The S-Aloha Capacity: Beyond the $e^{-1}$ Myth

Luca Barletta, Flaminio Borgonovo, and Ilario Filippini
Politecnico di Milano
{lucabarletta, flaminio.borgonovo, ilario.filippini}@polimi.it

Abstract—The stability and throughput of the Slotted Aloha protocol have been studied at length, yielding results that depend on the environment and channel assumptions, in many cases indicating $e^{-1}$ as the S-Aloha capacity. When users can detect only their own collisions, and the number of users $N$ goes to infinity, no definite capacity result exists. Approximated models have been introduced to study the exponential back-off mechanism, which seem to indicate an asymptotic capacity of $\ln(2)\div 2$ when binary back-off is used, and again $e^{-1}$ when the exponential base is optimized. Here we introduce a more accurate and flexible model that shows that past results miss their mark. In fact, we prove that with binary back-off the capacity is practically $0.370$, slightly greater than $e^{-1}$; furthermore, and more important, we prove that using $1.35$ as exponential back-off base, the capacity reaches $0.4303$ with an infinite number of users, and up to $0.4096$ with $N = 2$ users.

Index Terms—S-Aloha, Exponential Back-off, Collision Resolution, Capacity, Maximum Throughput, Random Access.

I. INTRODUCTION

The Aloha protocol, since its appearance in 1970 [1], has been perhaps the most studied topic in the multiple-access area, since its introduction has revolutionized the multiple-access world. Its application covers important fields such as satellite, cellular and local-area communications, and recently has been applied to radio frequency identification (RFID). However, questions regarding its stability and capacity, under many circumstances and channel assumptions, remain still unanswered, especially when dealing with an unlimited number of users and no channel feedback is available, as is the case here considered.

First studies assumed the stability of the channel traffic and its Poisson distribution, yielding for the slotted Aloha version here considered, a maximum throughput of $e^{-1}$ [2] [3], a figure that, even under other assumptions, has turned out quite often, as if a magic number for Aloha. It was soon realized that, from a practical point of view, the re-transmission probability of collided packets should be decoupled from first-time transmissions [4], but it was through a Markov model, proved to be non-recurrent [5], that instability was first mathematically proved.

In [5] it was also suggested that, in order to get a stable system and optimal throughput, the re-transmission probability should be tailored, when possible, over the number of transmitting users $n$. For example, denoting by $\beta$ the transmission probability, the throughput with $n$ users is

$$s = n\beta(1 - \beta)^{n-1},$$  \hspace{1cm} (1)

and is maximized by $\beta = 1/n$, providing throughput $s = e^{-1}$ as $n \rightarrow \infty$. Assuming that some additional information is available from the observation of the channel, an estimate $\hat{n}$ of $n$ can be attempted, and the throughput optimized by setting $\beta = 1/\hat{n}$. For example, in some broadcast channels, transmissions can be monitored and the outcome of the slot, i.e., empty, success, or collided, is made available. In [6]–[9] procedures and estimates have been suggested that are able to provide the theoretical throughput of $e^{-1}$.

In Dynamic Frame Aloha, where transmissions are organized in frames [10], the system can be stabilized by getting from the channel feedback some estimate $\hat{n}$ of the backlog $n$, in order to exploit the $\beta = 1/n$ optimality of (1). The maximum throughput attained in this way, $0.426$ pkt/slots, is greater than $e^{-1}$, owing to the fact that the number of ongoing collisions, and frame lengths, are quite often very small, taking advantage of the fact that the Aloha throughput is higher when fewer users are involved, as predicted by (1) in $\beta = 1/n$. In RFID, where Dynamic Frame Aloha has been recently adopted [11], [12], an interesting performance figure is the limiting throughput

$$\eta = \lim_{N \rightarrow \infty} \frac{N}{L(N)},$$

where $L(N)$ is the average number of slots needed to identify $N$ users. It has been recently proved in [13] that, if the backlog $n$ is known at each frame, the best strategy is to set the frame length equal to $n$, in which case the limiting throughput is, again, $e^{-1}$.

When channel feedback is available, however, a more efficient protocol family is that of Tree Algorithms [14]–[16], where collisions are solved by splitting colliding users into groups that are solved separately. The repetition of such splitting gives rise to groups organized as the leaves of a tree, hence the name. The best algorithm of this family is known as the Gallagher-Tsybakov algorithm [15], [16] that reaches throughput $0.487$, the highest ever attained by random-access algorithms, provided that the arrival frequency is known. In [17] it has been shown that, by randomly subdividing users into small groups, each of them solved by Dynamic Frame Aloha, throughput $0.469$ can be reached.

When no channel feedback is available, as is the case in standards for LANs and WLANs, the above mechanisms can...
The mechanism most often referred to is the exponential back-off (EB), which decreases the station transmitting probability according to the negative exponential law

$$\beta(i) = b^{-i-i_0},$$

where $i \geq 0$ counts the number of consecutive collisions experienced in transmitting a packet, $b > 1$, and $i_0$ is the transmission probability offset. EB has undergone many analysis attempts in order to assess its capabilities, especially in view of the fact that, with some variations, EB has been adopted in IEEE 802.3 and IEEE 802.11 standards in the binary EB (BEB) variation, i.e., with $b = 2$. Many efforts have been devoted to investigate issues such as stability and capacity. Unfortunately, the analysis of such protocol is quite complex and the results attained are partial and somehow contradictory and confusing, owing to the many differences in stability definitions and the assumptions underlying the analyzed models, as it appears from the bibliography cited below.

