Delay Analysis of Aloha Network

Soung Chang Liew, Ying Jun (Angela) Zhang, Da Rui Chen
Department of Information Engineering, The Chinese University of Hong Kong
Email:{soung, yjzhang, drchen}@ie.cuhk.edu.hk

***Abstract*—This paper provides a queueing analysis for the slotted Aloha network. We assume the use of an exponential backoff protocol. Most prior work on slotted Aloha focuses on the analysis of its saturation throughput. Good saturation throughput, however, does not automatically translate to good delay performance for the end users. For example, it is well-known that the maximum possible throughput of slotted Aloha with a large number of nodes is $e^{-1} = 0.3679$. Prior work showed that binary backoff factor of $r = 2$ can achieve a saturation throughput of 0.3466, which is very close to the e^{-1}. However, this paper shows that if mean queuing delay is to be bounded, then the offered load must be below 0.2158, a drastic 41% drop from e^{-1}. Fortunately, setting $r = 1.3757$ allows us to achieve bounded-mean-delay throughput of 0.3545, less than 4% lower than e^{-1}. A general conclusion is that the backoff factor r may significantly affect the queuing delay performance. Our analysis provides a framework to set system parameters properly.**

***Index Terms*—Multiple-access protocols, Aloha, wireless LAN.**

I. INTRODUCTION

The Aloha network has been studied extensively since the pioneer work by Abramson [1]. Prior work on the Aloha network has primarily focused on its overall system throughput. Many investigations [2][3][4] considered the access delay (i.e., service time incurred by a packet at the head-of-line (HOL) of its queue). In a saturated system, the reciprocal of the expected access delay is the saturation throughput. By contrast, the focus of this paper is the overall queuing delay (i.e., waiting time plus service time).

To achieve good system throughput, the transmission probabilities of the nodes must be adjusted dynamically according to the contention intensity in the network. An exponential backoff protocol can serve this purpose rather effectively [2].

Good system throughput, however, does not automatically translate to acceptable performance from the end user perspective. For example, if a real-time application such as a voice call is running on top the Aloha network, delay performance is important. This paper is devoted to the study of delay in a slotted Aloha network operated with an exponential backoff protocol. Within this context, this paper has three main contributions:

1. We establish a framework for queuing analysis in the Aloha network.
2. Based on the analytical framework, we derive the dependency of delay on the system parameters.
3. We discover that the queue-length distribution under exponential backoff is fractal.

With respect to contribution 1, we find that the saturation throughput is not a sound measure of performance if we care about delay. In particular, to achieve good delay performance, the offered load must be below another quantity called "safe-bounded-mean-delay (SBMD) throughput", which can be substantially lower than the saturation throughput.

With respect to contribution 2, we show that delay can be very sensitive to the system parameters. For example, it is well known that the maximum saturation throughput of a large slotted Aloha network is $e^{-1} = 0.3679$. An exponential backoff factor of 2 was commonly assumed in many prior studies [3][5], and it can achieve a saturation throughput of 0.3466 [2]. Thus, backoff factor 2 is quite satisfactory when it comes to saturation throughput performance. However, if we desire bounded mean delay, using backoff factor 2 requires us to limit the system offered load to below 0.2158. This corresponds to a drastic penalty of 41% with respect to e^{-1}. Fortunately, if we use a backoff factor of 1.3757 instead, the sustainable offered load can reach 0.3545, very close to e^{-1}.

With respect to contribution 3, we were prompted to investigate the possibility that the queue length is power-law distributed [6] (i.e., fractal) by our observation that some of our simulation results do not converge to the analytical results, although exactly the same system was being simulated. Additional simulations confirmed the power-law distribution, which suggests that the network is subject to the vagaries of rare events that can have high impact on the system statistics.

II. SYSTEM MODEL

A. Real System

We consider a slotted Aloha network with N nodes. We assume packet arrivals to each node are Poisson with rate $\lambda_o = S_o / N$, where S_o is the offered load to the overall system, where S_o is the offered load to the overall system, and thus S_o/N is the offered load to a single queue. Each node has a queue to hold its backlog packets. When a fresh packet enters the HOL of its queue, it transmits with probability $1/r_0$ in each time slot, where $r_0 \geq 1$. When more than one node transmits a packet in a time slot, a collision occurs and the packets are corrupted. A collided packet will be retransmitted in a future time slot. Each time a HOL packet suffers a collision, the transmission probability in the future is divided by the backoff factor $r > 1$. Thus, a HOL packet that has suffered i prior collisions will be transmitted in a future time slot with

This work was supported by the Competitive Earmarked Research Grant (Project Number 414305) established under the University Grant Committee of the Hong Kong Special Administrative Region, China.

