The Throughput and Access Delay of Slotted-Aloha With Exponential Backoff

Luca Barletta, *Member, IEEE*, Flaminio Borgonovo, *Member, IEEE*, and Ilario Filippini, *Senior Member, IEEE*

(*Invited Paper*)

***Abstract*— The behavior of exponential backoff (EB) has challenged researchers ever since its introduction, but only approximate and partial results have been produced up to this date. This paper presents accurate results about the effect of protocol parameters on throughput and delay, assuming queues in saturation. Among the manifold results, we first introduce a simple model that provides close-form results for the approximated model known as "decoupling assumption." Since the latter fails to provide well approximated results in many cases, we also introduce a Markovian model able to trade the precision of the results with complexity even with an infinite number of users, enabling us to get definite throughput results, such as 0.3706 with binary EB, and 0.4303 with an optimized base. Analytical considerations allow to derive the tail of the access-delay distribution, found to be slowly decreasing and with no variance as the number of users goes to infinity. Taking into account the overall performance, preliminary results seem to indicate that the exponential base $b = 1.35$ is more appealing than the standard value $b = 2$.**

***Index Terms*— Slotted-aloha, backoff, throughput, delay, unfairness, decoupling assumption, random access.**

I. Introduction

THE Slotted Aloha (S-Aloha) protocol, since its appearance in 1970 [3], [4], has been perhaps the most studied topic in the multiple-access area, as its introduction has revolutionized the multiple-access world. Its applications cover important fields such as satellite, cellular and local-area communications, recently being applied to radio frequency identification (RFID).

In S-Aloha N stations share slots in a time-slotted channel, and transmit at each slot, with some probability, either new or collided packets. If the transmission is successful a new packet can be accepted for transmission, otherwise, if it is collided, the packet is kept in the transmission buffer until success. This protocol is very attractive because of its simplicity, since it does not require coordination among stations, and its flexibility, since, in principle, it can accommodate any number of stations, up to infinity. However, issues regarding its stability and maximum throughput, under many settings and channel assumptions, remain largely unanswered, especially when dealing with an unlimited number of users and no channel feedback is available, as is the case here considered.

As it will appear in the core of the paper, the analysis of processes such as the traffic channel, i.e., the number of transmission in a slot, and the throughput, i.e., the number of successes in a slot, is rather complex. Therefore, first studies assumed that, for a large number of users, the channel traffic has an independent Poisson distribution at each slot with constant average Λ, in which case the throughput is a stationary process with average λ_0 which is easily shown to relate to Λ by [4], [5]

$$\lambda_0 = \Lambda e^{-\Lambda}, \tag{1}$$

providing a maximum throughput of e^{-1} in $\Lambda = 1$, a figure that has turned up quite often, even under different assumptions.

It was soon realized that the channel traffic can be unstable, reaching a large number of transmissions, and collisions, at each slot, yielding asymptotically zero throughput. Also, it was understood that in order to have stable operation, the retransmission probability of collided packets should be decoupled from first-time transmissions [6]. In [7] it was suggested that, in order to get a stable system and optimal throughput, the retransmission probability should be tailored over the number of transmitting users N. If the channel can be observed, an estimate of N can be drawn by keeping track of the slot outcomes, i.e., *empty*, *success*, or *collided*. Works in [8]–[11] suggest estimation and access procedures that are able to provide, again, the theoretical throughput of e^{-1}. Channel observations are possible also in RFID, where Dynamic Frame Aloha has been adopted in the last years [12], [13]. Here, 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 maximum throughput is, once again, e^{-1} [14].

When no channel feedback is available, as in LANs and WLANs, the above mechanisms cannot be used. Since a constant retransmission probability cannot stabilize the protocol, subsequent studies had to consider a retransmission probability that changes according to the user's own history. The only mechanism of this type so far considered, and called *backoff*, reduces the retransmission probability $\beta(i)$ as the number of collisions i suffered by the packet increases, on the ground that the number of collisions suffered measures the channel congestion level.

Manuscript received February 22, 2017; revised September 8, 2017; accepted December 6, 2017; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor J. Xie. Date of publication December 28, 2017; date of current version February 14, 2018. Some of the results in this paper were partially presented at IEEE INFOCOM 2016 [1] and at IEEE MedHocNet 2016 [2] conferences. *(Corresponding author: Luca Barletta.)*

The authors are with the Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, 20133 Milan, Italy. (e-mail: luca.barletta@polimi.it; flaminio.borgonovo@polimi.it; ilario.filippini@polimi.it).

Digital Object Identifier 10.1109/TNET.2017.2782696

First studies have considered exponential backoff (EB) of the memoryless type, which decreases the station transmission probability according to the negative exponential law

$$\beta(i) = b^{-i-i_0}, \tag{2}$$

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. This choice, with $b = 2$ and $i_0 = 0$, has been made in many past works, e.g. [15]–[18], mainly because it allows a Markovian model of the system. Unfortunately, due to the complexity of the process involved, the results attained in [15]–[18] appear to be limited and somehow contradictory and confusing. This is caused by the many differences in stability definitions and the assumptions underlying the analyzed models. Even resorting to simulations does not appear to be appropriate for a thorough analysis of the behavior of the system, as for some of the parameter settings here considered, the convergence to steady-state values is very slow.

Subsequently, a great number of studies have considered backoff dealing with IEEE 802.11. Actually, backoff was introduced in 1976 in the carrier sense multiple access with collision detection (CSMA-CD) protocol in the Ethernet LAN, later IEEE 802.3, and then adopted with CSMA-CA (collision avoidance) in IEEE 802.11. These applications use binary exponential backoff (BEB), in which the backoff value is taken as a uniform variable in a window with length 2^{i+i_0} slots. Although all those protocols seem to differ from S-Aloha, it happens that the latter is embedded in all of them, and their throughput S is strictly related to the S-Aloha throughput λ_0. In fact, if we define the slot of unit length equal to the signal propagation time, during which Carrier Sensing has no effect, the throughput of IEEE 802.11 can be expressed as

$$S = \frac{P_t P_s T}{P_t P_s T_b + P_t(1 - P_s)T_c + 1 - P_t}. \tag{3}$$

The formula above is taken from [19], a paper much referred to in the literature, where T is the packet transmission time in slots, T_b is the time the channel is sensed busy upon a success (it includes T and, possibly, RTS, CTS, ACK), T_c the time wasted on a collision, P_t is the transmission probability, P_s the success probability conditional to transmission. With this notation we have $\lambda_0 = P_t P_s$ and (3) can be rewritten as

$$S = \frac{\lambda_0\ T}{\lambda_0\ T_b + P_t(T_c - 1) - \lambda_0 T_c + 1}, \tag{4}$$

which yields λ_0 in the Aloha case, i.e., $T = T_c = T_b = 1$.

For the reasons already cited, studies on IEEE 802.11 have introduced simplified analytical models and approximations. Among the most relevant ones, the *saturation model*, first introduced in [19], is based on the assumption that queues are always full, such that as soon as a packet is successfully transmitted 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 throughput, thus guaranteeing the stable behavior and the throughput of the more realistic one.

To investigate throughput, an additional strong assumption, introduced in [19] and known as the *decoupling assumption*, has been largely used [20]–[24]. This assumption has a twofold implication, i.e., the *stationary* behavior of the model, and the *independence* in the behavior of the different transmitters. Using both assumptions above leads to a mean value analysis (MVA) and a fixed point equation [19], that provides in a simple way the (approximated) basic performance figures of the protocol.

A slightly different approach, again for IEEE 802.11, appears in [25], where law (2) is adopted, yielding a Markovian description of the system. The decoupling is not assumed, and, to face complexity, the authors resort to approximated analysis based on the equilibrium point, first introduced for S-Aloha in [6] and [8], and then largely used in subsequent works, notably in [26]. The authors show that their results are practically coincident with those attained by the decoupling assumption.

Finally, in [24], a slightly different version of the decoupling assumption is used, where the traffic on the channel is approximated by the Poisson distribution, rather than the Binomial distribution, as it is suitable for a large N.

It must be noted that the decoupling assumption, as well as the equilibrium point analysis, implies stationarity and, therefore, they cannot be adopted to investigate stability. Recently, the conditions for the stability/instability of S-Aloha with EB as in (2), and queues in saturation, have been given in [27].

In this paper we are interested in the analysis of the throughput and the delay of S-Aloha with unlimited backoff stages, when the number of users N is unknown and possibly very large, as it can be the case in future applications. To this purpose, the saturation model is a valid model to derive the performance in the congested situation; however the analyses in the literature that adopt the decoupling assumption, are unsatisfactory in some aspects. Most of them report throughput results directly related to IEEE 802.11, like in (4), or other metrics, that strongly depend on the choice of the specific parameters such as T, T_b and T_c in (4), making it difficult to extract information about the underlying S-Aloha. Furthermore, the elaborate passages required by the MVA, and the lack of close-from results, often obscure the role of parameters.

Moreover, the decoupling assumption itself is highly questionable in some cases. In fact, the most appreciated characteristic of Aloha is its flexibility, i.e., the ability to assign a large amount of the bandwidth when contention is limited, which requires small initial window size W_0, or i_0 in (2). On the other side, the decoupling assumption provides good results for high window size W_0, or i_0, as proven in Section V; it is highly questionable, or actually wrong, however, when small i_0 are concerned. As an example, Kwak *et al.* [20] use this assumption to derive the throughput of Aloha with unlimited backoff index, and the window-type EB. They find a maximum throughput equal to $\ln(b)/b$, independent of i_0, which yields $\ln(2)/2 \approx 0.34$ when $b = 2$, and e^{-1} for the optimized value of b, as compared to the better approximated values 0.3706 and 0.4303 attained, for $i_0 = 2$, with the model proposed in Section IV.