BEB for Slotted-Aloha under the infinite population model has been proved unstable by Aldous in [18]. Here, the authors assume that users arrive, transmit their packet according to law (2) with $i_0 = 0$, and after success leave. He proves that BEB asymptotically provides zero throughput under any positive packet arrival rate $\lambda$.

When a finite number $N$ of station is assumed, capacity evaluation is still an open issue, although some bounds have been obtained, still with $i_0 = 0$. In [19], Goodman et al. prove that an arrival frequency $\lambda^*(N) > 0$ does exist such that the system is stable if $\lambda(N) < \lambda^*(N)$, where $\lambda^*(N) \geq 1/N^{1+\alpha} N$ for some constant $\alpha$. In [20], Al-Ammal et al. improve the bound in [19] proving that BEB is stable for arrival rates smaller than $1/(\alpha N^{1-\eta})$, where $\eta < 0.25$. Finally, in [21], Håstad et al. show, using the same analytical model as in [19], that BEB is unstable whenever $\lambda_i > \lambda/N$ for $1 \leq i \leq N$, and $\lambda > 0.567 + 1/(4N - 2)$, where $\lambda$ is the system arrival rate and $\lambda_i$ is the arrival rate at node $i$.

The capacity in the $N = 2$ case under Bernoulli arrivals and $i_0 = 0$ has been proved to be 0.6096 in [22]. Here it has been shown that capture periods, one for each station, alternate on the channel. During a period the capturing station successfully transmit in subsequent slots until its queue is emptied.

Due to the complexity of an exact analysis, further attempts have introduced simplified models and approximations. Among these models, the saturation model has been first introduced in [23]. This model tries to analyze stability and capacity issues by assuming that queues are always full, such that once a successful transmission has occurred at a station, immediately a new one is available for transmission. This model is somewhat simpler and pessimistic with respect to the one with real queues, and has been adopted in the hope that it presented a stable behavior and positive capacity, thus guaranteeing the stable behavior and the capacity of the more realistic one.

With the saturation model, an approximate analysis is made possible by a further strong assumption, first introduced in [23], and then largely used, known as the decoupling assumption. This assumption has a twofold implication, i.e., the stationary behavior of the model, and the independence in the behavior of the different transmitters. These assumptions lead to a mean value analysis (MVA) and a fixed point equation [23], that provides in a simple way the basic performance figures of the protocol.

Though the above assumptions can be reasonable in some cases, for example when the back-off index is limited, as in LAN standards, or when the offset $i_0$ is large, they are highly questionable, or actually wrong, when unlimited backoff index and small $i_0$ are concerned, which are the cases relevant to Aloha. In [24] the authors provide the asymptotic capacity of the Aloha EB with unlimited back-off index as $\ln(b)/b$, independent of $i_0$, which yields $\ln(2)/2 = 0.34$ when $b = 2$, and $e^{-1}$ for the optimized value of $b$. As a matter of fact, up to now such capacity results have never been checked against a more realistic model or simulation.

In more recent years, many interesting works dealing with the decoupling assumption have appeared [25], [26]. They leverage on the fact that the decoupling property is observed in a scaled model of the system that for $N \to \infty$ converges to a deterministic system described by ordinary differential equations that define the mean field limit [27]. Although this model is quite elegant, flexible and can provide a lot of insights in the many possible variations, in the variation we are interested in, i.e. an infinite number of back-off stages and a small offset $i_0$, it provides the same approximated model as in [24], which fails to be accurate enough for small $i_0$.

In this paper we consider the saturation model with the back-off mechanism (2), focusing on the S-Aloha protocol, which means that we look for the smallest value of $i_0$ allowed by stability, i.e., $i_0 = 2$, to minimize delays and maximize the available bandwidth with few users. We first derive some results about the Markov chain that describes the content of each back-off stage, a chain too complex to allow direct throughput results. We then prove that the results attained by the decoupling assumption can be exactly re-derived assuming that the distribution of users among the different stages is jointly Poisson (Poisson assumption). This presents the advantage to provide all the results of the decoupling assumption, e.g., those derived in [24], with very simple passages. As main result we introduce a model that assumes a joint Poisson distribution of users only for high index stages, say from stage $s$ onward. In this way we show that the joint distribution of the lower index stages is numerically computable, since the number of users involved is quite low. Also, the performance
is practically provided by users in the lower index stages, so that the model is quite accurate, its accuracy increasing with \( s \). For \( N \to \infty \) the Poisson assumption turns out to be pessimistic and therefore the model provides a lower bound to real performance. We find that with \( s = 5 \) the approximation is practically close to the real values, and provides for BEB a capacity close to 0.3706, slightly greater than \( e^{-1} \), and 0.4303 when the exponential base is optimized. Finally, we numerically show that stationary behavior is not possible with \( i_0 \leq 1 \), and give insights into why the decoupling assumption provides a lower bound to the performance.

The paper is organized as follows. In Section II we derive some basic relations and results about the model. In Section III we discuss the decoupling assumption and present the Poisson model that is able to represent it. In Section IV we present the semi-Poisson model (SPM) and the numerical procedure that provides the throughput. Stability issues are also discussed. Finally, in Section V numerical results are presented.