978-1-4244-2324-8/08/$25.00 © 2008 IEEE.
This full text paper was peer reviewed at the direction of IEEE Communications Society subject matter experts for publication in the IEEE "GLOBECOM" 2008 proceedings.

probability $1/(r_0 r^i)$. We refer to i as the backoff stage of a node. A HOL packet will be retransmitted until it is successfully cleared without a collision, at which point the next-in-line packet, if any, will proceed to the HOL.

With our model, the "local state" of a queue can be described by a duple (Q,B), where Q is the number of backlog packets in the queue, including the HOL packet; and B is the backoff stage of the HOL packet. The "global state" of the overall system consists of the aggregate local states of all N queues. One can in principle construct a multi-dimensional Markov chain to analyze the system. However, the analysis for even modest-size N is prohibitively complex and not much insight can be gained from this brute-force analysis. Detailed and exact results, for example, are only available for the 2-node case [5].

B. *Proxy System*

For large N, an approximation technique that has been often used in saturation analysis is to replace the actual system model with a "proxy model" (e.g., [2]). This paper adopts the same approximations. The proxy system makes two approximations: (i) the probability of collision p_c experienced by a node is independent of its local state; (ii) as far as a local node is concerned, each of the other nodes transmits with a probability p_t in a given time slot. Certainly these approximations are only valid under large N when each local node only has a small effect on the overall system. Simulations of the actual system, referred to as the "real system" in this paper, can be used to check against the accuracy of the proxy-system analysis. This paper will show such verification results.

III. Queueing Delay Analysis

We decompose the analysis of the proxy system into three steps. The first step is a "global analysis" linking p_c and p_t. The second step is a "local analysis" focusing on the local dynamic of a node assuming a fixed p_c. The third step is a "coupling analysis" which combines the results from the first two steps.

A. *Global Analysis*

Consider the proxy system from the global viewpoint. Given that there are $(N-1)$ other nodes, each transmitting with probability p_t, the probability of collision when a node transmits is

$$\begin{aligned} p_c &= 1-(1-p_t)^{N-1} \\ &\to 1-e^{-Np_t} \quad \text{as} \quad N\to\infty \end{aligned} \tag{1}$$

Let $G_o = Np_t$ be the transmission attempt rate of the overall system when the offered load is S_o. For a non-saturated system under equilibrium, the output rate (i.e., throughput) is equal to the input rate (i.e., offered load). Thus, S_o is also the throughput (i.e, the average number of packets successfully transmitted per time slot). Therefore, $p_c = 1-S_o/G_o$. Applying the above relationships on (1) yields

$$\begin{aligned} S_o &= G_o\left(1-\frac{G_o}{N}\right)^{N-1} \\ &= G_o e^{-G_o} \qquad \text{as } N\to\infty \end{aligned} \tag{2}$$

B. *Local Analysis*

For Poisson arrival, a packet of a local queue generally arrives between the boundaries of two adjacent time slots. If it arrives to an empty queue, it must wait until the beginning of the next time slot before it can contend for transmission. Conceptually, it does not enter the HOL until the next time slot. It turns out that this local queue specification fits under the M/G/1 multiple-vacation queue model [7]. In the multiple-vacation queue model, the server leaves for a vacation when the queue becomes empty. Upon returning from a vacation, if the queue remains empty, the server immediately departs for another vacation. When a packet arrives to an empty queue in the Aloha network, the time until the beginning of the next time slot is part of the vacation time taken by the server. For slotted Aloha, the vacation time is fixed and equal to one slot time. The access delay incurred by a packet at the HOL corresponds to the service time of the M/G/1 vacation queue.