In our investigations we adopt the memoryless backoff (2), which allows the N backoff indexes to be modeled by a Markov Chain; on the other side, at least for the analysis with the decoupling assumption, it provides exactly the same results derived with the window-type backoff.

Our results are manifold. We first discuss the general model, stating new results and conjectures, although its complexity prevents the derivation of performance figures. Then we introduce an analytical approximated model, the Poisson model, which can be solved analytically, providing, in simple passages, close-form formulas for most of the metrics of interest. The numerical results of the latter model are the same as those in [24], and closely approximate, in absolute precision for $N \to \infty$, those attained by the decoupling assumption, and the window-type EB. Though still providing approximated results, the Poisson model has the advantage of simplicity, allowing an immediate comprehension of the effects of different parameters. As main result of the paper, we introduce a model that assumes a joint Poisson distribution of users only for high index stages, say from stage s onward, while retaining the actual joint distribution for stages before s, hence the name Semi-Poisson Model (SPM). The SPM provides results with an accuracy that increases with s. We show that for moderate values of s the joint stationary distribution is numerically computable, and that the results, for the parameter setting as used in practice, is quite accurate. Furthermore, we provide analytical tools for evaluating the access delay distribution and moments, proving its heavy tail property. Finally, preliminary results about the optimal parameter setting are discussed.

Numerical results show that choosing base $b = 1.35$ with $i_0 = 2$ provides better throughput than the standard value $b = 2$, in fact reaching 0.43 for an infinite number of users, as compared to 0.37. The access-delay analysis shows that, in any cases, delay variance does not exist when $N \to \infty$ for any i_0, and that a limit N_v on N must be considered to guarantee finite delay variance. Our results show that N_v increases with i_0, and that, constraining the bandwidth provided to a user alone on the channel, b^{-i_0}, the value $b = 1.35$ provides, again, better performance.

The paper is organized as follows. In Sec. II we derive some basic relations and results about the model. Section III is devoted to the Poisson model and its relationship with the decoupling assumption. In Sec. IV we present the semi-Poisson model (SPM) and the numerical procedure that provides the throughput, whereas in Sec. V some throughput results are presented. In Sec. VI we characterize the access-delay distribution and the constraint it poses on the number of users N, finally in Sec. VII we investigate the optimal parameter setting. Conclusions are given in Sec. VIII.

II. Preliminary Results

The system is described by a Markov chain whose Markovian state is $(N_0, N_1, \ldots, N_i, \ldots)$, where N_i, $\sum N_i = N$, denotes the number of users with backoff index i, whose expectation is n_i. Markovity is guaranteed by law (2), since knowing $(N_0, N_1, \ldots, N_i, \ldots)$ at time t, according to the outcomes of transmissions, we can determine

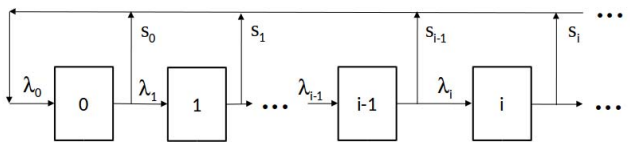

Fig. 1. Scheme of the backoff mechanism as a network of stages.

$(N_0, N_1, \ldots, N_i, \ldots)$ at time $t+1$. In the following we assume $b > 1$ and $i_0 > 1$, which guarantee [27] a stationary behavior of the chain. The whole system can be conveniently represented by a network of stages, as shown in Fig. 1, where λ_i, the flow rate across stage i, is given by the transmission rate of users in stage i, and s_i is the corresponding throughput.

Some performance metrics can easily be derived as averages once the joint distribution of the N_i's is known. The throughput of stage i can be expressed as

$$s_i = \mathsf{E}\left[\gamma_i(1, N_i) \prod_{k=0, k\neq i}^{\infty} \gamma_k(0, N_k)\right], \tag{5}$$

where

$$\gamma_i(k, N_i) = \binom{N_i}{k} \left(b^{-i-i_0}\right)^k \left(1 - b^{-i-i_0}\right)^{N_i - k}$$

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

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

The "routing" probability out of stage i, defined as $\alpha_i = \frac{\lambda_{i+1}}{\lambda_i}$, is given by:

$$1 - \alpha_i = \frac{s_i}{\lambda_i} = \frac{s_i}{n_i} b^{i+i_0}, \tag{6}$$

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

We also have

$$\lambda_i = \lambda_0 \alpha_0 \alpha_1 \ldots \alpha_{i-1}, \qquad i \geq 1, \tag{7}$$

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

$$N = \sum_{i=0}^{\infty} n_i = \sum_{i=0}^{\infty} \lambda_i T_i = \sum_{i=0}^{\infty} \lambda_i b^{i+i_0}, \tag{8}$$

where $T_i = b^{i+i_0}$ represents the average time the user stays in stage i. Since N is a given parameter, and sequence $\{\lambda_i\}$ exists because we are in stationary conditions, the convergence of (8) must be assured. In turn, this requires, by the ratio test,

$$\frac{\lambda_{i+1}}{\lambda_i} = \alpha_i < 1/b,$$

which can be translated into the existence of the limit:

$$\lim_{i \to \infty} \alpha_i = \alpha \leq 1/b, \tag{9}$$

where the equality sign holds for $N = \infty$.

Considering the throughput expression (5), we can write

$$1 - \alpha_i = \frac{\mathsf{E}\left[\gamma_i(1, N_i) \prod_{k=0, k\neq i}^{\infty} \gamma_k(0, N_k)\right]}{n_i b^{-i-i_0}}. \tag{10}$$

The following theorem, whose proof is given in the Appendix, provides a useful asymptotic result:

Theorem 1: In stationary conditions, i.e., for $b > 1$ and $i_0 > 1$, we have

$$\lim_{i\to\infty}(1-\alpha_i^{(N)}) = \lim_{i\to\infty} \mathsf{E}^{(N)}\left[\prod_{k=0,k\neq i}^{\infty}\gamma_k(0,N_k)\right] = P^{(N-1)}(idle) \tag{11}$$

where the superscript $^{(N)}$ denotes that the statistics have been evaluated with N users.

From (11) and (9) we get also

$$P(\text{idle}) \geq \frac{b-1}{b}, \tag{12}$$

where the equality sign holds for $N = \infty$.

Eq. (12) shows that BEB requires $P(\text{idle}) \geq 0.5$, showing that it can not completely exploit the channel, and that a small value of b is needed to relax this constraint.

Other interesting properties are suggested by strong arguments, but cannot be proved without knowing the behavior of the joint distribution of $(N_0, N_1, \ldots, N_i, \ldots)$. Unfortunately, this distribution can not be analytically studied and, therefore, these properties, useful to understand results, can only be stated as conjectures, as follows.

Defined the channel traffic as $\Lambda = \sum_i \lambda_i$, we have:

Conjecture 1: For any $N \geq 2$ and a given parameters setting, if N increases, then traffic Λ increases.

Conjecture 2: For any $N \geq 2$, an increase in b or i_0 decreases traffic Λ.

Conjecture 3: For any $N \geq 2$, if the change in b or i_0 increases traffic Λ, then $P(idle)$ decreases and vice-versa.

We note that result (12) explicitly proves Conjecture 2, with respect to parameter b and with an infinite number of users.

Although no proof can be provided, all the above conjectures are confirmed by our simulations and the approximated analysis presented in the next section. Furthermore, numerical results, derived in subsequent sections, have never confuted any of them.

Unfortunately, even the numerical solution of the chain is prevented by the complexity of the state space, which has an infinite number of dimensions and, looking at a possibly infinite N, each of them with an infinite number of states. We see in Sec. IV how some of these difficulties can be overcome in some cases of interest.

III. Poisson Model

The Poisson model assumes that the stationary distribution of $(N_0, N_1, \ldots, N_i, \ldots)$, of the chain in the previous section, is a joint Poisson distribution, independent from stage to stage. Then, denoting by $\Lambda = \sum_{k=0}^{\infty}\lambda_k$ the average traffic on the channel, we have

Theorem 2:

$$\alpha_i = \alpha = 1 - e^{-\Lambda}, \tag{13}$$

$$\lambda_0 = \Lambda e^{-\Lambda}, \tag{14}$$

$$N = b^{i_0}\frac{\Lambda e^{-\Lambda}}{1-b(1-e^{-\Lambda})}, \tag{15}$$

$$\Lambda \leq \ln\frac{b}{b-1} = \Lambda^*. \tag{16}$$

where the equality in (16) holds for $N = \infty$.

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 $\lambda_i = n_i b^{-i-i_0}$. Since the sum of independent Poisson variables is still Poisson distributed, results (13) and (14) immediately follow. By substituting into (8) and using (14) we prove (15) as

$$N = \lambda_0\sum_{i=0}^{\infty}(1-e^{-\Lambda})^i b^{i+i_0} = \lambda_0\ b^{i_0}\frac{1}{1-b(1-e^{-\Lambda})}. \tag{17}$$

Since N is given, and the convergence of (17) must hold, inequality (16) is verified. ∎

Given the number of users in the system, N, equation (15) provides the channel traffic Λ, while (14) gives the throughput.