II. PRELIMINARY RESULTS

The system is described by the chain \((N_0, N_1, \ldots, N_i, \ldots)\), where \( N_i \) denotes the number of users with index \( i \). In the following we assume a stationary behavior, unless otherwise proved. The throughput of stage \( i \) can be expressed as

\[
\begin{align*}
s_i = \mathbb{E} \left[ \gamma_i(1, N_i) \prod_{k=0, k \neq i}^{\infty} \gamma_k(0, N_k) \right],
\end{align*}
\]

where the average is taken over the joint distribution of the \( N_i \)'s, and

\[
\gamma_i(k, N_i) = \binom{N_i}{k} (b^{-i-i_0})^k (1 - b^{-i-i_0})^{N_i-k}
\]

the probability of having \( k \) transmissions in stage \( i \). The throughput is evaluated as

\[
\lambda_0 = \sum_{i=0}^{\infty} s_i.
\]

The rate of the flow of users across stage \( i \) is \( \lambda_i = \lambda_{i-1} - s_{i-1}, \quad i \geq 1 \), and the “routing” probability out of stage \( i \), defined as \( \alpha_i = \frac{\lambda_{i+1}}{\lambda_i} \), is given by:

\[
1 - \alpha_i = \frac{s_i}{\lambda_i} = \frac{s_i}{\mathbb{E}[N_i] b^{-i-i_0}},
\]

where the last passage comes from Little's result [28].

We also have

\[
\lambda_i = \lambda_0 \alpha_0 \alpha_1 \ldots \alpha_{i-1}, \quad i \geq 1,
\]

and again using Little's result, we can express \( N \) as

\[
N = \sum_{i=0}^{\infty} \lambda_i T_i = \sum_{i=0}^{\infty} \lambda_i b^{i+i_0},
\]

where \( T_i = b^{i+i_0} \) represents the average time the user stays in stage \( i \). The convergence of (3) requires, by the ratio test,

\[
\lim_{i \to \infty} \frac{\lambda_{i+1}}{\lambda_i} = \lim_{i \to \infty} \alpha_i < 1/b.
\]

Other interesting properties, such as the increasing behavior of sequence \( \{\alpha_i\} \), though intuitive, can only be proved by looking at the properties of the joint distribution of \((N_0, N_1, \ldots, N_i, \ldots)\). Unfortunately, the chain has an infinite number of dimensions, each of them with an infinite number of states, which prevents not only an analytical investigation, but also a numerical solution of the chain. We see in Sec. IV how some of these difficulties can be overcome. A worth mentioning result we are able to derive is

\[
P(\text{idle}) > \frac{b - 1}{b}, \quad (4)
\]

where \( P(\text{idle}) \) represents the probability that the channel is idle. From this we can immediately argue that BEB can not completely exploit the channel, and that a lower value of \( b \) is needed to ease this constraint.

III. DECOUPLING-ASSUMPTION DISCUSSION

The decoupling assumption makes \( \alpha_i = \alpha \) for all \( i \). With this assumption, the MVA in [23] and [24] is able to derive the performance figures of interest, such as throughput and rates \( \lambda_i \). These papers refer to the window-type back-off mechanism, as assumed in standards, although exactly the same results are attained with geometric back-off, as adopted in [25], [26] and the present paper.

We now prove the following:

**Theorem 1** The decoupling assumption provides exactly the same results as assuming that the stationary distribution of \((N_0, N_1, \ldots, N_i, \ldots)\) is a joint Poisson distribution, independent from stage to stage. Denoting by \( \Lambda = \sum_{k=0}^{\infty} \lambda_k \) the average traffic on the channel, we further have

\[
\alpha_i = 1 - e^{-\Lambda}, \quad \lambda_i = \lambda_0 (1 - e^{-\Lambda})^i, \quad (5)
\]

\[
\lambda_0 = \Lambda e^{-\Lambda}, \quad (6)
\]

\[
N = b^{i_0} \frac{\Lambda e^{-\Lambda}}{1 - b(1 - e^{-\Lambda})}, \quad (7)
\]

\[
\Lambda < \ln \frac{b}{b - 1}, \quad P(\text{idle}) > \frac{b - 1}{b}, \quad (8)
\]

\[
\lambda_0 < \frac{b - 1}{b} \ln \frac{b}{b - 1} \quad (9)
\]

**Proof:** Because of the assumed distribution, the \( N_i \)'s are statistically independent. Since a user with index \( i \) transmits independently of others with probability \( b^{-i-i_0} \), the distribution of transmitting users at stage \( i \) is still Poisson, independent of other \( N_j, j \neq i \), with average \( \lambda_i = \mathbb{E}[N_i] b^{-i-i_0} \). Since the sum of independent Poisson variables is still Poisson distributed, results (5) and (6) immediately follow. By substituting into (3) and using (6) we prove (7) as

\[
N = \lambda_0 \sum_{i=0}^{\infty} (1 - e^{-\Lambda})^i b^{i+i_0} = \lambda_0 b^{i_0} \frac{1}{1 - b(1 - e^{-\Lambda})}. \quad (9)
\]

In order for (9) to hold, inequality (8) must be verified; the second version comes from \( P(\text{idle}) = e^{-\Lambda} \). Finally, the last
point of the Theorem is proved by substituting the limit value
\( \Lambda = \Lambda^* = \ln \frac{b}{b-1} \) in (6).

Equation (7) provides the channel traffic \( \Lambda \), while (6) gives the throughput. The results we get with the Poisson assumption are exactly those attained in the cited papers, in particular the value of capacity, reached when \( N \to \infty \). If those results can be achieved by MVA, then the joint distribution of \( (N_0, N_1, \ldots, N_i, \ldots) \) only intervenes in determining the decoupling property. In other words, any distribution that results in the decoupling property provides exactly the same results. The Poisson distribution does so, and furthermore provides a simpler and more detailed model, where deriving results, such as the “canonical” curve \( \lambda_0(\Lambda) \) that provides the throughput as a function of channel traffic, is straightforward.