For notation purposes, in the following, $F(z)=\sum_{i=0}^{\infty}\Pr[F=i]z^i$ denotes the z-transform of a discrete non-negative random variable F, and $G^*(s)=\int_0^\infty f_G(x)e^{-sx}dx$ denotes the Laplace transform of a continuous non-negative random variable G. The M/G/1 vacation queue has the following solution:

$$\begin{aligned} Q(z) &= \frac{(1-\lambda_o\bar{X})}{\lambda\bar{V}}\cdot\frac{X^*(\lambda_o(1-z))[V^*(\lambda_o(1-z)-1]}{z-X^*(\lambda_o(1-z))} \\ D^*(s) &= Q(1-s/\lambda_o) = \frac{(1-\lambda_o\bar{X})}{\bar{V}}\cdot\frac{X^*(s)[V^*(s)-1]}{\lambda-s-\lambda_o X^*(s)} \end{aligned} \tag{3}$$

where

$Q =$ number of packets in the queue including the HOL packet
$D =$ queueing delay including the service time
$X =$ service time of a packet
$V =$ vacation time taken by the server when the queue is empty

Expressions (3) are generic expressions relating Q and D to X and V. To use (3), we need to derive the distributions of X and V specific to our system. For slotted Aloha, each vacation lasts exactly one time slot, so that

$$V^*(s) = e^{-s} \tag{4}$$

Recall that an approximation in the proxy system is a constant p_c independent of the local state. We now derive X in terms of p_c. Mathematically, the Laplace transform $X^*(s)$ in (3) is related to the z-transform $X(z)$ by

$$X^*(s) = X(e^{-s}) \tag{5}$$

To derive $X(z)$, let C be the number of collisions experienced by a HOL packet before it is successfully transmitted. By conditional-probability argument, we have

$$X(z)=\sum_{k=0}^{\infty}X(z\,|\,C=k)(1-p_c)p_c^{\,k},$$
$$X(z\,|\,C=k)=X_0(z)X_1(z)\cdots X_k(z)$$
where
$$X_j(z)= z\text{-transform of the time between the } j^{\text{th}} \text{ and } (j+1)^{\text{th}} \text{ transmissions} = \sum_{i=1}^{\infty}\left(\frac{1}{r_0r^j}\right)\left(1-\frac{1}{r_0r^j}\right)^{i-1}z^i=\frac{z}{r_0r^j-(r_0r^j-1)z} \tag{6}$$

Thus,

$$X(z)=\sum_{k=0}^{\infty}(1-p_c)p_c^{\,k}\prod_{j=0}^{k}\frac{z}{r_0r^j-(r_0r^j-1)z} \tag{7}$$

Eqns. (3), (4), (5), and (7) allow us to derive moments of D in terms of $r_0,\ r,\ p_c$. For the first moment $E[D]$, after some equation crunching, we can get

$$\begin{aligned}E[D]&=-D^{*}{}'(0)=X'(1)+\frac{\lambda_oX''(1)}{2(1-\lambda_oX'(1))}+\frac{\lambda_oX'(1)}{2(1-\lambda_oX'(1))}+\frac{\overline{V^2}}{2\overline{V}}\\&=\frac{r_0}{1-p_cr}+\frac{\lambda_or_0(p_cr^2+r_0-1)}{(1-p_cr^2)(1-p_cr-\lambda_or_0)}+\frac{\lambda_or_0}{2(1-p_cr-\lambda_or_0)}+\frac{1}{2}\\&=\frac{r_0}{1-p_cr}+\frac{\lambda_or_0(p_cr^2+2r_0-1)}{2(1-p_cr^2)(1-p_cr-\lambda_or_0)}+\frac{1}{2}\end{aligned} \tag{8}$$

We note that independently [9] obtained $X'(1)$ and $X''(1)$ for the $r_0=1$ case.

Bounded Mean-Delay Conditions

Let us next consider the implications of (8). From (8), convergence of $E[D]$ requires $p_cr<1,\ p_cr+\lambda_or_0<1$ and $p_cr^2<1$, but the first inequality is satisfied if the second is and can be eliminated. Thus, we have the following conditions for bounded $E[D]$:

$$p_cr+\lambda_or_0=p_cr+\frac{r_0S_o}{N}<1 \quad\text{and}\quad p_cr^2<1 \tag{9}$$

Note that at equilibrium, the mean service time is $X'(1)=r_0/(1-p_cr)$. Applying Little's law and the physical requirement that the average HOL occupancy is less than 1 when the system is non-saturated, we have $\lambda_o r_0/(1-p_cr)<1$, which is the same as the first inequality in (9). Thus, the first inequality is also the condition for non-saturation. The second inequality in (9), on the other hand, arises from the requirement to bound $Var(X)=X''(1)+X'(1)$ in (8).

