Analysis of the Stability and Performance of Exponential Backoff

Byung-Jae Kwak
bjkwak@antd.nist.gov

Nah-Oak Song
nsong@antd.nist.gov

Leonard E. Miller
lmiller@antd.nist.gov

***Abstract*— New analytical results are given for the stability and performance of the exponential backoff (EB) algorithm. Previous studies on the stability of the (binary) EB have produced contradictory results instead of a consensus: some proved instability, others showed stability under certain conditions. In these studies, simplified and/or modified models of the backoff algorithm were often used to make analysis more tractable. In this paper, care is taken to use a model for backoff that reflects the actual behavior of backoff algorithms. We show that EB is stable under a throughput definition of stability; the throughput of the network converges to a non-zero constant as the offered load N goes to infinity. We also obtain the analytical expressions for the saturation throughput and the medium access delay of a packet for a given number of nodes, N. The analysis considers the general case of EB with backoff factor r, where BEB is the special case with $r = 2$. The accuracy of the analysis is checked against simulation results.**

I. Introduction

Random access schemes for packet networks featuring distributed control require algorithms and protocols for resolving packet collisions that occur as the uncoordinated terminals contend for the channel. A widely used collision resolution protocol is the *binary exponential backoff* (BEB), forms of which are included in Ethernet [1] and Wireless LAN [2] standards. In this paper, we assess the stability of BEB and analyze its performance using a model that closely resembles practical network transmission schemes and therefore is useful for system planning and analysis.

A. Stability of Backoff Algorithms

Many papers study the stability of backoff algorithms, including BEB, in terms of their effect on network performance as the offered load increases. However, these studies have produced contradictory results instead of a consensus: some prove instability, others show stability under certain conditions. The mixed results are due to differences in the analytical models and the definitions of stability used in the analysis.

Simplified and/or modified models of the backoff algorithm are often used to make analysis more tractable. However, simplification or modification can lead to very different analytical results. For example, Aldous [3] proved that BEB is *unstable* for an infinite-node model (a simplified model) for any non-zero arrival rate, while Goodman et al. [4] showed using a modified finite-node model that BEB is *stable* for sufficiently small arrival rates. Also, while modification of the model can make the analysis much simpler, the analytical result may have limited relevance because it cannot be guaranteed that the modified model exhibits the same behavior as the actual algorithm.

The various definitions of stability used in the studies of backoff algorithms can be classified into two groups. One group of studies uses a definition based on throughput and the other uses delay to define stability. Under the throughput definition, the algorithm is stable if the throughput does not collapse as the offered load goes to infinity [3] or is an increasing function of the offered load [5]. Under the delay definition, the protocol is stable if the waiting time is bounded. Systems that are stable under the delay definition can be characterized by bounded backlog of packets in the queue, or recurrent property of Markov chain [6].

Most of the analytical and simulation studies on stability treat BEB in the context of a specific network medium access control (MAC) protocol such as Ethernet etc. [10], [11], [12], [13], [14]. However, the characteristics of this protocol seem to have as much or more effect on the network performance results than the intrinsic behavior of BEB. Some of the analytical works that focus on BEB itself are summarized as follows:

Kelly and MacPhee [7] prove that "for a general acknowledgment based random access scheme there exists a critical value $\nu_c \in [0, \infty]$, with the property that the number of packets successfully transmitted is finite with probability 0 or 1 according as $\nu < \nu_c$ or $\nu > \nu_c$," where ν is the arrival rate of the system. It is also shown that $\nu_c = 0$ for any scheme with slower than exponential backoff, and $\nu_c = \log 2$ for BEB. They use an infinite-node model with Poisson arrivals, assuming that no node ever has more than one packet arrive at it. This result proves that BEB is unstable for $\nu > \nu_c$, but leaves open the stability for $\nu < \nu_c$.