The throughput curve (6) as function of \( \Lambda \) reaches its maximum in \( \Lambda = 1 \), but the model is stable up to \( \Lambda^* = \ln \frac{b}{b-1} \), approximately equal to 0.7 for \( b = 2 \). This reflects the well-known fact that the BEB is over-reacting: As \( \Lambda \) increases some users are pushed toward high back-off indexes so that, practically, they can not contribute to \( \Lambda \), although an increase of \( \Lambda \) would mean an increase in throughput.

Note that if we want to maximize the throughput of this model in \( \ln \frac{b}{b-1} \) we need
\[
\ln \frac{b}{b-1} = 1, \\
\]
which proves the following:

**Corollary 1** The maximum throughput of the Poisson model, equal to \( e^{-1} \), is provided by assuming a back-off basis equal to
\[
b = \ln \frac{1}{1 - e^{-1}} \approx 1.52,
\]
Again, the above result has already been derived in [24] with MVA.

### IV. Semi-Poisson Model

We have seen that the system is modeled by a Markov chain \( (N_0, N_1, \ldots, N_i, \ldots) \), which has an infinite number of dimensions, each of them with an infinite number of states; this prevents not only an analytical investigation, but also a numerical solution of the chain. On the other hand, the decoupling assumption makes the system analyzable, but introduces an approximation that only provides a lower bound to capacity, as shown in Sec. IV-D. The fact that the average number of users decreases at higher index stages suggests adopting the decoupling/Poisson assumption only for such stages, say from stage \( s \) onward, introducing the SPM, whose accuracy in describing the actual system increases as \( s \) increases.

The SPM assumes that starting from index \( s > 0 \), the stationary distribution of \( (N_s, N_{s+1}, \ldots) \) is jointly Poisson distributed with averages \( (n_s, n_{s+1}, \ldots) \). Therefore, the distribution of the number of transmitting users with index \( i \geq s \) is given by the Poisson distribution, as in Th. 1, with average
\[
\Lambda_s = \sum_{i=s}^{\infty} \lambda_i = \sum_{i=s}^{\infty} n_i b^{-i-i_0}. \tag{10}
\]

This shows that stages \( s, s+1, \ldots \), from the transmission point of view, can be lumped together into a single enlarged stage \( s \), as shown in Fig. 1. Assuming \( \Lambda_s \) as a known constant, we can determine the behavior of the Markov process \( (N_0, N_1, \ldots, N_{s-1}) \), whose stationary distribution and cumulative average,
\[
n^i(<s) = \sum_{i=0}^{s-1} \mathbb{E} [N_i],
\]
develop depend on \( \Lambda_s \). To complete the model, we need a relationship between \( \Lambda_s \) and \( n^i(\geq s) = \sum_{i=s}^{\infty} n_i \) such that we can impose the constraint \( n^i(<s) + n^i(\geq s) = N \). Let denote by \( \tau_s \), the stationary probability that no user of index \( i \leq s - 1 \) transmits, i.e.,
\[
\tau_s = \mathbb{E} \left[ \prod_{i=0}^{s-1} \gamma_i(0, N_i) \right],
\]
where the average is taken over the stationary joint distribution of \( (N_0, N_1, \ldots, N_{s-1}) \). Owing to the Poisson assumption for stages \( i \geq s \), and (10), we have
\[
s_i = \lambda_i - \lambda_{i+1}
\]
\[
= \tau_s \lambda_i b^{-i-i_0} e^{-\sum_{i=s_i}^{\infty} n_j b^{-j-i_0}}
\]
\[
= \tau_s \lambda_s e^{-\Lambda_s}, \quad i \geq s.
\]

Recognizing that \( \tau_s e^{-\Lambda_s} = P(\text{idle}) \) is the probability that the channel is idle, we finally have
\[
1 - \alpha_i = \frac{s_i}{\lambda_i} = P(\text{idle}), \quad i \geq s,
\]
independent of \( i \) as expected.

Using the passages as in Th. 1, we get
\[
\lambda_i = \lambda_s \alpha^{i-s} = \lambda_s (1 - P(\text{idle}))^{i-s}, \quad i \geq s,
\]
\[
n_i = \lambda_s (1 - P(\text{idle}))^{i-s} b^{i+s_0}, \quad i \geq s,
\]
\[
n^i(\geq s) = \sum_{i=s}^{\infty} n_i = \lambda_s \sum_{i=s}^{\infty} (1 - P(\text{idle}))^{i-s} b^{i+s_0}
\]
\[
= \lambda_s b^{s+s_0}
\]
\[
\geq s.
\]

where the summation exists for \( b(1 - P(\text{idle})) < 1 \).

Referring to Fig. 1, we have
\[
\Lambda_s = \lambda'_{s} = \tau_s \Lambda_s e^{-\Lambda_s} = \Lambda_s P(\text{idle}),
\]
and using this into (11) we finally have
\[
n^i(\geq s) = \frac{\Lambda_s P(\text{idle}) b^{i+s_0}}{b P(\text{idle}) - (b - 1)}. \tag{12}
\]
The existence of (12) only for $P(\text{idle}) > \frac{b-1}{b}$ proves (4), since by increasing $s$ the real system is approximated as closely as wanted by the SPM.