The analysis thus far assumes steady-state equilibrium can be achieved. For a queuing system, steady state can be achieved if and only if $\Pr[Q=0]>0$ (see [8]). Since $\Pr[Q=0]>0$ means the queue is not saturated, the first inequality of (9) is also the necessary and sufficient condition for steady state operation

C. Coupling Analysis

The local analysis leaves us with (8), where mean delay is expressed in terms of r, r_0 and p_c. We need to use the result from the global analysis to remove the dependency on p_c. This section contains a condensed coupling analysis. The reader is referred to [10] for a more rigorous and detailed treatment.

Numerically, $E[D]$ can be obtained as follows. For a given S_o, we compute G_o from $S_o=G_o(1-G_o/N)^{N-1}$. A subtlety is that there are two possible solutions for G_o. In [10], we argue that the smaller G_o is the correct solution. Using the smaller G_o, we then substitute $p_c=1-S_o/G_o$ and $\lambda_o=S_o/N$ into (8) to find $E[D]$.

We next consider the limit on S_o to ensure bounded mean delay.

$N\to\infty$ Case

We first consider the asymptotic $N\to\infty$ case. Condition (9) implies that there is an upper bound that S_o should not exceed if $E[D]$ is to be finite. For $N\to\infty$, $\lambda_or_0=S_or_0/N\to0$ and the first inequality in (9) becomes $p_cr<1$. This inequality is satisfied if the second inequality $p_cr^2<1$ is satisfied, since $r>1$. Thus, we only need to look at the second inequality of (9). Let us examine the "boundary case" where $p_cr^2=1$. Applying (2) on $p_cr^2=(1-S_o/G_o)r^2=1$, we can get $S_o=\frac{r^2-1}{r^2}\ln\left(\frac{r^2}{r^2-1}\right)$. We shall refer to this quantity as the "boundary-bounded-mean-delay" throughput, denoted by

$$S_{BBMD}=\frac{r^2-1}{r^2}\ln\left(\frac{r^2}{r^2-1}\right) \tag{10}$$

The corresponding attempt rate is

$$G_{BBMD}=\ln\left(\frac{r^2}{r^2-1}\right) \tag{11}$$

The point (G_{BBMD},S_{BBMD}) is a point on the $S_o=G_oe^{-G_o}$ curve. Depending on r, it may lie to the left or right of the peak of the curve. In general, w require $S_o<S_{BBMD}$ for bounded mean delay operation, as elaborated below.

First, let us consider the case where r is such that (G_{BBMD},S_{BBMD}) lies to the left of the peak on the $S_o=G_oe^{-G_o}$ curve. We note that $dp_c/dG_o=e^{-G_o}>0$ so that **p_c is an increasing function of G_o on the $S_o=G_oe^{-G_o}$ curve.** Since (G_{BBMD},S_{BBMD}) is the point at which $p_cr^2=1$, we require the operating point (G_o,S_o) to be to the left of (G_{BBMD},S_{BBMD}) in order that $p_cr^2<1$, and this in turns means that $S_o<S_{BBMD}$.

Next, consider the case where r is such that (G_{BBMD}, S_{BBMD}) is to the right of the peak of $S_o = G_o e^{-G_o}$. It is now possible to find an $S_o > S_{BBMD}$ such that $p_c r^2 = (1 - S_o / G_o) r^2 < 1$ (e.g., see Fig. 1, where the p_c at (G_o, S_o) is smaller than the p_c at (G_{BBMD}, S_{BBMD})). Thus, it may appear at first glance that it is "safe" to operate the system with an offered load larger than S_{BBMD} for the small r case. In the following paragraph, we argue that this is not so.

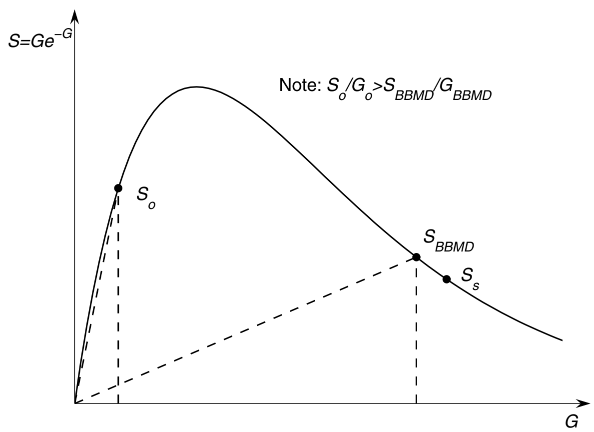


Fig. 1. An example where offered load S_o can be larger than S_{BBMD} while fulfilling $p_c r^2 = (1 - S_0 / G_0) r^2 < (1 - S_{BBMD} / G_{BBMD}) r^2 < 1$.