In [3], Aldous shows that, with infinite-node and Poisson arrival assumptions, BEB is *unstable* in the sense that $N(t)/t$ converges to zero as t goes to infinity for any non-zero arrival rate, where $N(t)$ is the number of the successful transmissions made during the time $[0, t]$. This result solves the open problem left in [7], but the model Aldous uses is slightly different from Kelly and MacPhee's model.

Goodman et al. prove in [4] that BEB is *stable* if the arrival rate of the system is sufficiently small in the sense that the backlog of packets awaiting transmission remains bounded in time. More specifically, they show that BEB is stable if the arrival rate is smaller than $\lambda^*(n)$, where $\lambda^*(n) \geq 1/n^{\alpha \log n}$ for some constant α and n is the number of nodes. They assume that each of the finite number of nodes n has a queue of infinite capacity.

In [8], Al-Ammal et al. give a tighter (greater) upper bound of the arrival rate than that given in [4] for the stability of BEB under the delay definition of stability. By using the same analytical model as in [4], they show that there is positive constant α such that, as long as n is sufficiently large and the system arrival rate is smaller than $1/\alpha n^{0.9}$ then the system is stable for the n-user system. The upper bound in [8] is further improved in [9], where it is proved that BEB is stable for arrival rate smaller than $1/\alpha n^{1-\eta}$, where $\eta < 0.25$. The main point of their work is that BEB is stable for an arrival rate that is inverse of a sublinear polynomial in n.

Finally, in [6], Håstad et al. show, using the same analytical model as in [4], that BEB is *unstable* whenever $\lambda_i \geq \lambda/n$ for $1 \leq i \leq n$ and $\lambda > 0.567 + 1/(4n-2)$, or when $\lambda > 0.5$ and n is sufficiently large under the delay definition of stability, where λ is the system arrival rate and λ_i is the arrival rate at node i.

In summary, these representative analyses indicate that BEB is unstable for an infinite-node model, and for a finite-node model it is stable if the system arrival rate is small enough but unstable if the arrival rate is too large. We note that they all assume slotted transmissions. While these analytical results are well established, because they are contradictory and do not represent the actual system, there remains doubt about the stability of BEB so that this question continues to be an open problem. As noted in [6] and [15], the infinite node model used in [7] and [3] is a mathematical abstraction with limited practical application, and except for [7], all of the studies cited actually analyze a modified version of BEB. For example, in BEB, after i consecutive packet transmission failures (collisions) a node selects for the next transmission *a single random slot* from the next 2^i slots with equal probability, while in the modified versions after i collisions a node is assumed to transmit in each slot with probability 2^{-i}. Clearly, it is easier to analyze such a modified version of BEB because of its memoryless nature, but it is not guaranteed that it has the same stability characteristics as BEB.

B. Approach of This Paper

In this paper, we analyze the stability of the original BEB algorithm by showing that the network throughput continues to be non-zero even when the number of nodes goes to infinity. The analysis considers the general case of exponential backoff (EB) with factor r; BEB is the special case with $r = 2$. In the notation of [3], we show that $N(t)/t$ converges to a non-zero value as t goes to infinity, and we show that p_c, the probability that a transmitted packet will experience a collision, is always smaller than $1/r$.

Network performance measures are usually given as a function of the offered load. A commonly used definition of offered load is N, the number of nodes waiting to transmit; this concept underlies EB [10], which indirectly estimates the number of nodes contending by counting consecutive collisions. A second definition of offered load is the total packet arrival rate of the system, relative to the channel capacity. Since the purpose of EB is to alleviate the effects of contention among the nodes and to adapt the system to the number of nodes, the first definition of offered load is more appropriate for analyzing EB and is used in this paper. For the same reason, the performance of EB should be evaluated based on its effect on the measures of network efficiency, such as throughput.