Using $n(\leq s) = N - n(\geq s)$, (12) can be written as

$$
\Lambda_s = (N - n(\leq s))b^{-s-i_0} \left( b - \frac{b-1}{P(\text{idle})} \right).
$$

Either (12) and (13) represent the required relationship that makes the system determined. Assuming $\Lambda_s$ as a known constant, then $(N_0, N_1, \ldots, N_{s-1})$ is a Markov process whose stationary distribution provides $\Lambda_s$, $\tau_s$, $n(\leq s)$, and $P(\text{idle})$, all of which depend on $\Lambda_s$. Therefore, the correct $\Lambda_s$ for the model is the one that makes equal both sides of (13). An iterative procedure can be envisaged to this purpose. The convergence of the procedure is discussed in Sec. IV-C. Here we anticipate that the solution does not exists if $i_0 < 1$, showing that in these cases the chain can not reach a stationary behavior. This is, in fact, the reason we derive, in Sec. V, numerical results for $i_0 \geq 2$.

Once convergence is assured the throughput can be evaluated as

$$
\lambda_0 = e^{-\Lambda_s} \sum_{j=0}^{s-1} E [\gamma_j(1, N_j)] \prod_{k=0, k \neq j}^{s-1} \gamma_k(0, N_k) + \Lambda_s e^{-\Lambda_s} E \left[ \prod_{k=0}^{s-1} \gamma_k(0, N_k) \right],
$$

and for $i \leq s - 1$

$$
1 - \alpha_i = \frac{e^{-\Lambda_s} E \left[ \gamma_i(1, N_i) \prod_{k=0, k \neq i}^{s-1} \gamma_k(0, N_k) \right]}{E [N_i] b^{-i-i_0}}.
$$

### A. Model Complexity

In order to make the model numerically solvable we must limit the range of the variates $N_0, N_1, \ldots, N_{s-1}$, i.e., the size of the stages, say to a common value $N_{\text{max}}$. This is a reasonable approach, as the averages $E [N_i]$ are finite. The transitions that lead to states with $N_i > N_{\text{max}}$ are not allowed, i.e., deleted from the original chain. This introduces a further approximation in the model that, however, can be made negligible by increasing $N_{\text{max}}$ and thus the computational cost. The accuracy of the approximation can be estimated by the low value $P(N_i = N_{\text{max}})$ that the model provides.

From the practical point of view, complexity is dictated by the number of states of the Markov chain, which is given by $(N_{\text{max}})^s$. The limits of the model are discussed in Sec. V. However, the fact that for $b = 2$ the sequence $\{E [N_i]\}$, starts with $E [N_0] \approx 1.5$ and decreases with $i$, suggests that limited values of $N_{\text{max}}$, such as 5 or 6, are enough to make the approximation error negligible.

### B. Transition Probabilities

The non-zero transition probabilities (TPs) with initial state $(N_0, N_1, \ldots, N_{s-1})$ can be partitioned into groups corresponding to one of the following 4 events:

1. Success generated by a stage with $i \geq s$. The final state is $(N_0 + 1, N_1, \ldots, N_{s-1})$, and the TP is

$$
\Lambda_s e^{-\Lambda_s} \sum_{j=0}^{s-1} \gamma_j(0, N_j).
$$

2. Success generated by a stage with $1 \leq i < s$. The final state is $(N_0 + 1, \ldots, N_i - 1, \ldots)$, and the TP is

$$
e^{-\Lambda_s \gamma_i(1, N_i)} \sum_{j=0, j \neq i}^{s-1} \gamma_j(0, N_j).
$$

3. Collision, where $C_i$ is the number of transmissions generated by stage $i$. The final state is $(N_0 - C_i, N_1 + C_0 - C_1, N_2 + C_1 - C_2, \ldots, N_{s-1} + C_{s-2} - C_{s-1})$, and the TP is

$$
\left\{ \begin{array}{ll}
\sum_{j=0}^{s-1} \gamma_j(C_j, N_j), & \text{if } \sum_{j=0}^{s-1} C_j \geq 2, \\
(1 - e^{-\Lambda_s}) \sum_{j=0}^{s-1} \gamma_j(C_j, N_j), & \text{if } \sum_{j=0}^{s-1} C_j = 1.
\end{array} \right.
$$

4. The state does not change. This happens when there is a success generated by stage $i = 0$, or when no one in the stages $i < s$ transmits and there is no success from the stages $i \geq s$. The TP is

$$
\left( 1 - \Lambda_s e^{-\Lambda_s} + e^{-\Lambda_s} \frac{\gamma_0(1, N_0)}{\gamma_0(0, N_0)} \right) \prod_{j=0}^{s-1} \gamma_j(0, N_j).
$$

When at least one of the components of the final state exceeds $N_{\text{max}}$, the TP is set to zero. In this way, the TP matrix is square of order $(N_{\text{max}})^s$.