Recall that our local analysis in Section III.B presumes that the system in not saturated, and that saturation will cause unbounded $E[D]$. It can be shown that the saturation throughput S_s and the corresponding system transmission attempt rate G_s are given by (see Section 2 of [10]):

$$G_s = \ln\left(\frac{r}{r-1}\right)$$
$$S_s = \left(\frac{r-1}{r}\right)\ln\left(\frac{r}{r-1}\right) \quad (12)$$

Now, $G_s = \ln\frac{r}{r-1} > \ln\frac{r^2}{(r-1)(r+1)} = G_{BBMD}$. Thus, given an r, (G_s, S_s) is always to the right of (G_{BBMD}, S_{BBMD}) on the $S_o = G_o e^{-G_o}$ curve (see Fig. 1 for illustration). If we load the system with an $S_o > S_s$, there is the danger that the system will run into saturation, at which point $E[D]$ will go to infinity because the saturation throughput S_s cannot keep up with the input rate S_o. When that happens, the first inequality of (9) will not be satisfied. In a simulation experiment performed by us, for a situation such as that depicted in Fig. 1, we intentionally caused the system to go into saturation with a sudden increase in the offered load, and then decreased the offered load back to S_o. The simulation results show that $E[D]$ then becomes unbounded. In other words, such an S_o which is larger than S_s is not a "safe" offered load, and it is obtained with an a priori assumption of equilibrium and non-saturation. If the system is already in saturation, $E[D]$ is unbounded for such an S_o. On the other hand, in the simulation experiment, if we decreased the offered load further to below S_s, then the system did clear up and $E[D]$ became bounded. This is because S_s is higher than the input rate and sooner or later, the backlog will be cleared. Thus, S_o below S_s is safe.

Based on the above argument, we therefore define the "safe" bounded-mean-delay throughput as

$$\begin{aligned} S_{SBMD}(r) &= \min\left[S_{BBMD}(r),\ S_s(r)\right] \\ &= \min\left[\frac{r^2-1}{r^2}\ln\left(\frac{r^2}{r^2-1}\right),\ \frac{r-1}{r}\ln\left(\frac{r}{r-1}\right)\right] \end{aligned} \quad (13)$$

Non-asymptotic Case

We now consider the finite-N case. For the $N \to \infty$ case, the second inequality of (9) is always more stringent than the first inequality. For the finite-N case, depending on the values of N, r, and r_0, the first inequality in (9) could be more stringent than the second inequality. When the first inequality is more stringent, then S_{BBMD} should be found by setting the first inequality to equality. Recall that this is the saturation point so that $S_{BBMD} = S_s$. When the second inequality is more stringent, using $p_c r^2 = (1 - S_o / G_o) r^2 = 1$ to eliminate G_o in the first line of (2) gives $S_{BBMD} = N(1 - 1/r^2)\left[1 - \left(1 - 1/r^2\right)^{1/(N-1)}\right]$. In summary, for the non-asymptotic case,

$$S_{SBMD}(r, r_0, N) = \min\left[S_{BBMD}(r, r_0, N),\ S_s(r, r_0, N)\right] \quad (14)$$
$$= \begin{cases} S_s(r, r_0, N) & \text{or} \\ \min\left\{\dfrac{N(r^2-1)}{r^2}\left[1 - \left(\dfrac{r^2-1}{r^2}\right)^{1/(N-1)}\right],\ S_s(r, r_0, N)\right\} \end{cases}$$

where the first (second) case applies when the first (second) inequality in (9) is more stringent. Also, there is no closed-form solution to $S_s(r, r_0, N)$ and it must be found numerically from the following expression (see [10] for details):

$$\left(1 + \frac{r_0}{r-1}\cdot\frac{S_s}{N}\right)^N = \left(\frac{r}{r-1}\right)\left(1 + \frac{r_0 - r}{r-1}\cdot\frac{S_s}{N}\right)^{N-1} \quad (15)$$

IV. Numerical results and Effects of Backoff Factor

A. Maximum SBMD Throughput

Let us now examine how S_{SBMD} varies as r is varied. We focus on the asymptotic $N \to \infty$ case here. Similar argument applies to the non-asymptotic case. Fig. 2 plots $S_{SBMD}(r)$, $S_{BBMD}(r)$, and $S_s(r)$ versus r according to (13).

