because collided packets are retransmitted according to the particular backoff policy. However, if the retransmissions from the backlogged nodes are sufficiently randomized, it is plausible to approximate the total number of retransmissions and new transmissions in a given slot as a Poisson variable with a total traffic load $G$, which corresponds to the sum of the new packet arrival rate $G_N$ and the collided packets arrival rate $G_R$, then $G > G_N$. With this approximation the probability of a successful transmission in a slot is $G e^{-G}$ as it was done in [6], [11], [12]. The problem with this simple approach is that the specific values of $G_N$ and $G_R$ are not known, and the only parameter used for the analysis is $G$. Hence, this model does not provide any insight into the dynamics of the system because the value of $G$ is neither related to the particular backoff policy nor to the particular parameters of that policy such as the retransmission probability in the GB policy. In a real system, the number of backlogged packets is time-variant, therefore, the value of $G$ also changes; this leads to a feedback effect, generating further changes in the number of backlogged packets. These changes are not reflected on the infinite model with $G$ constant. Under the infinite population model, if the arrival rate exceeds a certain value, the number of backlogged packets would grow without limit since the departure rate would be lower than the arrival rate, then the access delay of packets would also grow with no bound [2].

Based on the fact that under light and moderate traffic load, the probability that a high number of packets being simultaneously backlogged is not significant, a maximum limited number of packets in backlog state is considered. Then, a finite Markov chain is employed to approximately analyze the system performance. Simulation results are used to validate the results of this mathematical approach. A similar model to that presented in [18], [23], [25] is used, where an infinite population of terminals is considered and generate a Poisson process for the new packet arrivals with rate $G_N$. This arrival rate remains constant during the busiest hours of the day. Also, it is assumed that the active terminals collectively generate new packet arrivals according to a Poisson process with mean rate $G_N$ and this rate is not decreased with the number of collided packets; that is, an infinite buffer is considered at each node in the system. Hence, the total traffic which is offered to the channel consists of that from both the retransmitted and new packets. It has been proven in [9] that any random access protocol that uses a fix value for the retransmission probability, i.e., that does not varies according to the traffic conditions of the system, is unstable. Additionally, it is possible to find a value for the retransmission probability that maintains the system stable for small values of the traffic load. Hence, a dynamic value for the retransmission probability that maximizes the throughput while maintaining the system stable is calculated. The retransmission probability is chosen to be moderately large for small arrival rates, so as to avoid large delays after collisions and if the arrival rate is high the retransmission probability should be small in order to avoid further collisions.

It is considered that all packets have the same number of bits and each node takes $T$ seconds to transmit one packet. Packets that have suffered at least one collision are retransmitted in the next slot with probability $r$ until they are successfully received by the base station, hence the backlogged packets will be retransmitted in a random time after the collision occurred. The value of $G_N$ is not decreased with the number of collided packets.

It is assumed that there could be as much as $M$ simultaneously backlogged packets in the system, hence the packet population waiting to be transmitted is finite. The mathematical model implies that if a new packet suffers a collision and there are already $M$ backlogged packets, then this new packet would be discarded, i.e., in the mathematical analyses, the effect of more than $M$ backlogged packets on the random access channel is neglected whereas in the simulation results there are not discarded packets. The value of $M$ is set large enough such that the analytical results are close to the simulation results, where no packets are discarded. Then, the analytical model corresponds to an approximation of a practical system. For low values of $M$, the retransmitted traffic load, $G_R$, is also low and the analytical throughput is also lower than the simulation results. For high values of $M$, there is no variation in the value of the total arrival rate, hence, the analytical throughput remains close to
the simulation results. However, the processing time is incremented considerably. This model allows us to analyze both throughput and access delay simultaneously because now both the retransmission probability and the backoff policy, have a direct impact on the value of the total arrival rate \( G \). This model differs from the conventional infinite population model where the value of the total arrival rate is assumed to be constant regardless of the backoff policy and the retransmission probability. In order to have stability conditions the new packet arrival rate \( (G_N) \) has to be less than 0.36 new packets per time slot because the maximum throughput is 0.36. Otherwise, the incoming packets would be much more than the outgoing packets, entering in an unstable operation. The backoff delay after \( i \) collisions is \( W_i \). Then the \( i^{th} \) retransmission takes place after \( W_i \) time slots from the moment of the \( i \)th collision. The probability distribution of this random time is determined according to the GB policy: \( W_i \) is geometrically distributed with probability \( \tau \).