In this paper, we assume a fixed number of nodes N in saturation condition. Here *saturation condition* means that each node always has a packet to transmit. Thus, N represents the offered load of the network. We also assume no errors on the channel. Under this assumption, we analyze network *throughput* and *medium access delay* for a slotted system with EB and compare the analytical results with simulation. Note that all the previously cited works do not evaluate delay, even though some of the papers define stability in terms of delay. The saturation condition assumption is also made by Bianchi in [16], where he used his own approach to analyze the throughput of the distributed coordination function (DCF) mode of the IEEE 802.11 wireless LAN standards.

This paper is organized as follows. In Section II, we analyze EB to obtain the performance measures and establish stability. The analysis is carried out in several steps, which consists of modeling of EB, analysis of saturation throughput and medium access delay, and analysis of asymptotic behavior. Section III discusses some aspects of the simulation used to validate the analytical results. Section IV is the conclusion of the paper.

II. Analysis of EB

In our analysis, the time is divided into time slots of equal length, and all packets are assumed to be of the same duration, equal to the slot time. Furthermore, all nodes are assumed to be synchronized so that every transmission starts at the beginning of a slot and ends before the next slot. At its first transmission, a packet is transmitted after waiting the number of slots randomly selected from $\{0, 1, \cdots, W_0 - 1\}$, where $W_0 \geq 1$ is an integer representing the minimum contention window size. Every time a node's packet is involved in a collision, the contention window size for that node is multiplied by the *backoff factor* r and a random number is generated within the contention window for the next transmission attempt. Thus, on a packet's i-th retransmission, a random number is selected from $\{0, 1, \cdots, r^i W_0 - 1\}$ with equal probability, where $i = 0$ represents the first transmission attempt. With $r = 2$, this procedure is called *binary* exponential backoff.

A. Analytical Model of EB

The characteristic behavior of a backoff algorithm is critical when the channel is heavily loaded, and in fact, the very idea of EB is to cope with the heavily loaded channel condition. Thus, we analyze EB under saturation conditions. The saturation condition represents the largest possible load offered by the given number of nodes, which is a reasonable assumption for investigating EB.

We model the operation at an individual node using the state diagram in Fig. 1, in which each node is in one of an infinite number of backoff states and p_c denotes the probability

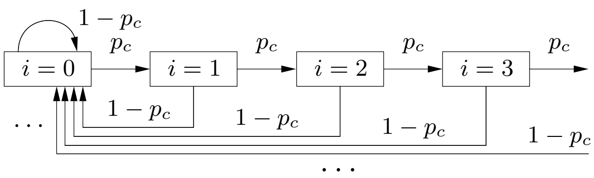

Fig. 1. State transition diagram in a node for exponential backoff (EB).

that a transmission experiences a collision. In backoff state i, $i = 0, 1, 2, \cdots$, the contention window size for a node is $W_i = r^i W_0$, where W_0 is the minimum contention window size. As the diagram in Fig. 1 indicates, after a successful transmission, which occurs with probability $1 - p_c$, from any other state a node enters backoff state $i = 0$ with contention window size W_0. While in backoff state $i = k$, after an unsuccessful transmission, a node enters backoff state $i = k + 1$ with probability p_c.

Denote B_k as the k-th state that a node enters. Then, B_k is a Markov chain with the transition probabilities $p_{i,j}$, $i, j = 0, 1, \cdots$, given as follows.

$$p_{i,0} = \Pr\{B_{k+1} = 0 \mid B_k = i\} = 1 - p_c, \tag{1}$$
$$p_{i,i+1} = \Pr\{B_{k+1} = i+1 \mid B_k = i\} = p_c, \tag{2}$$
$$p_{i,j} = 0, \quad j \neq 0,\ j \neq i+1. \tag{3}$$

Define a probability $P_i = \lim_{k\to\infty} \Pr\{B_k = i\}$, $i = 0, 1, \cdots$, then P_i is the relative frequency that a node will enter state i in the steady state. Since $\sum_{i=0}^{\infty} P_i = 1$, from Fig. 1, P_i can be obtained as follows

$$P_i = (1 - p_c)p_c^i, \quad i = 0, 1, \cdots . \tag{4}$$

B. Throughput

The main performance measure in evaluating a network is its throughput. We analyze the saturation throughput by calculating the probability that there is a successful transmission in a time slot.