### C. SPM Solution and Convergence

The solution of the iterative procedure in Sec. IV comes at the intersection of $\Lambda_s$ and the right-hand side $f(\Lambda_s)$ of (13), which is a decreasing function of $\Lambda_s$. This decrease occurs until $P(\text{idle}) = (b-1)/b$, after which no operating point can exist. In some cases, however, we can not have operating points, i.e., $P(\text{idle}) > (b-1)/b$, even with the smallest $\Lambda_s$. Therefore, denoting $\lim_{\Lambda_s \to 0} P(\text{idle}) = \varepsilon(s, i_0)$, the solution exists if $\varepsilon(s, i_0) > (b-1)/b$. In our numerical evaluations this condition always occurred when $i_0 \geq 2$, suggesting that the solution always exists. When $i_0 = 0$, we can analytically solve the SPM for $s = 1$, since it reduces to a two-state $0/1$ chain whose transition probabilities can be derived from those given in Sec. IV-B. In this case it is easy to prove that $\varepsilon(1, 0) = 1/2$, meaning that a solution exists when $b < 2$; otherwise, no solution exists. Also, we have numerically verified that as $s$ increases, $\varepsilon(s, 0)$ decreases approaching zero. This suggests that no solution exists in the limit $s \to \infty$, meaning that the real system behavior can not be stationary.

Similar arguments hold for the case $i_0 = 1$; here, our evaluations indicate that $\varepsilon(s, 1)$ decreases approaching $(b-1)/b$. Therefore, we can not argue about stability in this case.
D. SPM Lower Bound

The numerical results given in Sec. V provide throughputs that increase as the parameter \( s \) of the SPM increases from zero. This seems to indicate that SPM provides a lower bound to the throughput of the real system. The reason for this is buried into the complex shape of the joint distribution of the numbers of users in the stages. In this section we provide some insights that explain how SPM with \( s = 1 \) provides greater throughput than the Poisson model.

In Appendix A we prove that, with \( s = 1 \), the tails of the distribution \( \{\pi_k\} \) of the first-stage content of SPM can be expressed as

\[
\pi_k = \pi_{k-1} \Lambda_1 e^{-\Lambda_1} (1-b^{-i_0})^k. \tag{14}
\]

On the other side, the Poisson distribution for the Poisson model is such that

\[
\pi_k' = \pi_{k-1}' E[N_0] / k.
\]

Thus, we see that the tail of the distribution \( \{\pi_k\} \) decreases much faster than the Poisson’s. Also, being \( \{\pi_k\} \) unimodal, if the average \( E[N_0] \) is the same for the two models, we have \( \pi_0 < \pi_0' \), meaning that \( \{\pi_k\} \) is more grouped around its average.

In Figure 2 we report the conditional throughput of BEB

\[
s(k) = \Lambda_1 e^{-\Lambda_1} (1-2^{-i_0})^k + k2^{-i_0}(1-2^{-i_0})^{k-1} e^{-\Lambda_1}, \tag{15}
\]

when \( i_0 = 2 \) and \( \Lambda_1 = 0.3291 \), that correspond to the maximum throughput. Due to the differences noted above, the average of the throughput with respect to \( \{\pi_k\} \), whose mean is about 1.5, presents higher throughput than with \( \{\pi_k'\} \). Fig. 2 also reports the curve (15) for \( i_0 = 4 \) and \( \Lambda_1 = 0.3417 \), again corresponding to maximum throughput, being the average of the distribution about 5.6. Since in this case the throughput curve becomes more levelled, we expect a smaller difference with the Poisson case, as confirmed by results in Sec. V.

In comparing with the Poisson case, however, we must consider that an higher throughput cause an increase of the average number of users in stage 0 and a reduction in the other stages; this, in the iteration procedure, requires an adjustment of rate \( \Lambda_1 \) and a new evaluation of the throughput. Nevertheless, if we consider the limit \( N \to \infty \), the increase in the average number of users in stage 0 does neither alter the number in other stages nor \( \Lambda_1 \). Then we may conclude that the Poisson model provides a maximum throughput that lower bounds the maximum throughput of the Semi-Poisson model with \( s = 1 \).

Although we conjecture that similar mechanism causes the maximum throughput of SPM to increase with \( s \), an analytical proof is difficult due to the complexity of the Markov chain, as the number of its states increases exponentially with \( s \).

V. Model Results

The model allows to derive the throughput as function of \( N \). However, it is perhaps more interesting to draw, as usual, the throughput against the channel traffic \( \Lambda \). We have seen that the channel traffic increases with \( N \) and reaches a maximum, \( \Lambda^* \), as \( N \to \infty \), where \( P(\text{idle}) \) reaches the limiting value \( (b-1)/b \). It is also interesting to plot the throughput for values beyond \( \Lambda^* \), which can highlight whether the system is operating in the constrained (increasing) or in the congested (decreasing) part of the curve. This is possible if we consider the lumped state as decoupled from lower index stages. In fact, the lumped state generates traffic \( \Lambda_s \), that, when successful, feeds users into stage 0. Thus, \( \Lambda_s \) is seen as a forcing parameter whose increase makes \( \Lambda \) increase, which can be any value, even in the range \( \Lambda > \Lambda^* \). However, it must be noted that increasing \( \Lambda \) over \( \Lambda^* \) in the end increases congestion and the average content of lower index stages. This in turn makes the truncation \( N_i \leq N_{\max} \), \( i \leq s-1 \), operate more often, thus decreasing the accuracy of the numerical solution.

Since our model is only numerically solvable, the capacity \( \Lambda^* \) can be approximated by looking for the value \( \Lambda_s \), the model’s input variable, that makes \( P(\text{idle}) \) approaching from above the limiting value \( (b-1)/b \). In this critical range the model is highly sensitive, since a very small increase in \( \Lambda_s \) causes great changes in \( N \), but negligible changes in the other variables, so that \( \Lambda^* \) can be approximated very finely.

A. Binary Exponential Back-off

Here we refer to the BEB law with \( i_0 = 2 \). The results have been attained with a number of stages up to \( s = 5 \), limiting the content of each stage to \( N_{\max} = 10 \) for \( s < 5 \), and \( N_{\max} = 6 \) for \( s = 5 \).