III. THROUGHPUT ANALYSIS

The following analysis is based on the throughput analysis for finite population found in [3]. However, our analysis differs to that of [3] because the new traffic load is considered to be constant regardless of the number of nodes with collided packets. That is, it is considered that the number of backlogged packets in the cell is finite but the total number of nodes is infinite. Thus, the arrival rate of new packets is not decremented according to the number of backlogged packets. Then, for infinite population where \( M \) backlogged packets can be in the system (the value of \( M \) has to be high enough in order to closely approximate the behavior of a system where there is no limit in the number of backlogged packets), let every packet be in one of two states: New: The packets that have never been transmitted before, hence they have not suffered any collision. New packets arrive according to a Poisson process with constant arrival rate of \( \lambda \).

Let the number of slots be numbered sequentially and \( N(k) \) represents the number of packets in the backlogged state at the beginning of the \( k \)th slot. \( N(k) \) depends on the number of packets in the \( (k-1)\)th slot and the number of packets that moved from the new state to the backlog state or the number of backlogged packets successfully transmitted within that slot. This system can be modeled using a Markov chain as that depicted in Figure 1 where the random variable \( N(k) \) corresponds to the state of the chain. State transitions upward are possible between all states to any superior state, except from state 0 to state 1 since in any collision there is at least two packets involved. Downward transitions are only possible between neighbor states since there can only be one successful transmission per time slot. Let \( \pi_i \) the steady state probability of finding the system in state \( i \), i.e.:

\[
\pi_i = \lim_{k \to \infty} P[N(k) = i] \quad (1)
\]

and, the transition probability form state \( i \) to state \( j \), i.e.:

\[
p_{ij} = \lim_{k \to \infty} P[N(k) = j | N(k-1) = i] \quad (2)
\]

The transition probabilities for the different cases depicted in Figure 1 are derived, given by:

\[
p_{ij} = \begin{cases} 
\frac{\lambda}{M} & \text{if } j = i-1 \\
\frac{\lambda}{M} \times (1 - \tau)^i \times e^{-\lambda} & \text{if } j = i \\
\frac{\lambda}{M} \times (1 - \tau)^{i-1} \times e^{-\lambda} \times (1 - \tau) & \text{if } j = i+1 \\
0 & \text{otherwise}
\end{cases}
\]

Taking the probabilities in (3), the Markov chain depicted in Figure 1 can be solved using:

\[
\pi P = \pi; \sum_{i=0}^{M} \pi_i = 1 \quad (4)
\]

where \( \pi = [\pi_0, \pi_1, \pi_2, ..., \pi_M] \) and \( P \) is the probability transition matrix with elements \( p_{ij} \).

Figure 1. Markov Chain of the S-ALOHA System.

Then the steady state probabilities are found by proposing a co-lineal vector \( x = [x_0, x_1, ..., x_M] \) with \( x_0 = 1 \) and solving:

\[
x (1 - P) = 0 \quad (5)
\]

where \( I \) is the Identity Matrix. Then, (4) is solved recursively and it is found to be:

\[
x_i = \frac{1}{w - e^{-\lambda \tau} (1 + G_N) x_0} \quad i = 1
\]

\[
x_i = \frac{1}{2\tau (1 - \tau)^{i-1} e^{-\lambda \tau} (1 - \tau + (1 - \tau) G_N) x_i} \quad i = 2
\]

\[
x_i = \frac{1}{i\tau (1 - \tau)^{i-1}} \left[ \left( e^{-\lambda \tau} - 1 \right) x_{i-1} - G_N x_{i-2} - \sum_{k=2}^{i} \frac{G_N^{k-1} \tau^{i-k}}{k!} \right]
\]