For $r > 1.3757$, $S_{BBMD}(r) < S_s(r)$; and for $r \le 1.3757$, $S_{BBMD}(r) \ge S_s(r)$. Specifically, the r which maximizes $S_{SBMD}(r)$ is $r^*_{SBMD} = 1.3757$, which is obtained by setting $S_{BBMD}(r) = S_s(r)$:

$$\frac{{r^*_{SBMD}}^2 - 1}{{r^*_{SBMD}}^2}\ln\left(\frac{{r^*_{SBMD}}^2}{{r^*_{SBMD}}^2 - 1}\right) = \frac{r^*_{SBMD} - 1}{r^*_{SBMD}}\ln\left(\frac{r^*_{SBMD}}{r^*_{SBMD} - 1}\right) \quad (16)$$

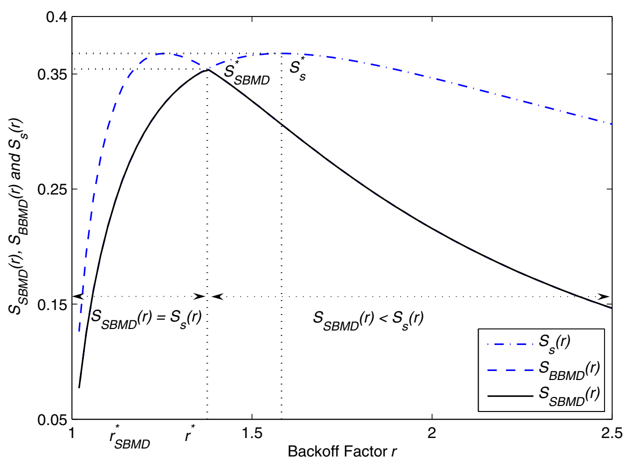


Fig. 2. $S_{SBMD}(r)$, $S_{BBMD}(r)$, and $S_s(r)$ versus r for the case of $N \to \infty$.

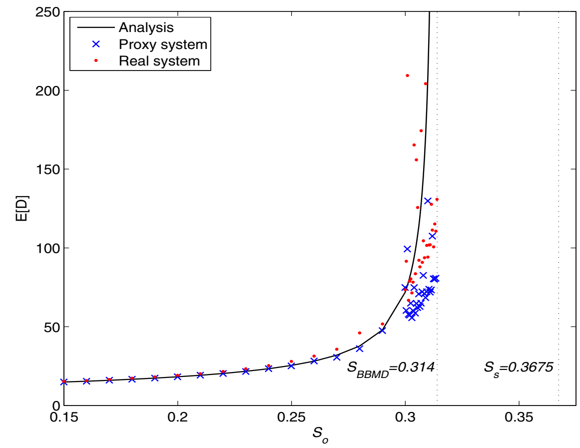
Note that $r^*_{SBMD} \ne r^*_s = e/(e-1)$, where the r^*_s is the value of r value that maximizes the saturation throughput $S_s(r)$. The maximum saturation throughput is $S^*_s = S_s(r^*_s) = e^{-1} = 0.3679$. However, $S_{SBMD}(r^*_s) = 0.3063$, which is 17% below S^*_s. That is, if we set $r = r^*_s$, S_o must be at least 17% below S^*_s to ensure bounded delay.

The binary backoff factor of $r = 2$ is assumed in the majority of prior work, and in many practical multiple-access networks such as the Ethernet and WiFi. For slotted Aloha, the corresponding saturation throughput $S_s(2) = 0.3466$ is reasonably close to $S^*_s = 0.3679$, and one could hardly raise objection to adopting $r = 2$ on the basis of saturation throughput. However, if bounded mean delay is desired, we have $S_{SBMD}(2) = 0.2158$. That is, there is a drastic 41% penalty with respect to S^*_s. Therefore, $r = 2$ is a bad choice