The throughput curve (14) as function of Λ reaches its maximum in $\Lambda = 1$, but the model works up to $\Lambda = \Lambda^* = \ln\frac{b}{b-1}$, approximately equal to 0.693 for $b = 2$, where the throughput is $\lambda_0 = \ln(2)/2 \approx 0.346$. This reflects the well known fact that the BEB is overreacting: As N increases some users are pushed toward high backoff indexes so that, practically, they cannot contribute to Λ, although an increase of Λ would mean an increase in throughput.

Note that, if we want to maximize the throughput of this model in $\Lambda^* = \ln\frac{b}{b-1}$, 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 backoff basis equal to

$$b = \frac{1}{1-e^{-1}} \approx 1.582,$$

regardless of i_0.

Finally, the analysis of relation (15) proves the following

Theorem 3: With the Poisson model, Conjectures (1)*,* (2) *and* (3) *hold true.*

We note that the reduction in Λ predicted by Conjecture (2) can increase the throughput with N users if the change makes Λ closer to one, otherwise the throughput decreases.

At this point we can prove the

Theorem 4: The results provided in Theorem 2 and those derived with the decoupling assumption converge as $N \to \infty$.

Proof: According to [19]–[21], the decoupling assumption makes $\alpha_i = \alpha$, for all i, and also makes the asymptotic transmission probability of a single user $\hat{\beta}$ independent from user to user, in such a way that, with N users we have

$$\alpha = 1 - (1-\hat{\beta})^{N-1} \approx 1 - e^{-(N-1)\hat{\beta}}. \tag{18}$$

For N such as $N-1 \approx N$, the Binomial distribution approaches the Poisson distribution [21], and $\Lambda \approx (N-1)\hat{\beta}$, i.e., returning (13). ■

Approximation (18) was also used in [24]. The maximum throughput and channel traffic we have derived above for $b=2$ and $b=1.58$ were already derived in [20].

Remark 1: From all the above, it appears that the Poisson model is less accurate for small values of N, due to the fact that in (13) Λ represents the whole traffic and not the one due to the $N-1$ users possibly interfering. To take care of this, in Theorem 2, expression $e^{-\Lambda}$ should be changed into $e^{-\Lambda\frac{N-1}{N}}$. With this correction, the Poisson model can be made exactly equivalent to the decoupling assumption.

Remark 2: The decoupling assumption provides the same results if the window-type backoff is replaced by the memoryless one.

In fact, memoryless backoff was used, for example, in [21] and [22].

In concluding this section, we note that the Poisson model changes (5) into

$$s_i = \gamma_i(1,n_i)\prod_{k=0,k\neq i}^{\infty}\gamma_k(0,n_k). \tag{19}$$

With the real distribution of N_i, the passage from (5) to (19) is allowed only when functions $\gamma_i(n_i)$ are linear, which is not, actually, true. Nevertheless, at high-index stages functions $\gamma_i(n_i)$ become indeed linear in the region where N_i is concerned, showing that using the Decoupling/Poisson assumption only for these stages, provides asymptotically correct results.

IV. Semi-Poisson Model

We have seen that the system is modeled by a Markov chain with state $(N_0,N_1,\ldots,N_i,\ldots)$, which has an infinite number of dimensions, each of them with $N+1$ values; since we are interested in large and possibly infinite N, this prevents not only an analytical investigation, but also a numerical solution of the chain. On the other hand, the decoupling/Poisson assumption makes the system analyzable, but introduces an approximation that only provides a lower bound to maximum throughput, as shown in Sec. V-A.

Here we exploit the fact that the average number of users decreases at higher index stages; this, in addition to the remark at the end of Sec. III, suggests adopting the Poisson model only for such stages, say from stage s onward, introducing the Semi-Poisson Model (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. 2, with average

$$\Lambda_s = \sum_{i=s}^{\infty}\lambda_i = \sum_{i=s}^{\infty} n_i b^{-i-i_0}. \tag{20}$$

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. 2. Assuming Λ_s as a known constant, we can determine the behavior of the Markov process with state $(N_0,N_1,\ldots,N_{s-1})$, whose stationary distribution and cumulative average,

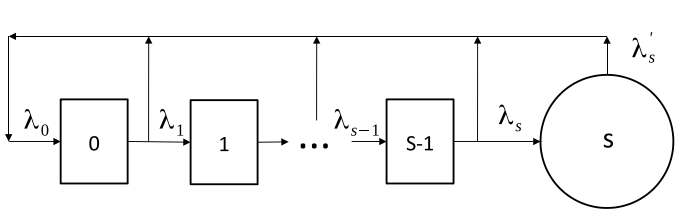

Fig. 2. Scheme of the SPM with stages $0,1,\ldots s-1$, and the lumped stage modeled with a Poisson distribution.

$$n^{(<s)} = \sum_{i=0}^{s-1} n_i,$$

depend on Λ_s. To complete the model, we need to set a relationship between Λ_s and $n^{(\geq s)}=\sum_{i=s}^{\infty} n_i$ such that we can fix the number in the system:

$$n^{(<s)} + n^{(\geq s)} = N(\Lambda_s). \tag{21}$$

Since the number N is given, (21) allows the determination of $\Lambda_s(N)$ through an iterative procedure, exactly as it happens with (14).

In order to derive the sought relationship, let us denote by τ_s the stationary probability that no user with index $i\leq s-1$ transmits, i.e.,

$$\tau_s = \mathsf{E}\left[\prod_{i=0}^{s-1}\gamma_i(0,N_i)\right],$$

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 (20), we have

$$\begin{aligned} s_i &= \lambda_i - \lambda_{i+1} \\ &= \tau_s n_i b^{-i-i_0} e^{-n_i b^{-i-i_0}} e^{-\sum_{j\neq i, i=s}^{\infty} n_j b^{-j-i_0}} \\ &= \tau_s \lambda_i e^{-\Lambda_s}, \quad i\geq s. \end{aligned}$$

Recognizing that $\tau_s e^{-\Lambda_s} = P^{(N)}(\text{idle})$ is the probability that the channel is idle, we finally have

$$1-\alpha_i^{(N)} = 1-\alpha_s^{(N)} = \frac{s_i}{\lambda_i} = P^{(N)}(\text{idle}), \quad i\geq s, \tag{22}$$

independent of i as expected.

Using the passages as in Th. 2, we get

$$\begin{aligned} \lambda_i &= \lambda_s \alpha^{i-s} = \lambda_s \left(1-P(\text{idle})\right)^{i-s}, \quad i\geq s, \\ n_i &= \lambda_s \left(1-P(\text{idle})\right)^{i-s} b^{i+i_0}, \qquad i\geq s, \\ n^{(\geq s)} &= \sum_{i=s}^{\infty} n_i = \lambda_s \sum_{i=s}^{\infty} \left(1-P(\text{idle})\right)^{i-s} b^{i+i_0} \\ &= \lambda_s b^{s+i_0} \frac{1}{1-b\left(1-P(\text{idle})\right)}, \end{aligned} \tag{23}$$

where the summation, which must converge due to stationary conditions, exists for $b\left(1-P(\text{idle})\right) < 1$, again confirming (12).

Referring to Fig. 2, we have

$$\lambda_s = \lambda'_s = \tau_s \Lambda_s e^{-\Lambda_s} = \Lambda_s P(\text{idle}),$$

and using this into (23) we finally have

$$n^{(\geq s)} = \frac{\Lambda_s P(\text{idle}) b^{s+i_0}}{bP(\text{idle}) - (b-1)}. \tag{24}$$

The existence of (24) only for $P(\text{idle}) > \frac{b-1}{b}$ still confirms (12).

Using $n^{(<s)} = N - n^{(\geq s)}$, (24) can be written as

$$\Lambda_s = (N - n^{(<s)}) b^{-s-i_0} \left(b - \frac{b-1}{P(\text{idle})} \right). \tag{25}$$

Either (24) or (25) represent the required relationship that makes the system determined. Assuming Λ_s as a known constant, then $(N_0, N_1, \ldots, N_{s-1})$ is the state of a Markov process whose stationary distribution provides λ_s, τ_s, $n^{(<s)}$, and $P(\text{idle})$, all of which depend on Λ_s. Therefore, the correct Λ_s for the model is the one that makes equal both sides of (25), which is reached through an iterative procedure. Once the correct Λ_s is found, the throughput can be evaluated as

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

and for $i \leq s-1$

$$1 - \alpha_i = \frac{e^{-\Lambda_s} \mathsf{E}\left[\gamma_i(1, N_i) \prod_{k=0, k\neq i}^{s-1} \gamma_k(0, N_k) \right]}{n_i b^{-i-i_0}}. \tag{26}$$

A. Model Accuracy

The accuracy of the model in evaluating the throughput of the system is expected to increase with s since, in stationary conditions, the contribution Λ_s of the lumped stage to the traffic decreases to zero as s increases to infinity, where the model produces exact throughput results. Therefore we expect the ratio Λ_s/Λ to provide a good accuracy indicator.

On the other hand, we must remark that, in evaluating other performance metrics, the SPM model can introduce small discrepancies, even in the limit $s \to \infty$. For example, we see in Sec. VI that the tail of the access-delay distribution is derived from the asymptotic value of α_i as $i \to \infty$. This, in the SPM, is given by (22), which equals $1 - P^{(N)}(\text{idle})$, whereas the real asymptotic value $\alpha^{(N)}$ is given by (11), and is equal to $1 - P^{(N-1)}(\text{idle})$. Therefore, to get $\alpha^{(N)}$ we obviously use the latter, where the required $1 - P^{(N-1)}(\text{idle})$ is attained by the SPM (22). The above discrepancy vanishes as $N \to \infty$.