\[
+ \left[ \frac{(i-1)\tau}{(1 - \tau)} - G_N \right] x_{i-1} + \frac{G_N}{(1 - \tau)} x_{i-2} \quad 3 \leq i \leq M
\]
Now, the steady state probability $\pi_i$ can be found as:

$$\pi_i = \frac{x_i}{\sum_{j=0}^{M} x_j} \quad ; \quad i = 0, 1, 2, \ldots, M$$

Finally, the throughput is equivalent to the probability that one packet is transmitted in a time slot and this probability depends on the number of packets in the backlogged state, then:

$$S = \sum_{i=0}^{M} S(i)\pi_i$$

The probability of having only one transmission in a time slot when there are $i$ packets in the backlog state is:

$$S(i) = [G_N(1-\tau) + i\tau]e^{-G_N(1-\tau)}$$

And the traffic load due to the backlogged packets is:

$$G_R = \sum_{i=0}^{M} i\pi_k$$

Total traffic load is the sum of the new packets arrival rate plus the backlogged packets arrival rate, then $G_T = G_N + G_R$.

Two different approaches for the optimization of the retransmission probability are proposed:

A) Since the throughput depends on the number of collided packets, then, the retransmission probability should also depend on the number of potential retransmissions in a time slot. Then, differentiating (10) and equating to zero, the optimal retransmission probability when there are $i$ collided packets in the system is given as:

$$\tau_{opt}(i) = \frac{1-G_N}{i - G_N}$$

By using the optimum retransmission probability, the throughput is kept at the same value as the total arrival rate, i.e., the input packet rate in the system is the same as the total output rate. The difficulty of this method is that the value of the number of backlogged packets, $i$, is not easily known and is necessary to use an estimation algorithm such as the one presented in [25]. An alternative method to obtain the retransmission probability that achieves an acceptable performance is presented next.

B) Considering that in order to maximize throughput, the total arrival rate has to be limited to one packet (new or backlogged) per time slot, then:

$$G_T = G_N + \tau \sum_{i=0}^{M} i\pi_k = 1$$

Then, solving for $\tau$:

$$\tau_{opt}(i) = \frac{1-G_N}{i - G_N}$$

Since this alternative method considers an average value of the traffic load due to the backlogged packets, the retransmission probability would be lower than it should be when the instantaneous retransmission rate is higher than the average value, generating a higher collision probability in the system. Conversely, when the actual retransmission rate is lower than the average value, then the calculated retransmission probability would be higher than it should be and a large number of idle slots is generated. However, in average, the total output rate is very close to the total input rate, especially for low arrival rates. The advantage of this method is that several works in the literature have address the subject of estimation algorithms for the new packet mean arrival rate such as [11] and [25] and there is no need to estimate the instantaneous number of backlogged packets.

IV. ACCESS DELAY ANALYSIS

In [6], an expression for the cumulative distribution function (cdf) of the delay, $d$, was found as:

$$F_D(d) = \sum_{k=1}^{d} \sum_{m=1}^{k} P(X_k = m)P(R = k)$$

where $X_k$ is a random variable that represents the backoff interval in the $k^{th}$ retransmission and:

$$P[R = k] = \rho_{success}^{k}(1 - \rho_{success})$$

However in [6], the probability of success was considered to be $e^{-GT}$ and $GT$ was considered to be insensitive to the value of the retransmission probability. Considering the infinite population model in Section II, the success probability is found as:

$$P[No other transmissions|One Packet has transmitted]$$

and

$$P_{success} = \sum_{k=0}^{M} e^{-G_N(1-\tau)^{-1}(2-\tau)\pi_i}$$

Finally we found in [24] an expression for $P[X_k = n]$ for the GB policy.