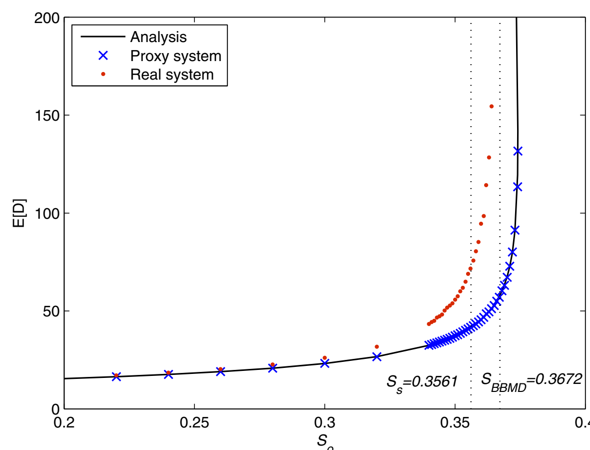
Fortunately, the maximum SBMD throughput, obtained by setting $r = r^*_{SBMD} = 1.3757$, is rather close to S^*_s. Specifically, $S^*_{SBMD} = S_{SBMD}(r^*_{SBMD}) = 0.3545$. The penalty with respect to S^*_s is only less than 4%. Overall, we conclude that using the proper r is important to ensuring a good throughput under the bounded-delay requirement, perhaps more so than when saturation throughput is the only concern. This can be seen from Fig. 2, which shows that $S_{SBMD}(r)$ rises and falls much more sharply with r than $S_s(r)$ does.

B. Mean Delay versus Offered Load

Fig. 3 plots $E[D]$ versus S_o for $N = 30$. In Fig. 3(a), $(r_0, r, N) = (10,\ 1.582,\ 30)$. For this case, $S_{BBMD} = 0.3140 < 0.3765 = S_s$. Thus, S_{SBMD} is limited by S_{BBMD} rather than the saturation throughput S_s. The solid line corresponds to the result obtained from analysis. The cross points are simulation results of the proxy system in which the dynamic of a single node is simulated with fixed $p_c = (G_o - S_o)/G_o$. The dotted points correspond to simulation results of the real system with 30 queues. The results are consistent in that for offered load S_o near S_{BBMD}, $E[D]$ begins to build up quickly.



(a)



(b)

Fig. 3. $E[D]$ versus S_o for (a) $(r_0, r, N) = (10,\ 1.582,\ 30)$; (b) $(r_0, r, N) = (10,\ 1.200,\ 30)$.

An interesting observation is that near S_{BBMD}, the simulated proxy-system $E[D]$ does not quite converge exactly to the analytical $E[D]$ near S_{BBMD}, although the simulation experiment simulates exactly the same proxy system as that in the analysis. So, this non-convergence is not due to the proxy system not approximating the real system well. In fact, for the same S_o near S_{BBMD}, different simulation runs will produce rather different $E[D]$ even if we let each run lasts a long time.

Interested readers are referred to [10] for a detailed discussion on this non-convergence phenomenon.

In Fig. 3(b), $(r_0, r, N) = (10,\ 1.200,\ 30)$. For this case, $S_{BBMD} = 0.3762 \ > 0.3561 = S_s$. Thus, S_{SBMD} is limited by S_s rather than S_{BBMD}. With respect to the S versus G curve (such as that in Fig. 1 but with finite N), this is the case where both S_{BBMD} and S_s are to the right of the peak of the curve. As discussed in Section III.C, it is possible to load the system with S_o above S_{BBMD} and yet satisfy the convergence condition as dictated by (9) (see Fig. 1). However, as explained, such an S_o not "safe" because the system may evolve into a saturated state. Thus, although Fig. 3(b) shows finite $E[D]$ for $S_o > S_s$, operating at such offered load for an extended period of time is not advisable.

C. *Fractal Distribution of Queue Length*

In Fig. 3(a), when S_o approaches S_{BBMD}, the simulated $E[D]$ shows large fluctuations no matter how long we run the simulation. This led us to suspect that the Aloha system is fractal in nature. In Fig. 4, we plot the simulation results of $\ln \Pr[Q \geq q]$ versus $\ln(q)$ under the same parameter setting as in Fig. 3(a). The linear relationship is indicative of power-law distributions [6]. When $S_o = 0.31$, the gradient of the curves is around -1.3, indicating that the non-convergence of $E[Q^m]$ for $m > 1.5$ [6] (i.e., variance of Q does not converge). For comparison, we also plot the curve for an M/M/1 queue, which does not exhibit power-law distribution.