The probability P_i given in (4) is the relative frequency that a node *enters* state i. However, the average time a node *stays* in a state is different for each state and is a function of the contention window size of the state. As illustrated in Fig. 2, if a node enters state i, an integer random variable D_i of uniform distribution between 0 and $W_i - 1$ including the boundaries is generated and after waiting for D_i time slots, the node will (re-)transmit the packet, after which the success or failure of the transmission will determine the next state of the node. Note that the node will stay in state i for $D_i + 1$ time slots until the node moves to the next state. On average, a node will stay in state i for $\overline{d}_i = \mathrm{E}[D_i{+}1] = (W_i{+}1)/2$ time slots. Let S_i be the probability that a node is in state i at a given time; then S_i specifies the distribution of nodes over the states. Since S_i is proportional to $P_i\overline{d}_i$, it is given by

$$S_i = \frac{P_i\overline{d}_i}{\sum_{j=0}^{\infty} P_j\overline{d}_j} = \frac{(1-p_c)p_c^i(1-rp_c)(W_i+1)}{W_0(1-p_c)+1-rp_c}, \tag{5}$$

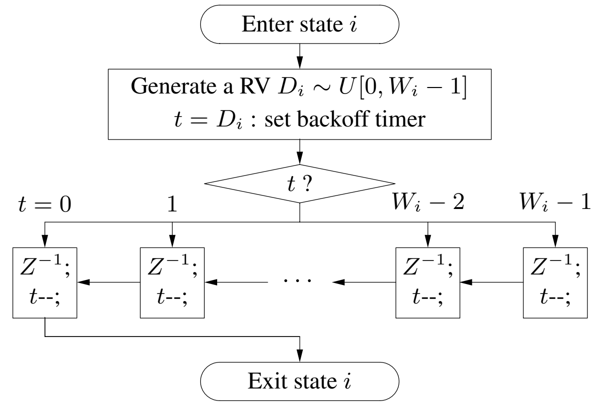

Fig. 2. Procedure taken by a node in state i. $U[0, W_i - 1]$ represents a uniform distribution of integer over the interval $[0, W_i - 1]$, Z^{-1} means time delay by one time slot, and t-- is a decrement operation on the backoff timer.

where the summation in the denominator does not exist if $rp_c \geq 1$. In fact, $rp_c < 1$ is a necessary condition for the system to reach steady state. Note that S_i is given as a function of p_c and W_0. Later, we show that the value of p_c is determined when the value of W_0, and the number of nodes N are given.

Define $\Pr\{t = k \mid i\}$ as the conditional probability that a node's backoff timer t will have value k given that the node is in state i. Since $\sum_{k=0}^{W_i-1} \Pr\{t = k \mid i\} = 1$, it follows that

$$S_i = \sum_{k=0}^{W_i-1} S_i \cdot \Pr\{t = k \mid i\} = \sum_{k=0}^{W_i-1} s_{i,k}, \tag{6}$$

where $s_{i,k}$ is the probability that the node is in state i *and* the backoff timer has value k. Since the backoff timer is decreased by one every slot time, $s_{i,k}$ satisfies

$$s_{i,k} = s_{i,W_i-1} \cdot (W_i - k), \quad k = 0, 1, \cdots, W_i - 1. \tag{7}$$

By substituting (7) into (6), it can be shown that $s_{i,W_i-1} = S_i/(\overline{d}_i\, W_i)$, and thus,

$$s_{i,k} = \frac{2(1-p_c)p_c^i(1-rp_c)}{W_0(1-p_c)+1-rp_c} \cdot \frac{W_i - k}{W_i}. \tag{8}$$

When $k = 0$, we have $s_{i,0}$, the probability that a node is in state i and the backoff timer is expired, that is, a node will transmit a packet in state i.