V. NUMERICAL RESULTS

In this section numerical results for the infinite population model considered is presented. Firstly, the infinite population model is verified through simulations in order to find the appropriate value of $M$ for which close results to the case when no limit in the number of simultaneous backlogged packets is considered are achieved. Then, throughput and access delay results are shown. In Figure 2, the traffic load due to the backlogged packets for different values of the new packet arrival rate for $\tau = 0.0001$ is shown. For low new packet arrival rates, the value of $M$ to achieve a good
approximation to the simulations (where no limit on the number of simultaneous backlogged packets is considered) can be low (in the order of 100) since the probability of finding a high number of simultaneous backlogged packets is low. However, for high new packet arrival rates, low values of \( M \) do not allow to achieve a good approximation to the simulations because the collision probability is higher and the maximum number of simultaneous backlogged packets is incremented; then, a high value of \( M \) should be considered (in the order of 9000). The benefit of using a low value of \( M \) for low new packet arrival rate is much less processing time for the analytical results. The value of \( M \) impacts directly on the arrival rate due to the backlogged packets which is incremented accordingly.

Even if in a practical cellular system it is not probable to have 9000 simultaneous backlogged packets, due to the sizes of the cells, it is important to consider a maximum value for \( M \) in order to illustrate the analysis methodology, even for high arrival rates, i.e., when the conventional infinite population model considers a high arrival rate it implicitly considers a very high number of potential backlogged packets. Since in a cellular system, new packet arrivals follow a Poisson process suitable for a very high number of potential users with a relative low activity each user, then the total number of backlogged packets is not expected to be close to 9000. In Figure 3, the analytical result are compared with simulation results with \( \tau = 0.0001 \) and different values of the maximum number of simultaneous backlogged packets in the system, \( M \). In the simulations a limit on the backlogged packets is not considered. The higher the new packet rate, the analytic results approximate the simulation results for high values of \( M \) since more packets have to be considered for the retransmission rate. If \( G_N \) is sufficiently low, then the value of \( M \) can be also small. When \( M = 9000 \) the analytical results coincide with the simulation results for any value of \( G_N \) shown in the figure. Then, an approximation considering that the probability to find more than 9000 backlogged packets in the system is very low and do not have a big impact on the total arrival load is considered. Then, for the remaining results, the value of \( M = 9000 \) is used unless otherwise stated. Throughput obtained using the optimum retransmission probability and the alternative retransmission probability is plotted in Figure 4. For both retransmission probabilities, throughput is very close to the input arrival rate. Hence, most of the packets that arrive to the system are successfully transmitted. For moderate arrival rates (up to 0.35 packets per time slot) the optimal retransmission probability is higher than the alternative retransmission probability since the first considers the instantaneous number of backlogged packets. For the alternative retransmission probability, the output packet rate is not always equal to the input packet rate as in the case of the optimum retransmission probability derived earlier, but the practical implementation of the alternative retransmission probability is much easier.

Results for the 95 percentile are also shown. In Figure 5, the access delay at the 95 percentile for the GB policy and with a constant value of \( \tau = 0.0001 \) and for different values of \( M \) is presented. It can be seen that for a low value of \( M \) there is a good approximation to the simulation results for low values of \( G_N \), but in order to achieve a good approximation for higher values of \( G_N \) then \( M \) has to be on the order of 9000. It is observed that when \( M \) is low for high new packet arrival rates, there is a low number of retransmissions, rendering a low value of the access delay and a linear increase can be seen, showing a poor approximation to the simulation results. However, as the value of \( M \) is increased, an exponential grow is detected closer to the simulation results.

**CONCLUSIONS**

The conventional infinite population model has the advantage of being simple but it does not provide any insight on the dynamics of the system. The mathematical analysis methodology in this paper considers that the new packets arrive according to a Poisson process with a constant rate
while there is a maximum number of simultaneous backlogged packets that generate a retransmission arrival rate. For a sufficiently high value of M, the analytical results are very close to the simulation results where no limit on the number of backlogged packets were considered. Two approaches to find the optimum retransmission probabilities are developed; in the first one, the number of backlogged approaches to find the optimum retransmission probabilities different retransmission policies. The disadvantages of this model are: In order to establish the value of the retransmission probability it is necessary to know the value of G_N which can be difficult to obtain, but it can be estimated using some of the proposed algorithms in the literature.

Figure 4. Average Throughput under the optimum retransmission probability (\( r_{opt} \)) and the alternative retransmission probability (\( r_{alt} \)).

Figure 5. Access Delay at the 95 Percentile for different values of \( M \).

REFERENCES