In Fig. 3 we report the throughput curves, as function of channel traffic \( \Lambda \), for different values of \( s \), namely \( s = 0, 1, 2, 3, 4, 5 \), corresponding to different accuracy degrees of the model. We note that the curve for \( s = 0 \) represents the throughput attained with the decoupling assumption, which correspond to (6), the well known curve of S-Aloha under the stationary infinite population model. Also reported on the curves are the circles corresponding to the limiting values \( \Lambda^* \), whose coordinates are reported for a precise comparison in Table I. We see that for the same \( \Lambda \) the throughput increases as \( s \) increases showing that, using the joint distribution provided by the model, the throughput converges to a different value with respect to the Poisson assumption. However, we must note that as \( s \) increases, the average content of stages increases as well, so that, if we refer to the same \( N \) in the different
curves, we must slightly decrease $\Lambda$. This is also the reason why $\Lambda^*$ decreases as $s$ increases.

Even with $s = 1$ we observe a noteworthy difference with respect of the Poisson model. The improvement increases with $s = 2$, probably due to the effect of the joint distribution of $(N_0, N_1)$. Adding another stage has a more limited effect, and when $s = 5$ the throughput improvement is very limited, due to the fact that transmissions at the fifth stage are very few.

A further increase in $s$ is expected to be barely noticeable, meaning that the convergence of the SPM, hence capacity, is reached. Remarkable is the estimate of the maximum throughput $0.3706$ reached with $s = 5$. Actually, this value has been reached with the forcing parameter $\Lambda_5 = 0.01379$, and provides $P(\text{idle}) = 0.50003$ and $\Lambda = 0.65016$, which correspond to $N = 62154$. The attained bound is well above the value $\ln(2)/2$, in $s = 0$, predicted by the decoupling assumption.

Figure 4 shows the performance of the model with $s = 5$ at capacity. In particular we report $\alpha_i$ as function of $i$, showing that indeed this sequence is an increasing sequence, compared to the constant $\alpha_1 = 0.5$ predicted by the Poisson model. We can appreciate that the difference in $\alpha_0$ is remarkable. We also report the curve $\lambda_i$, which corresponds to the flow in stage $i$. Here we see the dramatic decrease as $i$ increases, and the value $0.0034$ reached for $i = 4$. This low value confirms the limited impact of the performance of this stage and shows that the greater accuracy attained in adding a further stage is negligible.

Table II shows the marginal distributions of $N_i$, $i = 0, \ldots, 4$ for $s = 5$, $\Lambda = \Lambda^*$, and $N_{\text{max}} = 6$. Again, the very low values reached in $N_i = 6$ show that increasing $N_{\text{max}}$ above 6 does not appreciably increase the precision of the model. Then, we can conclude that the capacity of the BEB is very closely approximated by $0.3706$.

Figure 5 reports the throughput curves, as in Fig. 3, for the case $i_0 = 4$. The increase of the offset $i_0$ increases the sojourn time in stages and the average content $E[N_i]$, which now is about 5.6 in the first stage. The increase in $E[N_i]$ makes the Poisson assumption more accurate, and this explains why the curves of the model are closer to the Poisson curve than those of Fig. 3 for $i_0 = 2$. Also, the increase in $E[N_i]$ requires $N_{\text{max}} = 10$, thus increasing the complexity of the model. However, this is compensated by the reduction in the number of stages, which is now $s = 3$. Here capacity is about 0.351 at $\Lambda^* = 0.6845$.

B. Optimal Exponential Back-off

The Poisson model indicates $b \approx 1.52$ as the optimal back-off base with capacity $e^{-1}$. Here, given a base $b$, the capacity decreases as $i_0$ increases, so that our investigations can be limited to $i_0 = 2$. Figure 6 shows the throughput curves of the SPM model for the case $i_0 = 2$, and different values of $b$. Again, circles representing the maximum achievable
Figure 5. BEB, \( i_0 = 4 \): Throughput of the Poisson model (bottom line) and the SPM with \( s = 1, 2, 3 \), versus traffic channel \( \Lambda \). The dot-dash curve represents the throughput of BEB, \( i_0 = 2 \), with \( s = 5 \); circles represent the maximum achievable throughput.

Figure 6. \( i_0 = 2 \): Throughput of the SPM versus traffic channel \( \Lambda \), with \( s = 5 \) for \( b = 2 \), and \( s = 6 \) for \( b = 1.5, 1.45, 1.4, 1.35, 1.3 \). Circles represent the maximum achievable throughput. The throughput under Poisson assumption is reported in dashed line. Asterisks correspond to the case \( N = 2 \).

throughput are reported. Also reported on the curves are asterisks representing the throughput attained by simulating the real system for \( N = 2 \). As we can see, the simulated results lie almost exactly on the curves. Decreasing \( b \) allows for more traffic on the channel, although the average content \( E[N_i] \) diminishes due to the shorter time spent at each stage. This provides more efficiency and the overall effect is that the throughput curves become higher and shifted to the left, so that the maximum channel traffic becomes placed on the downward region of the throughput curve.

Table III reports the SPM maximum throughputs corresponding to the circles in Fig. 6. We see that the capacity is 0.4303, which is attained with \( b = 1.35 \). For smaller values of \( N \), however, the throughput is higher with smaller values of \( b \), and at \( N = 2 \) reaches 0.496 with \( b = 1.35 \) and 0.5295 with \( b = 1.15 \). The reduction of \( E[N_i] \) from \( b = 1.5 \) and lower, allows additional precision, and a further stage is added.