Let p_t be the probability that a given node will transmit in an arbitrary time slot. Then, since $s_{i,0}$, $i = 0, 1, \cdots$, are the probabilities of mutually exclusive events, $p_t = \sum_{i=0}^{\infty} s_{i,0}$. From (8), $s_{i,0} = s_{i-1,0}p_c$, $i = 1, 2, \cdots$. Thus,

$$p_t = \frac{s_{0,0}}{1-p_c} = \frac{2(1-rp_c)}{W_0(1-p_c)+1-rp_c}. \tag{9}$$

Note that p_t is a function of p_c and W_0, but also related to N through the value of p_c as will be shown later. As we shall see in the following, since p_c goes to $1/r$ as N goes to infinity, p_t converges to zero as the number of nodes goes to infinity.

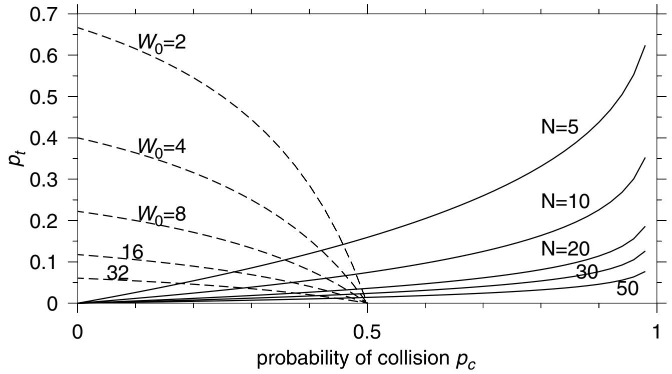

Fig. 3. Plots of p_t as a function of p_c; dashed lines: p_t in (9), solid lines: p_t in (11). $r = 2$.

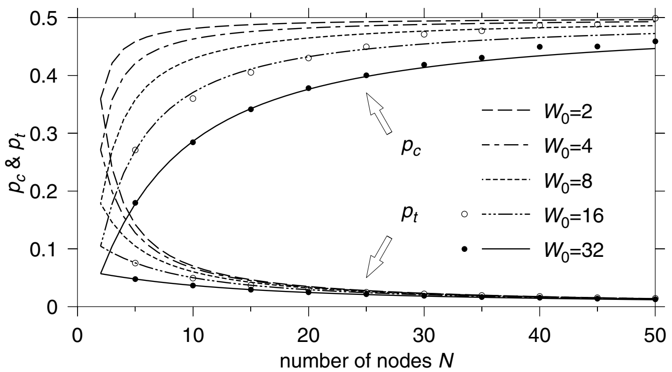

Fig. 4. Plots of the probability of collision p_c, and the probability of transmission p_t. $r = 2$.

As noted in [16], the numerical value of p_t is also constrained by the fact that p_c can be expressed in terms of p_t, that is

$$p_c = 1 - (1 - p_t)^{N-1}, \tag{10}$$

where $(1-p_t)^{N-1}$ is the probability that none of the other $N-1$ nodes transmits. Solving (10) for p_t, we have

$$p_t = 1 - (1 - p_c)^{1/(N-1)}. \tag{11}$$

Since (9) and (11) are two constraining equations for p_t as a function of p_c, the unique intersection of these two equations gives us the values of p_c and p_t for given N and W_0. Note that p_c is always less than $1/r$, which is the requirement for the existence of the summation in equation (5). Fig. 3 shows plots of p_t as a function of p_c given in equation (9) (dashed lines) for $r = 2$ and various values of W_0, and in equation (11) (solid lines) for various values of N. The probability of collision p_c and the probability of transmission p_t obtained numerically from (9) and (11) by calculating the intersection are plotted in Fig. 4. The plot shows p_c and p_t converging to $1/r$ $(= 0.5)$ and zero, respectively, as the number of nodes increases. The circles and bullets drawn along the curves are simulation results obtained for $W_0 = 16$ and $W_0 = 32$. Fig. 4 shows that the analytical and simulation results agree extremely well. (More discussion on the simulation is given in Section III.)