In any case, with finite s, the model only provides approximate results, whose accuracy depends on s, to be traded against complexity, as discussed in the next section. On the other side, the accuracy for a given s depends on the system parameters b and i_0. The results reported in this paper refer to parameters values that are encountered in practice, for which we can get almost absolute precision.

B. 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 $N_{\max} \leq N$. This is a reasonable approach, as the averages n_i and variances $\mathsf{Var}[N_i]$ are finite. The transitions that lead to states with $N_i > N_{\max}$ 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 $N_{\max}$ and thus the computational cost. The accuracy of the approximation can be estimated by the low values of $P(N_i = N_{\max})$ 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 $(N_{\max}+1)^s$. The numerical limits of the model are discussed in Sec. V. However, the fact that for $b = 2$ and $i_0 = 2$ the sequence $\{n_i\}_i$ starts with $\mathsf{E}[N_0] \approx 1.5$ and decreases with i, suggests that limited values of $N_{\max}$, such as 5 or 6, are enough to make the approximation error negligible in this case.

In Sec. VI we need the asymptotic value of α_i which, for some values of N is reached at higher values of s than considered above. To this purpose, for small N, the constraint $\sum_{i=0}^{s-1} N_i \leq N$ has been enforced, which results in a number of states equal to

$$\binom{N+s}{s}, \tag{27}$$

much smaller than $(N_{\max}+1)^s$ for $N < (N_{\max})^s$.

C. 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 four 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
$$\Lambda_s e^{-\Lambda_s} \prod_{j=0}^{s-1} \gamma_j(0, N_j).$$
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
$$e^{-\Lambda_s} \gamma_i(1, N_i) \prod_{j=0, j\neq i}^{s-1} \gamma_j(0, N_j).$$
3) Collision, where C_i is the number of transmissions generated by stage i. The final state is $(N_0 - C_0, N_1 + C_0 - C_1, N_2 + C_1 - C_2, \ldots, N_{s-1} + C_{s-2} - C_{s-1})$, and the TP is
$$\begin{cases} \prod_{j=0}^{s-1} \gamma_j(C_j, N_j), & \text{if } \sum_{j=0}^{s-1} C_j \geq 2, \\ (1 - e^{-\Lambda_s}) \prod_{j=0}^{s-1} \gamma_j(C_j, N_j), & \text{if } \sum_{j=0}^{s-1} C_j = 1. \end{cases}$$
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 $N_{\max}$, the TP is set to zero. In this way, the TP matrix is square of order $(N_{\max} + 1)^s$.

V. Throughput Results

In [27], we have proved that the system is stable for a large range of parameters values, namely for $b > 1$ and $i_0 > 1$. In this section we report the results of the SPM model for the parameter values that are most commonly used in the literature and in standards, for example using $b = 2$ and integer values of i_0, mostly $i_0 = 2$.

The model allows the derivation of the throughput as function of N. However, in this section we draw the throughput λ_0 against the channel traffic Λ, as it has been done historically with (1): Here Λ and λ_0 are assumed continuous, although in practice they are derived from N, and thus belong to a discrete set. The channel traffic increases with N and reaches a maximum, Λ^*, corresponding to $\Lambda_s = \Lambda_s^*$, as $N \to \infty$; here $P(\text{idle})$ reaches the limiting value $(b-1)/b$. It is also interesting to plot the throughput for values beyond Λ^*, 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 values of Λ_s beyond Λ_s^*, and disregard the constraint (21), which cannot be met in this zone with positive N. However, it must be noted that increasing Λ beyond Λ^* 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 maximum throughput Λ^* can be approximated by looking for the value Λ_s, the model's input variable, that makes $P(\text{idle})$ approach from above the limiting value $(b-1)/b$. In this critical range the model is highly sensitive, since a very small increase in Λ_s causes great changes in N, but negligible changes in the other variables, so that Λ^* can be approximated very finely.

A. Binary Exponential Backoff

The results for BEB law with $i_0 = 2$ 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 $s = 5$. These limitations do not introduce significant errors, as can be inferred by the content distribution of the first five stages, as shown in Table I.

In Fig. 3 we report the throughput curves, as function of channel traffic Λ, for different values of s, namely $s = 0 - 5$, corresponding to different accuracy degrees of the model. We note that the curve for $s = 0$ represents the throughput attained with the Poisson model, which corresponds to (14), 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 Λ^*, whose coordinates are reported for a precise comparison in Table II. We see that for the same Λ the throughput increases as s increases showing that, using the joint distribution provided by the model, the throughput converges to a different value compared 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 Λ. This is also the reason why Λ^* 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

TABLE I

BEB: Marginal Distributions of the Content of Stage i, for $s = 5$ and $\Lambda = \Lambda^*$

	$i = 0$	$i = 1$	$i = 2$	$i = 3$	$i = 4$
n_i	1.482	1.244	1.061	0.9549	0.9012
$N_i = 0$	0.1649	0.2649	0.3289	0.3705	0.3957
$N_i = 1$	0.3643	0.3640	0.3824	0.3829	0.3777
$N_i = 2$	0.3204	0.2570	0.2062	0.1817	0.1687
$N_i = 3$	0.1262	0.09251	0.06633	0.05272	0.04714
$N_i = 4$	0.02222	0.01910	0.0140	0.01049	0.00927
$N_i = 5$	0.001839	0.002260	0.00201	0.00151	0.00136
$N_i = 6$	0.000077	0.000154	0.000192	0.000154	0.000143

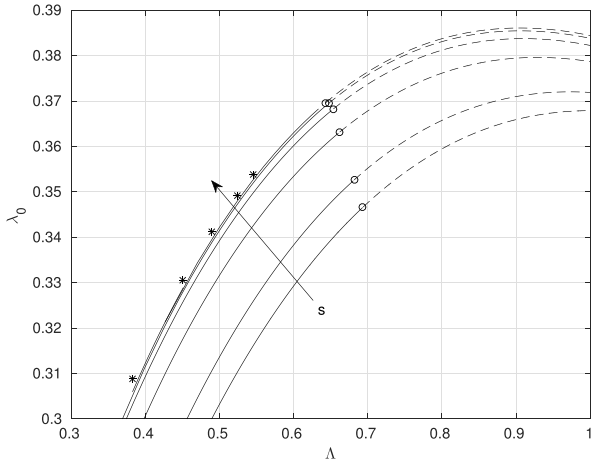

Fig. 3. BEB, $i_0 = 2$: Throughput of the SPM with $s = 0, 1, 2, 3, 4, 5$ versus traffic channel Λ. The arrow shows the direction of increase of s. The lowest curve, $s = 0$, represents the Poisson model, whereas circles represent the maximum achievable throughput, i.e., throughput values evaluated at different Λ^* for each s. From left to right, asterisks show Monte Carlo simulation results obtained for $N = 2, 3, 4, 5, 6$.

TABLE II

BEB: Throughput λ_0 of the SPM With $s = 0, 1, 2, 3, 4, 5$, With $\Lambda = \Lambda^*$. The Channel Traffics Λ^* and Λ_s^* Are Also Reported

	$s = 0$	$s = 1$	$s = 2$	$s = 3$	$s = 4$	$s = 5$
Λ^*	0.6931	0.6817	0.6629	0.6542	0.6512	0.6501
Λ_s^*	0.6931	0.3291	0.1415	0.0622	0.0287	0.0138
λ_0	0.3466	0.3526	0.3633	0.3683	0.3700	0.3706

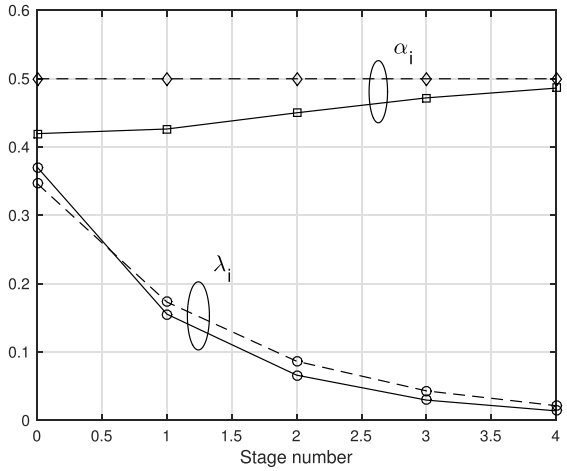

Fig. 4. BEB, $i_0 = 2$: The behavior of curves α_i and λ_i versus i with $s = 5$ and $\Lambda = \Lambda^*$. The same curves for the Poisson model (dashed line) are also reported.

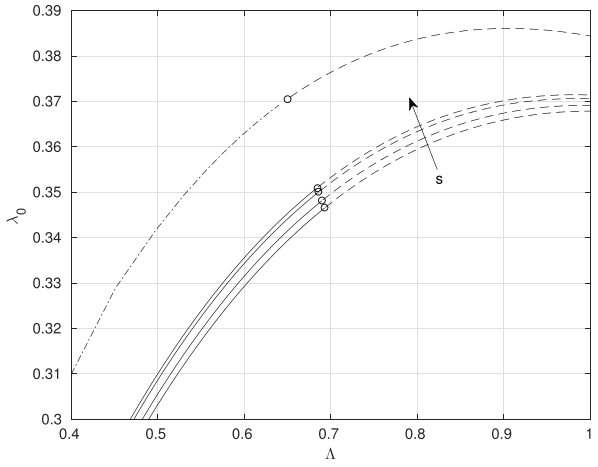

Fig. 5. BEB, $i_0 = 4$: Throughput of the Poisson model (bottom line) and the SPM with $s = 1, 2, 3$, versus traffic channel Λ. The dot-dash curve represents the throughput of BEB, $i_0 = 2$, with $s = 5$; circles represent the maximum achievable throughput.