Table III

<table>
<thead>
<tr>
<th>( b )</th>
<th>( \Lambda^* )</th>
<th>( \Lambda^* )</th>
<th>( b = 1.4 )</th>
<th>( b = 1.4 )</th>
<th>( b = 1.35 )</th>
<th>( b = 1.35 )</th>
<th>( b = 1.3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \lambda )</td>
<td>( 0.6502 )</td>
<td>( 0.9612 )</td>
<td>( 1.0109 )</td>
<td>( 1.0669 )</td>
<td>( 1.1309 )</td>
<td>( 1.2058 )</td>
<td>( 1.1858 )</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>( 0.0138 )</td>
<td>( 0.0511 )</td>
<td>( 0.0652 )</td>
<td>( 0.0847 )</td>
<td>( 0.1121 )</td>
<td>( 0.1519 )</td>
<td>( 0.1819 )</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>( 0.3706 )</td>
<td>( 0.4247 )</td>
<td>( 0.4279 )</td>
<td>( 0.4300 )</td>
<td>( 0.4303 )</td>
<td>( 0.4279 )</td>
<td>( 0.4247 )</td>
</tr>
</tbody>
</table>

Figure 7. \( i_0 = 2 \): Throughput of the SPM with \( b = 1.35 \), versus traffic channel \( \Lambda \) and for \( s = 0, 1, 2, 3, 4, 5, 6 \). Circles represent the maximum throughput. Asterisks represent simulation results for \( N = 2, \ldots, 6 \).

(\( s = 6 \)) with a content truncated to \( N_{max} = 4 \). Figure 7 shows the convergence of throughput curves in the optimal case for \( N \to \infty \) where \( b = 1.35 \).

To end this section, we must observe that the reported numerical values and capacity are the highest ever observed in S-Aloha, and even overcome the maximum throughput observed with channel feedback. Also, the fact that the throughput decreases as \( N \) increases vaguely resembles the case of \( N \) users that transmit with probability \( \beta = 1/N \), since throughput (1) equals 0.5 at \( N = 2 \) and decreases, as \( N \) and contentions increase, to reach \( e^{-1} \) as \( N \to \infty \).

VI. CONCLUSIONS

In this paper we have introduced the semi-Poisson model for the Aloha protocol with exponential back-off mechanism. This model provides a lower bound to the maximum throughput of the real system, and can be adjusted to closely approximate the protocol’s maximum throughput. In particular, we have shown that offset \( i_0 = 2 \) and base \( b = 1.35 \) provide a maximum throughput equal to 0.4303, a value that practically represents the capacity, since the model is very accurate. This result is the highest ever achieved with S-Aloha, even considering the versions based on channel feedback. We also argue that \( i_0 \leq 1 \) can not provide stable operation. Future work is ongoing to formally prove the latter point.

APPENDIX A

The Markov chain that represents the content of the first stage in the SPM with \( s = 1 \) presents the state diagram shown...
in Fig. 8, where the transition probabilities are

\[ p_{k,k+1} = \Lambda_0 e^{-\Lambda_0} (1 - p)^k, \]

\[ p_{k,k-h} = -\frac{k!}{h!(k-h)!} p^h (1-p)^{k-h}, \quad h = 2, 3, \ldots, k \]

\[ p_{k,k-1} = kp(1-p)^{k-1}(1-e^{-\Lambda_0}), \]

and \( p = b^{-20} \).

Letting \( \{\pi_k\} \) denote the distribution of the chain, the downward probability flow out of node \( k \) can be written as

\[ \Phi_1 = \pi_k \sum_{h=1}^{k} p_{k,k-h} = \pi_k \left( 1 - (1-p)^k - kp(1-p)^{k-1} e^{-\Lambda_0} \right), \]

and asymptotically, as \( k \to \infty \), we have \( \Phi_1 = \pi_k + o(k) \). The downward probability flow into node \( k \) can be written as

\[ \Phi_2 = \sum_{h=1}^{\infty} \pi_{k+h} p_{k+h,k} = \sum_{h=1}^{k} \pi_{k+h} \frac{(k+h)!}{h!k!} p^h (1-p)^k 
\]

\[ - \pi_{k-1} (k+1)p(1-p)^{k-1} e^{-\Lambda_0} \]

\[ \leq \pi_k (1-p)^k \sum_{h=1}^{k} \frac{(k+h)!}{h!k!} p^h = \pi_k (1-p)^k (1 - (1-p)^k), \]

where we have assumed \( \pi_{k+h} \leq \pi_k \), as in asymptotic conditions; the last equality comes from the negative binomial expansion

\[ (1-p)^{k+1} = \sum_{h=0}^{\infty} \frac{(k+h)!}{h!k!} p^h = 1 + \sum_{h=1}^{k} \frac{(k+h)!}{h!k!} p^h. \]

We see that, asymptotically, \( \Phi_2 \) becomes negligible with respect to \( \Phi_1 \), yielding \( \Phi_1 - \Phi_2 \approx \pi_k \). This difference must be balanced by the difference in the upward flows yielding:

\[ \pi_k \approx \pi_{k-1} \Lambda_0 e^{-\Lambda_0} (1-p)^{k-1} - \pi_k \Lambda_0 e^{-\Lambda_0} (1-p)^k, \]

which, asymptotically, provides (14).

REFERENCES