Since the channel is busy for a given time slot when there is at least one transmitting node in the slot time, the probability that the channel will be busy in a time slot is

$$P_{\text{busy}} = 1 - (1 - p_t)^N = 1 - P_{\text{idle}}, \tag{12}$$

where P_{idle} is the probability that a time slot is idle. On the other hand, a successful transmission occurs when there is only one transmitting node. Thus, the probability that there will be a successful transmission in a time slot is defined as

$$P_{\text{succ}} = {}_NC_1 \; p_t \; (1 - p_t)^{N-1} = Np_t(1 - p_t)^{N-1}, \tag{13}$$

where ${}_NC_1$ is the number of ways of choosing 1 out of N nodes. Note that a collision occurs if there are multiple nodes transmitting in the same time slot. Thus, the probability that a collision will occur in a time slot is given by $P_{\text{coll}} = P_{\text{busy}} - P_{\text{succ}}$.

If we normalize the slot time as the unit time, in any given unit time duration, the average number of frames that are successfully transmitted is P_{succ}. If we ignore the packet overhead, the normalized throughput is simply P_{succ}. In the notation of [3], $P_{\text{succ}} = \lim_{t\to\infty} N(t)/t$. Fig. 5 shows plots of P_{succ}, the normalized throughput, for various values of W_0. Note that P_{succ} converges to a non-zero constant ($\frac{1}{2}\ln 2$ to be precise when $r = 2$), which does not depend on W_0, as the number of nodes increases. Even when there are a lot of nodes contending for the medium access, BEB manages to control the transmission attempts in a slot to guarantee sustained probability of successful transmission. In fact, it is shown below in Section II-D that the average number of nodes that transmit in a slot converges to a constant less than 1 as the number of nodes goes to infinity. Note that, for a small number of nodes and large W_0, P_{succ} increases as the number of nodes increases. This behavior occurs because of the large number of idle time slots on the channel, and increasing the number of nodes increases the efficiency of the channel usage leading to a higher P_{succ}.

C. *Expected Medium Access Delay*

Delay is another key element in evaluating the performance of a network. We define the medium access delay as the time from the moment a packet is ready to be transmitted to the moment the packet starts its successful transmission. The medium access delay is obtained by analyzing the expected total number of backoff time slots.

P_i defined in (4) gives information regarding the behavior of a *node*. But the state transition information of *a packet* transmitted by a node is necessary to analyze the backoff profile of the packet. Let Q_i be the probability that a *packet* enters state i, $i = 0, 1, \cdots$, in steady state. Then $Q_0 = 1$ because every packet starts at state 0, and $Q_i = p_c Q_{i-1}$ since a packet enters state i when it experiences a collision in state $i - 1$. Thus, by mathematical induction, Q_i is given by $Q_i = p_c^i$, $i = 0, 1, 2, \cdots$. Note that Q_i is numerically identical to P_i after normalization, that is $P_i = Q_i \,/\, \sum_{j=0}^{\infty} Q_j$. Define T_n as the probability that a packet will be successfully transmitted on exactly the n-th retransmission, then

$$T_n = Q_n - Q_{n+1} = (1 - p_c)p_c^n, \quad n = 0, 1, 2, \cdots, \tag{14}$$

where T_0 is the probability that a packet will be successfully transmitted without retransmission. Let N_R be a random variable representing the number of retransmissions until success; then T_n is the probability mass function of N_R, and the average number of *re*-transmissions per packet is given by