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 effective convergence of the SPM is reached. Remarkable is the estimate of the maximum throughput 0.3706 reached with $s = 5$. Actually, this value has been reached with $\Lambda_5 = 0.01379$, and provides $P(\text{idle}) = 0.50003$ and $\Lambda = 0.65016$, which correspond to $N = 62154$. The attained throughput bound is well above the value $\ln(2)/2 = 0.3466$ predicted by the decoupling/Poisson assumption. We note also that the maximum channel traffic $\Lambda = 0.65016$ is below the one predicted by such assumption and equal to $\ln(2) = 0.6931$. Moreover, the results confirm that, as predicted by (11) and (9), in the limit, $P(\text{idle})$ is the same for the two models. Therefore, getting the same $P(\text{idle})$ with smaller traffic means that transmissions are more dispersed than with the Poisson model, while getting higher throughput reveals that among transmission slots the interference is more concentrated.

Figure 4 shows the performance of the model with $s = 5$ at $\Lambda = \Lambda^*$ with the same parameter setting as in Fig. 3. In particular we report α_i as function of i, showing that indeed this sequence is non-constant, compared to the constant $\alpha_i = 0.5$ predicted by the Poisson model. We can appreciate that the difference in α_0 is substantial. We also report the curve λ_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 I presents the marginal distributions of N_i, $i = 0, \ldots, 4$ for $s = 5$, $\Lambda = \Lambda^*$, and $N_{\max} = 6$. Again, the very low values reached in $N_i = 6$ show that increasing $N_{\max}$ above 6 does not appreciably increase the precision of the model. Then, we can conclude that the maximum throughput 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 n_i, which now is about 5.6 in the first stage. The increase in 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 n_i requires $N_{\max} = 10$, 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 the maximum throughput is about 0.351 at $\Lambda^* = 0.6845$.

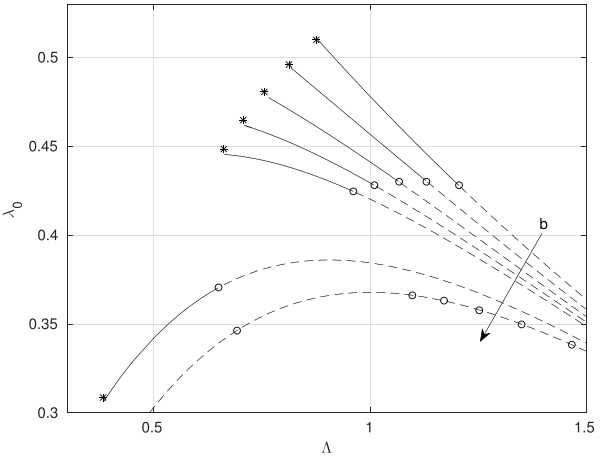

Fig. 6. $i_0 = 2$: Throughput of the SPM versus traffic channel Λ, with $s = 5$ for $b = 2$, and $s = 6$ for $b = 1.5, 1.45, 1.4, 1.35, 1.3$. Circles represent the maximum achievable throughput. The throughput under Poisson assumption is reported in dashed line. Asterisks correspond to Monte Carlo simulation with $N = 2$.

B. Optimal Exponential Backoff With $i_0 = 2$

The Poisson model indicates $b \approx 1.58$ as the optimal backoff base with maximum throughput e^{-1}, regardless of the value of i_0, which is not the case with our model, where, given a base b, the maximum throughput decreases as i_0 increases. 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 throughput are reported. Also reported on the curves are asterisks representing the throughput attained by simulating the real system for $N = 2$.

TABLE III

MAXIMUM THROUGHPUTS AND THE CORRESPONDING TRAFFIC CHANNELS FOR SEVERAL VALUES OF b AND FOR $i_0 = 2$

	$b = 2$	$b = 1.5$	$b = 1.45$	$b = 1.4$	$b = 1.35$	$b = 1.3$
Λ^*	0.6502	0.9612	1.0109	1.0669	1.1309	1.2058
Λ_s^*	0.0138	0.0511	0.0652	0.0847	0.1121	0.1519
λ_0^*	0.3706	0.4247	0.4279	0.4300	0.4303	0.4279

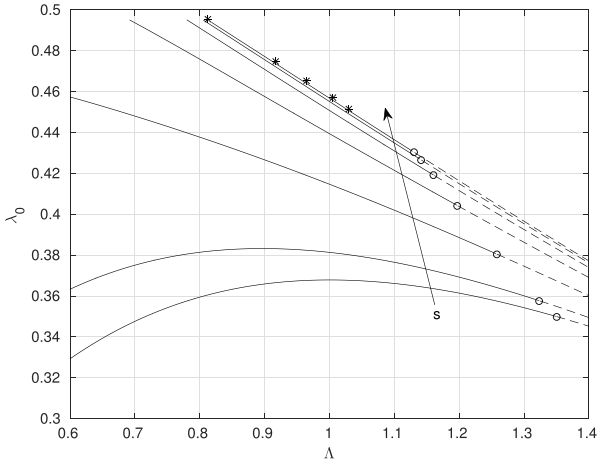

Fig. 7. $i_0 = 2$: Throughput of the SPM with $b = 1.35$, versus traffic channel Λ and for $s = 0, 1, 2, 3, 4, 5, 6$. Circles represent the maximum throughput. Asterisks represent simulation results for $N = 2, \ldots, 6$.

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 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.

Table III reports the SPM maximum throughputs corresponding to the circles in Fig. 6. We see that the maximum throughput 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 n_i from $b = 1.5$ and lower values allows additional precision, and a further stage is added ($s = 6$) 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 $b = 1.35$.

C. Further Results

The results shown in the preceding section agree with the Conjectures exposed in Section II. In particular, Conjecture 1 is confirmed by the asterisks reported in Figs. 3, 5 and 7, representing different values of N. Conjecture 2 is confirmed by the comparison between Figs. 3 and 5, and by Fig. 8. Numerical results, not reported here, also confirm Conjecture 3.

In the above figures the throughput λ_0 is plotted as function of the channel traffic Λ, as it was more suitable for the validation of the SPM model. Here, instead, we want to provide some insights on the parameter setting for the proper backoff operation. To this purpose we plot results against the number of users N, as it is required for practical purposes.

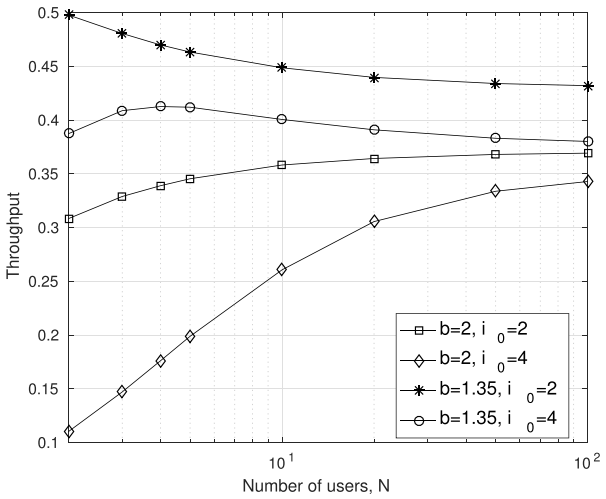

Fig. 8. Throughput versus number of users N.

In Fig. 8 the throughput curves for four different parameter choices are reported. The figure clearly shows the effects of parameters i_0 and b as predicted by Conjecture 2. From the throughput point of view the best choice seems to be $b = 1.35$ and $i_0 = 2$. However, as we see in the next section, this advantage does not come for free; in fact, the throughput increase is associated with an impairment of the access delay, since the capture of the channel by low index users is magnified. This effect is reduced when considering the curve with lowest throughput. However, in this case, the flexibility of the protocol, i.e., the capability to assign a large amount of bandwidth also when only a few users are active, is greatly reduced.

Things are somewhat different with many of the IEEE 802.11 protocols, whose throughput is given by (4). Since it usually happens that $T \gg 1$, while T_b and T_c are much smaller than T, a change in λ_0 corresponds to a smaller change in S, and, with only one user the throughput is close to 1 even at low values of λ_0.

VI. ACCESS DELAY

A. Analysis

The access delay is the time interval between two consecutive accesses by the same user. It can be evaluated as

$$D_R = \sum_{k=0}^{R} T_k, \tag{28}$$

where T_k is the geometric random variable (RV) that represents the time spent in stage k by a user, and R is the random index of the stage where the transmission is successful.

The throughput λ_0, evaluated as described in Sec. IV, immediately provides the first-order delay moment as

$$d_R \triangleq \mathsf{E}[D_R] = \frac{N}{\lambda_0}. \tag{29}$$

As simple as it is, relation (29) has been overlooked in many of the works appeared in the literature. It is readily proved by Little's Result and by observing that the throughput of each

user is λ_0/N. On the other side, with the saturation model, the average number of packets in the transmission buffer is always one. This result shows that the setting that provides the highest throughput also gives the smallest average access delay.