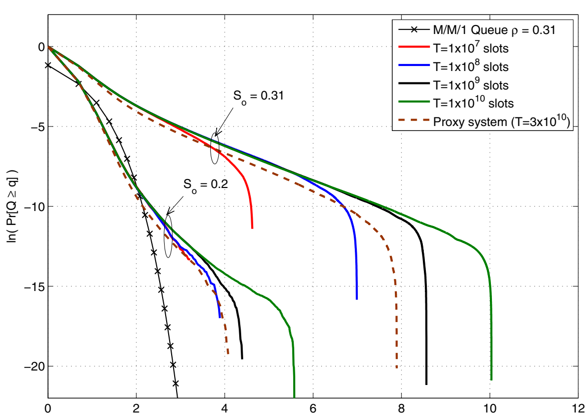


Fig. 4. Power-law queue-length distributions of real and proxy systems for $(r_0, r, N) = (10,\ 1.582,\ 30)$.

We intentionally ran a number of experiments with increasing simulated time slots. The role of "rare events" is clear. Note that for the same S_o, each linear curve plunges vertically at some $\ln(q)$. The point where it plunges depends on the number of simulate time slots. In particular, we may not see large queue length if the simulated time horizon is not long enough, but as the simulated time horizon increases, rare events come into effect. This implies that for the Aloha system, simulation results could be "misleading" if one does not frame them in the right context. With respect to Fig. 3(a), we may see many simulation runs producing lower $E[D]$ than the computed $E[D]$; however, a rare event may produce $E[D]$ much higher than the computed $E[D]$, changing the gathered statistics significantly.

V. Conclusion

We have presented a framework for the analysis of the queuing delay in the slotted Aloha system operating the exponential backoff protocol. The condition for bounded mean-delay operation has been established. Specifically, the system offered load must be below a "safe-bounded-mean-delay throughput", S_{SBMD}, which is distinct from the well-known saturation throughput, in order that bounded mean-delay can be achieved. For the case in which the number of nodes N is large, we have

$$S_{SBMD} = \min\left[\frac{r^2-1}{r^2}\ln\left(\frac{r^2}{r^2-1}\right),\ \frac{r-1}{r}\ln\left(\frac{r}{r-1}\right)\right] \tag{17}$$

where r is the backoff factor. A general conclusion from this work is that the setting of r is crucial as far as the delay.

Starvation can also occur in the Aloha system. When starvation occurs, some nodes are deprived of service for an extended period of time while other nodes hog the system. To limit the scope of this paper, we have not delved into the discussion of starvation here. Interested readers are referred to [10], in which we argue that the conditions leading to starvation and unbounded mean delay are one of the same, hence uniting these two notions.

Finally, a natural generalization of the results here are for carrier-sense multiple-access (CSMA) networks. A companion paper of ours [11] is an attempt in that direction.

References

[1] N. Abramson "The Aloha System – Another Alternative for Computer Communication," *Proc. Fall Joint Comput. Conf.*, AFIP Conference, vol. 44, pp. 281-285, 1970.

[2] B-J Kwak, N-O Song, L. E. Miller, "Performance Analysis of Exponential Backoff," *IEEE/ACM Trans. on Networking*, vol. 13., no. 2, pp. 343-353, Apr. 2005.

[3] Y. Yang, T-S. P. Yum, "Delay Distribution of Slotted Aloha and CSMA," *IEEE Trans on Comm.*, vol. 51, no. 11, pp. 1846-1857, Nov. 2003.

[4] F. A. Tobagi, "Distributions of Packet Delay and Interdeparture Time in Slotted Aloha and Carrier Sense Multiple Access," *JACM,* vol. 29, no. 4, pp. 907-927, Oct 1982.

[5] J. Goodman, A. G. Greenberg, N. Madras, P. March, "Stability of Binary Exponential Backoff," *JACM* vol. 35, no, 3, July 1988, pp. 579-602.

[6] http://en.wikipedia.org/wiki/Power_law

[7] B. T. Doshi, "Queueing Systems with Vacations – A Survey," *Queueing Systems, Theory and Application*, vol. 1, no. 1, 1986.

[8] L. Kleinrock, *Queueing Systems, Volume 1: Theory*, Wiley 1975.

[9] T. T. Lee and L. Dai, "A Statistical Theory of Wireless Networks – Part I. Queuing Analysis of Spatial Interferences," *Technical Report, The Chinese University of Hong Kong*.

[10] S. C. Liew, Y. J. Zhang, D. R. Chen, "Bounded Mean-Delay Throughput and Non-Starvation Conditions in Aloha Network," available at http://arxiv.org/pdf/0801.4054.pdf.

[11] Y. J. Zhang, S. C. Liew, D. R. Chen, "Delay Analysis for Wireless Local Area Networks with Multipacket Reception under Finite Load," *IEEE Globecom 2008.*