$$n_R = E[N_R] = \sum_{n=0}^{\infty} nT_n = \frac{p_c}{1-p_c}. \tag{15}$$

On average, it requires $n_R + 1 = 1/(1-p_c)$ transmissions per packet. If a packet is retransmitted N_R times, then the packet will be delayed by $\sum_{i=0}^{N_R-1}(D_i+1) + D_{N_R}$ time slots, where D_i is an integer random variable of uniform distribution between 0 and $W_i - 1$. Thus, the expected number of time slots of backoff per packet is given by

$$\begin{aligned}\overline{D}_\Sigma &= \mathrm{E}\left[\sum_{i=0}^{N_R-1}(D_i+1) + D_{N_R}\right] \\ &= \mathrm{E}_{N_R}\left[\frac{1}{2}\left(N_R + 1 + \frac{W_0(1-r^{N_R+1})}{1-r}\right)\right] - 1 \\ &= \frac{1}{2}\left(\frac{1}{1-p_c} + \frac{W_0}{1-rp_c}\right) - 1 = \frac{n_R+1}{p_t} - 1, \end{aligned} \tag{16}$$

where $E_{N_R}[\cdot]$ is an ensemble average over the random variable N_R. Since it takes on average $\overline{D}_\Sigma$ time slots for a packet until successful transmission, $\overline{D}_\Sigma$ is the medium access delay in time slots. Fig. 6 shows the expected medium access delay in time slots for various values of W_0 obtained by analysis as well as by simulation. It shows that the medium access delay increases almost linearly with the number of nodes N. It is shown in Section II-D that the medium access delay approaches a linear function of N as the number of nodes goes to infinity.

D. Stability and Asymptotic Behavior of EB

Now we investigate the asymptotic behavior of EB observed when the number of nodes N goes to infinity. As shown in Fig. 4, p_t converges to zero as the number of nodes increases, due to the increased contention window sizes which causes a smaller probability of transmission in a given time slot. The following theorem describes the asymptotic behavior of p_t.

Theorem 1: Define $n_t = N \cdot p_t$ as the expected number of nodes that will transmit in an arbitrary time slot. Then, n_t converges to the non-zero value $\ln[r/(r-1)]$ as the number of nodes N goes to infinity.

The theorem can be proven first by showing that $\lim_{N\to\infty} p_c = 1/r$, and $\lim_{N\to\infty} p_t = 0$, then using them in (9) and (10). This theorem tells us two very important facts. First, n_t converges to a finite positive constant. In fact, with $r = 2$, n_t converges to $\ln 2 < 1$. Thus, no matter how many nodes the network contains, it can be expected that, on average, less than one node will try to transmit in any time slot, which in turn guarantees non-zero throughput of the network regardless of the number of the nodes in the network as shown in the following corollary. Secondly, $\lim_{N\to\infty} n_t$ is not a function of W_0. Thus, the minimum contention window size does not affect the asymptotic behavior of the network.

Fig. 7 shows the plots of Np_t vs. N along with simulation results. With a larger minimum contention window size W_0, the expected number of transmitting nodes in a slot is smaller because of the longer average backoff by each node. But as the number of nodes increases, all curves converge to the asymptote $\lim_{N\to\infty} n_t = \ln 2$, which is shown with a thin line in Fig. 7.

Corollary 2: The probability P_{busy} that channel is busy, and the probability P_{succ} that there will be a successful transmission in a time slot converge as the number of nodes N goes to infinity as follows:

$$\lim_{N\to\infty} P_{\text{busy}} = \frac{1}{r}, \quad \lim_{N\to\infty} P_{\text{succ}} = \frac{r-1}{r}\ln\frac{r}{r-1}. \tag{17}$$