Given $R = r$, the variates $\{T_k\}_{k=0}^{r}$ are independent and Geometric-distributed, respectively with averages $\{b^{i_0+k}\}_{k=0}^{r}$. The distribution of D_r has been derived in [29] in closed form, and we report it here specialized to this context:

$$p_{D_r}(x) = \sum_{k=0}^{r} b^{-i_0-k}(1-b^{-i_0-k})^{x-1} \prod_{i=0, i\neq k}^{r} \frac{1}{1-b^{-(k-i)}}, \quad x \geq r+1. \tag{30}$$

The result above provides the complementary cumulative distribution function (CCDF) of D_R as follows

$$P(D_R > d) = \sum_{r=0}^{\infty} P(R=r)P(D_r > d), \tag{31}$$

where the probability mass function of R, by (6) and (7), is

$$P(R=i) = \begin{cases} 1-\alpha_0 & i=0, \\ \alpha_0\alpha_1 \ldots \alpha_{i-1}(1-\alpha_i) & i \geq 1. \end{cases} \tag{32}$$

In the following we derive concise results about the tail of the CCDF of D_R.

Theorem 5: The random variable D_r/d_r converges in distribution, for $r \to \infty$, to a random variable Δ with pdf

$$p_\Delta(x) = \sum_{k=0}^{\infty} \frac{Q}{\prod_{i=0}^{k}(1-b^i)} b^{k+1} e^{-b^{k+1}x}, \quad x > 0, \tag{33}$$

where

$$Q = \prod_{i=1}^{\infty} \frac{1}{1-b^{-i}}. \tag{34}$$

Proof: First of all,

$$d_r = \sum_{k=0}^{r} b^{i_0+k} = b^{i_0} \frac{b^{r+1}-1}{b-1}. \tag{35}$$

For large r, it is standard to show that the scaled Geometric random variable T_r/d_r converges in distribution to an Exponential random variable with average b^{-1}. Pretty much in the same way, one can proceed backwards to show that T_{r-k}/d_r converges in distribution to an Exponential random variable with average b^{-1-k}, for any finite $k = 0, 1, 2, \cdots$. By (30), the weight associated with $\lim_{r\to\infty} T_{r-k}/d_r$ is

$$\prod_{i=0}^{k} \frac{1}{1-b^i} \prod_{i=1}^{\infty} \frac{1}{1-b^{-i}} = \frac{Q}{\prod_{i=0}^{k}(1-b^i)} \tag{36}$$

for $k = 0, 1, \cdots$, hence the thesis. ∎

Since the weights in (36) decrease quickly with k, one needs to consider only few terms to numerically evaluate (33). As an example, Fig. 9 shows the probability distribution of D_r/d_r with $r = 10$ in the two cases $b = 2$ and $b = 1.35$ and offset $i_0 = 2$. Evaluations of D_r/d_r for values of $r > 10$ show no changes.

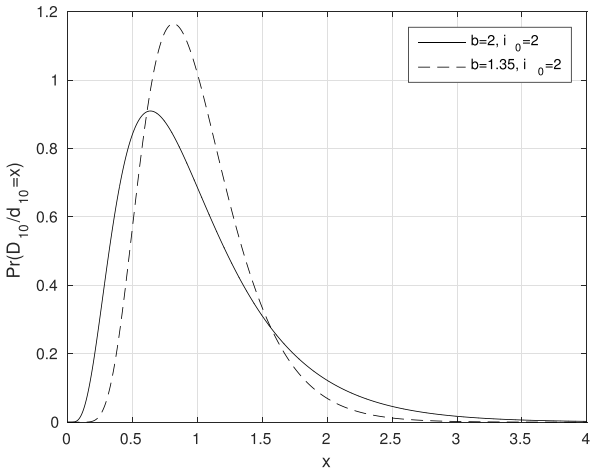

Fig. 9. Probability distribution of D_r/d_r with $r = 10$ in the two cases $b = 2$ and $b = 1.35$ and offset $i_0 = 2$.

Theorem 6: The tail of the CCDF of the access delay can be expressed as

$$P(D_R > d) \approx H(b, \zeta, i_0) d^{-\zeta}, \tag{37}$$

where ζ, which is a function of N, is defined by

$$\alpha = b^{-\zeta(N)}. \tag{38}$$

Proof: Given r_1 we have

$$\lim_{d\to\infty} P(D_{r_1} > d) = 0,$$

which shows that given any $\varepsilon > 0$ we can always find $d'(r_1, \varepsilon)$ such that for $d > d'$

$$\sum_{r=0}^{r_1-1} P(R=r)P(D_r > d) \leq \varepsilon.$$

Also, given d we have

$$\lim_{r\to\infty} P(D_r > d) = 1,$$

which shows that given any $\varepsilon' > 0$ we can always find $r_2(d, \varepsilon')$ such that for $r > r_2$

$$\sum_{r=r_2+1}^{\infty} P(R=r) - \varepsilon' \leq \sum_{r=r_2+1}^{\infty} P(R=r)P(D_r > d) \leq \sum_{r=r_2+1}^{\infty} P(R=r). \tag{39}$$

Since ε and ε' can be made as small as wanted, for $d > d'(r_1, \varepsilon)$ we can write

$$P(D_R > d) \approx \sum_{r=r_1}^{r_2} P(R=r)P(D_r > d) + \sum_{r=r_2+1}^{\infty} P(R=r).$$

Then, by Theorem 5, we can always find an integer r_1 so that delay (31) can be written as

$$P(D_R > d) \approx \sum_{r=r_1}^{r_2} P(R=r) P\left(\Delta > \frac{d}{d_r}\right) + P(R > r_2),$$

for $d \geq d'(r_1)$, where the approximation becomes equality in the limit $d \to \infty$. For $d = d_i$, by (35), we have

$$P(D_R > d_i) \approx \sum_{r=i-a_1}^{i+a_2} P(R=r) P\left(\Delta > \frac{b^{i+1}-1}{b^{r+1}-1}\right) + P(R > r_2) \quad (40)$$

where we have denoted $r_1 = i - a_1$ and $r_2 = i + a_2$. For large values of r_1, hence of r_2 and i, we can exploit (9) in (32) and write

$$P(R=r) \approx \mu_{r_1}(1-\alpha)\alpha^r, \quad r \geq r_1,$$
$$P(R>r) \approx \mu_{r_1}\alpha^{r+1}, \quad r \geq r_1,$$

where $\mu_{r_1} = \alpha_0\alpha_1 \cdots \alpha_{r_1-1}/\alpha^{r_1}$, so that (40) can be written as

$$P(D_R > d_i) \approx A\alpha^i, \quad i \geq r_1,\ d > d', \quad (41)$$

where

$$A = \mu_{i-a_1}(1-\alpha) \sum_{k=-a_1}^{a_2} \alpha^k P\left(\Delta > \frac{b^{i+1}-1}{b^{i+k+1}-1}\right).$$

Using (38) into (41) we have

$$P(D_R > d_i) \approx A(b,\zeta) b^{-\zeta i}, \quad i \geq r_1,\ d > d'. \quad (42)$$

From (35) we get

$$d_i^{\zeta} \approx \left(\frac{b^{i_0+1}}{b-1}\right)^{\zeta} b^{\zeta i},$$

which replaced into (42) provides

$$P(D_R > d_i) \approx A(b,\zeta)\left(\frac{b^{i_0+1}}{b-1}\right)^{\zeta} d_i^{-\zeta}. \quad (43)$$

Since we expect asymptotically a slow change in $P(D_R > d)$, almost a straight line in a log-log scale, we can safely extrapolate $P(D_R > d)$ for any large d from $P(D_R > d_i)$, which represents, in the same scale, a uniform sampling. This extrapolation procedure gives rise to the term $H(b,\zeta,i_0)$ in (37). ∎

Since (37) is represented by a straight line in a log-log scale, we dub ζ the *slope* of the tail. We then have:

Corollary 2: For $N = 2$ we have $\zeta = i_0$, and $\zeta = 1$ for $N = \infty$.

Proof: From (38) and (11) ζ can be written as

$$\zeta = -\frac{\ln \alpha}{\ln b} = -\frac{\ln(P^{(N-1)}(\text{idle}))}{\ln b}. \quad (44)$$

The thesis comes from

$$\alpha^{(2)} = 1 - P^{(1)}(\text{idle}) = b^{-i_0},$$
$$\alpha^{(\infty)} = 1 - P^{(\infty)}(\text{idle}) = \frac{1}{b}$$

∎

Conjecture 4: The CCDF tail has slope ζ that decreases from i_0 to 1 as N increases from 2 to infinity.

This comes from Conjectures 1 and 3 applied to (44).

Corollary 3: Necessary condition for the existence of the k-th order moment of the access delay is $\zeta > k$.

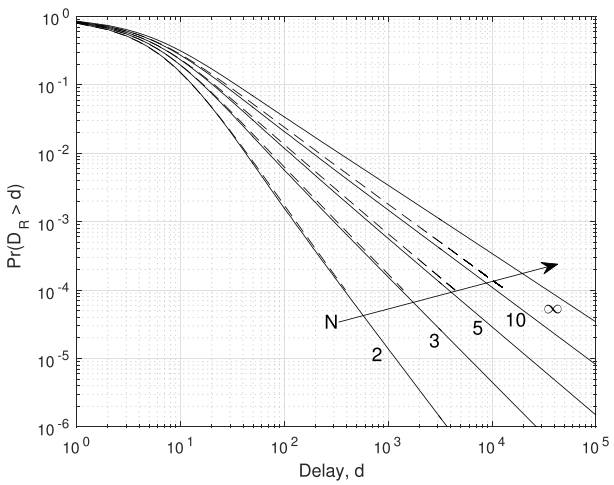

Fig. 10. The complementary cumulative distribution function of the access delay D_R for the cases $N = \{2, 3, 5, 10, \infty\}$, with parameters $b = 2$ and $i_0 = 2$. Solid line: analytic evaluation. Dashed line: Monte Carlo simulation.

Proof: From distribution (37) we can derive the moments of D_R as

$$\mathsf{E}\left[(D_R)^k\right] = \int_0^{\infty} k d^{k-1} P(D_R > d)\, \mathrm{d}\, d \approx H(b,\zeta,i_0) \int_0^{\infty} k d^{-(\zeta-k+1)}\, \mathrm{d}\, d, \quad (45)$$

from which the thesis descends as condition for the integral above to exists. ∎

Corollary 3 shows, for example, that by using $i_0 = 2$ we do not have any second order moment for $N \geq 3$. In any cases, with a given i_0, as N increases ζ tends to one, so that no second order moment can exist for any b.

We note that all the theoretical results stated in this section can be derived in the same way with the Poisson model, where α_i is assumed constant; however, numerical results are different. For example, using the Poisson model with $N = 2$, by (13) and (15) we get $\alpha = 0.27$ and $\zeta = 1.89$. Hence, once again, the Poisson model is pessimistic.

Before ending this section we must notice that the result of Corollary 3, expressed in the implicit form $\alpha < b^{-k}$, has already been derived in [30] referring to the IEEE 802.11 DCF access protocol, and window-type backoff, using the decoupling assumption.

B. Numerical Evaluations

The CCDF of D_R for the cases $N = \{2, 3, 5, 10, \infty\}$, obtained with (30) and (31), with parameters $b = 2$ and $i_0 = 2$, is shown in Fig. 10. We note again that for $N = 2$ and $N = \infty$ exact slopes are provided by the analysis, while in the other cases we have used the approximated solution provided by the SPM. The approximation encounters its limit in the maximum value of s allowed by the numerical complexity, and diminishes as N increases. To validate the curves we have also reported the dashed lines representing the sample CCDF's derived by simulating the BEB for 5×10^8 time slots, and for moderate values of N. As we can see, the measured results

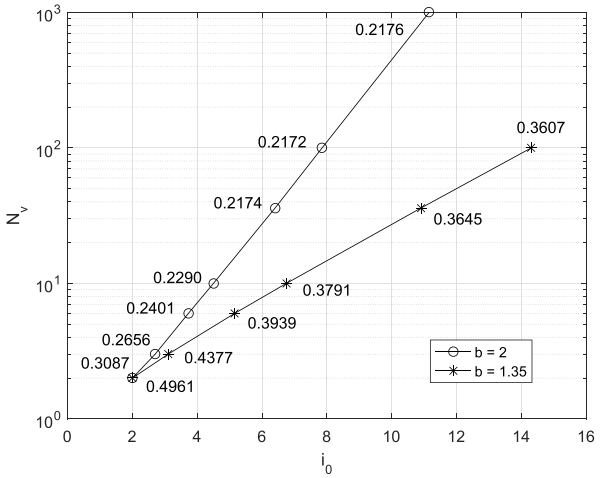

Fig. 11. Maximum value of N allowed for the existence of the variance as function of i_0, for two values of b.

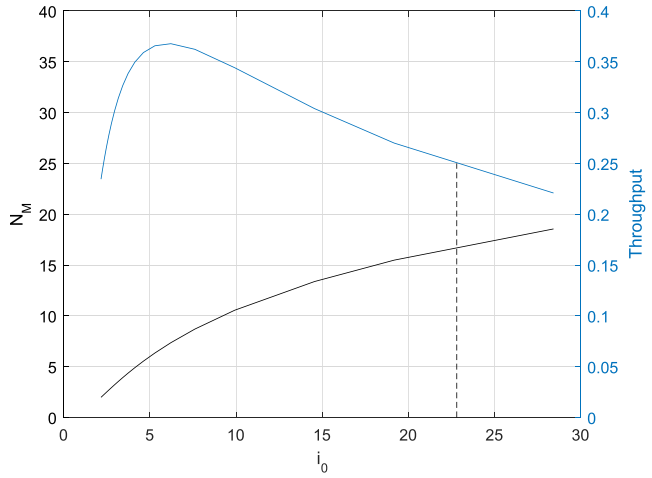

Fig. 12. Maximum number of users that can be admitted requiring the variance to exist, and assuming flexibility $b^{-i_0} = 0.25$, as function of i_0. The throughput curve is also reported on the right scale. The curves have been derived with the Poisson Model.

almost perfectly match with the evaluations attained by SPM. The small mismatch observed for $N = 5, 10$ cannot be safely ascribed either to lack of approximation of SPM or to slow convergence of the simulation.

As predicted by (37), the tails of the distributions are represented, in the log-log scale, by straight lines of constant slope, which decreases as N increases. In any case, the heavy-tail behavior of the delay distribution is confirmed.

VII. Optimal Parameter Setting

In the preceding section, Corollaries 2, 3, and Conjecture 4 show that, to improve the access delay, parameter i_0 should be chosen as high as possible, the opposite that is required to maximize throughput and flexibility. It appears that further investigations are required to determine optimal parameter setting, and what follows is a first step in that direction.

In order to set some delay bounds, we require the delay variance to exist. This forces the slope of the delay distribution to equal at least two. As the slope decreases as N increases, it appears that the number of users N must be limited. We expect that the limit on N increases with i_0. However, in increasing i_0 we cause, as a further impairment, the reduction in the protocol flexibility, i.e., the bandwidth assigned when only few users, among N, are active. In the following we define the protocol flexibility as the bandwidth assigned when only one user is active, i.e., b^{-i_0}.

Figure 11 shows the upper bound on N, $N_v(i_0)$, that guarantees the existence of the delay variance versus i_0, with $b = 2$ and $b = 1.35$. On the same figure we have reported on the curves the throughput values for some N_v.

We see that, in order to guarantee the existence of the variance, higher values of i_0 should be chosen as N_v increases, but this reduces the flexibility. For example, with $N_v = 10$ users, the variance exists with $i_0 \geq 4.5$ for $b = 2$, yielding $b^{-i_0} \approx 0.044$, while for $b = 1.35$ we get $i_0 = 7$, yielding $b^{-i_0} \approx 0.122$, almost three times as greater. Using the same flexibility parameter, 0.044, $b = 1.35$ provides $i_0 = 10.4$, corresponding to almost $N_v = 30$, again three times as greater than with $b = 2$, and with a 50% increase in throughput. This makes base $b = 1.35$ much more appealing than the standard value $b = 2$.

The straight-line behavior of curves in Fig. 11 can be explained by observing that on these curves, by Corollary 2, we require a constant value for the slope $\zeta = 2$, i.e., a constant value for α (38). By Theorem 1 this means a constant value for $P^{(N-1)}(\text{idle}) \simeq P^{(N)}(\text{idle})$, which in turn, by Conjecture 3, requires a constant value for the channel traffic Λ. Therefore, referring to the Poisson model, that becomes more accurate as i_0 increases and N is large, we can refer to relation (15), that reduces to

$$N = b^{i_0} H, \tag{46}$$

where H is a constant. Hence the straight line in Fig. 11 with a log-log scale.

The results above suggest the investigation of values of b less than 1.35, and greater i_0 values. Unfortunately, in this range of the parameters the SPM loses accuracy, since an higher approximation requires a large s, which increases the complexity beyond feasibility. We leave this investigation for further work. However, to enlighten the reader with the results we can derive, we produce some evaluations using the Poisson model.

Requiring the variance to exist constraints the slope in (38) to be at least two. Then by (13) we can write

$$1 - e^{-\Lambda(N)} \leq b^{-2}. \tag{47}$$

The above can be solved with respect to N by using (15), obtaining the maximum number of users N_M that can be admitted in order the variance to exist.

The results are shown in Fig. 12 where N_M is plotted against i_0, with flexibility $b^{-i_0} = 0.25$. We see that N_M increases with i_0; however, increasing i_0 too much has the effect of reducing the throughput, as it appears in the figure by the throughput curve. As an example, if we require a throughput not less than 0.25, we see that we can admit at most $N_M = 17$ users, using $i_0 = 22$ and $b = 1.065$. We can further

increase N_M by requiring smaller throughput constraint and/or flexibility; however, before claiming definite results, additional investigations are required, taking into account, for example, the transient behavior, i.e., the ability of the protocol to promptly react to changes.

VIII. Conclusion

In this paper we have introduced two analytical models, and produced analytical and numerical results, to investigate the Exponential Backoff, the most popular random-access control mechanism so far introduced. The Poisson model provides close-form results for the approximate model that uses the well known "decoupling assumption." The semi-Poisson model is able to produce fairly accurate numerical results for throughput and access delay, while keeping the complexity at a minimum. Our results show that in order to satisfy access-delay constraints and provide a fair throughput the number of users must be limited. In this respect, we find that the usual setting of parameters, such as the binary base of backoff, is far from being optimum. Further work is needed to investigate the optimal setting and, perhaps, a better backoff law.

Appendix
Proof of Theorem 1

Referring to (10), let focus on times $k_1, k_2, \ldots, k_r, \ldots$ where a user with index i transmits. Loosely speaking, since we are interested in the limit $i \to \infty$, we can take i so large that no more than one user has ever an index greater or equal than i. This means that, when a user has index i, this index is the largest among all users. In the time interval $[k_r, k_{r+1}]$, between two consecutive transmissions of index i, the largest index reach again i, say at time k'_r, and from this point up to k_{r+1} the remaining $N-1$ smaller indexes behave as in a system with exactly $N-1$ users. Next we prove that in the limit $i \to \infty$ the system with the $N-1$ smaller indexes reaches steady state in k_{r+1}. In fact, assuming $k'_r = 0$, and denoting by $\mathbf{X}_k^{(N-1)}$ the first $N-1$ backoff indexes at time instant $k \geq 0$, and by $f_k = f(\mathbf{X}_k^{(N-1)}|\mathbf{x}_0)$ any probability measure on $\mathbf{X}_k^{(N-1)}$, conditional to the initial value $\mathbf{x}_0$, due to the ergodicity of $\{\mathbf{X}_k^{(N-1)}\}$, for all $\delta > 0$ and $\mathbf{x}_0$, there exists a time instant $L(\delta, \mathbf{x}_0) < \infty$ such that

$$|f^\star - f_L| < \delta, \tag{48}$$

where $f^\star = f(\mathbf{X}_\infty^{(N-1)})$ is the stationary value of the probability measure. Since we are dealing with the limit $i \to \infty$, index i can always be chosen in such a way that $k_{r+1} - k'_r$, whose average is b^{i+i_0}, is, with probability one, greater than any $L(\delta, \mathbf{x}_0)$. This means that the probability measure $f(\mathbf{X}_{k_{r+1}}^{(N-1)})$ has become independent of k_{r+1}. By applying this argument to the product inside expectation (10), the expectation itself splits in the product of expectations, yielding (11).

References

[1] L. Barletta, F. Borgonovo, and I. Filippini, "The S-Aloha capacity: Beyond the e^{-1} myth," in *Proc. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, Apr. 2016, pp. 1–9.

[2] L. Barletta, F. Borgonovo, and I. Filippini, "The access delay of Aloha with exponential back-off strategies," in *Proc. Medit. Ad Hoc Netw. Workshop (Med-Hoc-Net)*, Jun. 2016, pp. 1–6.

[3] N. Abramson, "The ALOHA SYSTEM: Another alternative for computer communications," in *Proc. Fall Joint Comput. Conf.*, vol. 37. Nov. 1970, pp. 281–285.

[4] L. G. Roberts, "ALOHA packet system with and without slots and capture," *SIGCOMM Comput. Commun. Rev.*, vol. 5, no. 2, pp. 28–42, 1975.

[5] L. Kleinrock and S. S. Lam, "Packet-switching in a slotted satellite channel," in *Proc. Jun. 4–8, 1973, Nat. Comput. Conf. Expo. (AFIPS)*, 1973, pp. 703–710.

[6] L. Kleinrock and S. Lam, "Packet switching in a multiaccess broadcast channel: Performance evaluation," *IEEE Trans. Commun.*, vol. COM-23, no. 4, pp. 410–423, Apr. 1975.

[7] G. Fayolle, E. Gelenbe, and J. Labetoulle, "Stability and optimal control of the packet switching broadcast channel," *J. ACM*, vol. 24, no. 3, pp. 375–386, Jul. 1977.

[8] S. Lam and L. Kleinrock, "Packet switching in a multiaccess broadcast channel: Dynamic control procedures," *IEEE Trans. Commun.*, vol. COM-23, no. 9, pp. 891–904, Sep. 1975.

[9] B. Hajek and T. van Loon, "Decentralized dynamic control of a multiaccess broadcast channel," *IEEE Trans. Autom. Control*, vol. AC-27, no. 3, pp. 559–569, Jun. 1982.

[10] R. Rivest, "Network control by Bayesian broadcast," *IEEE Trans. Inf. Theory*, vol. IT-33, no. 3, pp. 323–328, May 1987.

[11] L. P. Clare, "Control procedures for slotted Aloha systems that achieve stability," in *Proc. ACM SIGCOMM Conf. Commun. Archit. Protocols*, 1986, pp. 302–309.

[12] *Information Technology—Radio Frequency Identification for Item Management—Part 6: Parameters for Air Interface Communications at 860 MHz to 960 MHz*, ISO/IEC Standard 18000-6:2010, International Organization for Standardization, 2004.

[13] *Class 1 Generation 2 UHF Air Interface Protocol Standard*, Lawrenceville, NJ, USA, EPCglobal, 2005.

[14] L. Barletta, F. Borgonovo, and M. Cesana, "A formal proof of the optimal frame setting for dynamic-frame Aloha with known population size," *IEEE Trans. Inf. Theory*, vol. 60, no. 11, pp. 7221–7230, Nov. 2014.

[15] D. J. Aldous, "Ultimate instability of exponential back-off protocol for acknowledgment-based transmission control of random access communication channels," *IEEE Trans. Inf. Theory*, vol. IT-33, no. 2, pp. 219–223, Mar. 1987.

[16] J. Goodman, A. G. Greenberg, N. Madras, and P. March, "Stability of binary exponential backoff," *J. ACM*, vol. 35, no. 3, pp. 579–602, 1988.

[17] H. Al-Ammal, L. A. Goldberg, and P. MacKenzie, "An improved stability bound for binary exponential backoff," *Theory Comput. Syst.*, vol. 30, no. 3, pp. 229–244, 2001.

[18] J. Hastad, T. Leighton, and B. Rogoff, "Analysis of backoff protocols for multiple access channels," *SIAM J. Comput.*, vol. 25, no. 4, pp. 740–744, 1996.

[19] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 3, pp. 535–547, Mar. 2000.

[20] B.-J. Kwak, N.-O. Song, and L. E. Miller, "Performance analysis of exponential backoff," *IEEE/ACM Trans. Netw.*, vol. 13, no. 2, pp. 343–355, Apr. 2005.

[21] A. Kumar, E. Altman, D. Miorandi, and M. Goyal, "New insights from a fixed-point analysis of single cell IEEE 802.11 WLANs," *IEEE/ACM Trans. Netw.*, vol. 15, no. 3, pp. 588–601, Jun. 2007.

[22] J. W. Cho, J. Y. L. Boudec, and Y. Jiang, "On the asymptotic validity of the decoupling assumption for analyzing 802.11 MAC protocol," *IEEE Trans. Inf. Theory*, vol. 58, no. 11, pp. 6879–6893, Nov. 2012.

[23] C. Bordenave, D. McDonald, and A. Proutiere. (Nov. 2013). "A particle system in interaction with a rapidly varying environment: Mean field limits and applications." [Online]. Available: https://arxiv.org/abs/math/0701363

[24] L. Dai and X. Sun, "A unified analysis of IEEE 802.11 DCF networks: Stability, throughput, and delay," *IEEE Trans. Mobile Comput.*, vol. 12, no. 8, pp. 1558–1572, Aug. 2013.

[25] G. Sharma, A. Ganesh, and P. Key, "Performance analysis of contention based medium access control protocols," *IEEE Trans. Inf. Theory*, vol. 55, no. 4, pp. 1665–1682, Apr. 2009.

[26] A. Carleial and M. Hellman, "Bistable behavior of Aloha-type systems," *IEEE Trans. Commun.*, vol. COM-23, no. 4, pp. 401–410, Apr. 1975.

[27] L. Barletta and F. Borgonovo, "The stability of exponential backoff protocols for slotted-Aloha with saturated queues," in *Proc. IEEE Inf. Theory Workshop (ITW)*, Nov. 2017, pp. 1–5.
[28] J. D. C. Little, "A proof for the queuing formula: $L = \lambda$W," *Oper. Res.*, vol. 9, no. 3, pp. 383–387, 1961.
[29] A. Sen and N. Balakrishnan, "Convolution of geometrics and a reliability problem," *Statist. Probab. Lett.*, vol. 43, no. 4, pp. 421–426, 1999.
[30] T. Sakurai and H. L. Vu, "MAC access delay of IEEE 802.11 DCF," *IEEE Trans. Wireless Commun.*, vol. 6, no. 5, pp. 1702–1710, May 2007.

Luca Barletta (M'13) received the M.S. degree *(cum laude)* in telecommunications engineering and the Ph.D. degree in information engineering from the Politecnico di Milano, Milan, Italy, in 2007 and 2010, respectively.

From 2012 to 2014, he was a Post-Doctoral Researcher with the Institute for Advanced Study, Technische Universitaet Muenchen, Munich, Germany, and from 2014 to 2015, he was a Senior Researcher with the Institute for Communications Engineering. From 2011 to 2012, he was a Post-Doctoral Researcher with the Politecnico di Milano and a Visiting Researcher with Bell Labs, Alcatel-Lucent, Holmdel, NJ, USA, in 2012. He is currently an Assistant Professor with the Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano.

His research interests include digital communications and information theory, with emphasis on fiber-optic and wireless communications, and analysis of random access protocols.

Flaminio Borgonovo (M'90) received the Laurea degree in electronic engineering from the Politecnico di Milano, Milan, Italy, in 1971.

From 1973 to 1979, he was with the Italian National Research Council, where he was involved in the field of computer communications. In 1979, he became an Associate Professor with the Electronics Department, Politecnico di Milano, and after a period of working with the Università di Catania in 1990, he has been a Full Professor of electrical communications with the Politecnico di Milano. He has coauthored different national and international patents. In his research activity, he has investigated the multiple access problem in different environments, such as satellite, local area networks, metropolitan area networks, cellular networks, sensor networks, and more recently, in vehicular networks and radio frequency identification. In his professional activity, he has served as a consultant for the planning and the deployment of some private telecommunication networks for both data and voice traffic.

Ilario Filippini (S'06–M'10–SM'16) received the B.S., M.S., and Ph.D. degrees in electrical engineering from the Politecnico di Milan, in 2003, 2005, and 2009, respectively. He is currently an Assistant Professor with the Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano. His research interests include planning and optimization of wired and wireless networks, traffic management in software defined networks, and performance evaluation and resource management in wireless access networks. He serves as an Editor of *Wireless Communications and Mobile Computing* and *Computer Networks*.