The proof of Corollary 2 is straightforward from Theorem 1. As noted in Section II-B, P_{succ} represents the normalized throughput. Thus, the asymptote of P_{succ} in (17), drawn in Fig. 5 with thin solid line, shows that EB is stable under the throughput definition. Note that even with a large number of nodes, the channel is idle about 50% of the time (for $r = 2$), which guarantees sustained non-zero probability of successful transmission. This is due to the backoff mechanism controlling transmission attempts by the nodes. Consequently, as the number of nodes increases, the medium access delay also increases; each node has to wait longer to have its turn. In Section II-C, $\overline{D}_\Sigma$, the expected medium access delay was derived. To see the asymptotic behavior of $\overline{D}_\Sigma$, note that (16) can be written as $\overline{D}_\Sigma = \frac{N}{(1-p_c)\,Np_t} - 1$. Since, $\lim_{N\to\infty} Np_t = \ln(r/(r-1))$ and $\lim_{N\to\infty} p_c = 1/r$, $\overline{D}_\Sigma$ approaches an asymptote that is, a linear function of N, as N goes to infinity:

$$\lim_{N\to\infty} \overline{D}_\Sigma = \frac{N}{\frac{r-1}{r}\ln\frac{r}{r-1}} - 1 = \frac{1}{\lim_{N\to\infty} P_{\text{succ}}} N - 1 \tag{18}$$

Note that (18) is not a function of W_0. The thin solid line in Fig. 6 shows a plot of the asymptote (18).

III. Simulation

To support our analysis, simulation results are added in Figures 4–7, which are represented by circles and bullets, along with the curves of analytical results. The simulator is written in the C++ programming language, and simulation results were obtained by running 500,000 time slots after 10,000 time slots of warming up for $W_0 = 16, 32$, and $N = 5, 10, \cdots, 50$.

The simulation results in Figures 4–7 agree with those obtained from our analysis. However, when there are relatively many nodes, slight differences between the analytical and simulation results are observed which can be attributed to a *starvation* effect. A starvation effect is different from a capture effect. A capture effect makes only a few nodes consume the whole transmission channel, but a starvation effect gives a few nodes little chance to transmit their packets while most of nodes have fairly good chances. For example, for $N = 50$ and $W_0 = 16$,

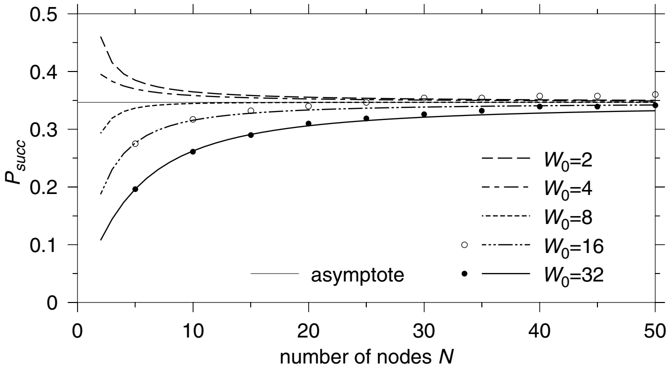

Fig. 5. The probability of successful transmission in a slot (normalized saturation throughput). $r = 2$.

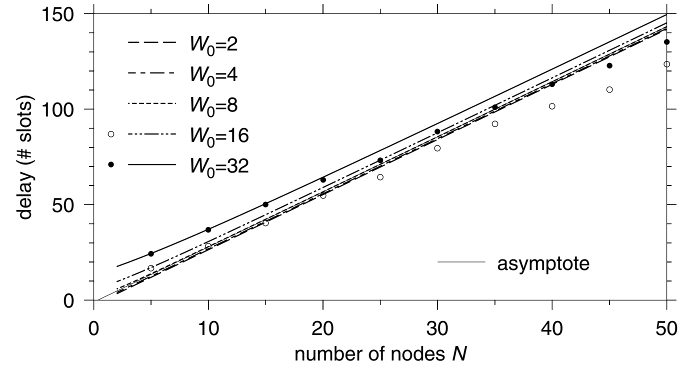

Fig. 6. The medium access delay in the number of time slots. $r = 2$.

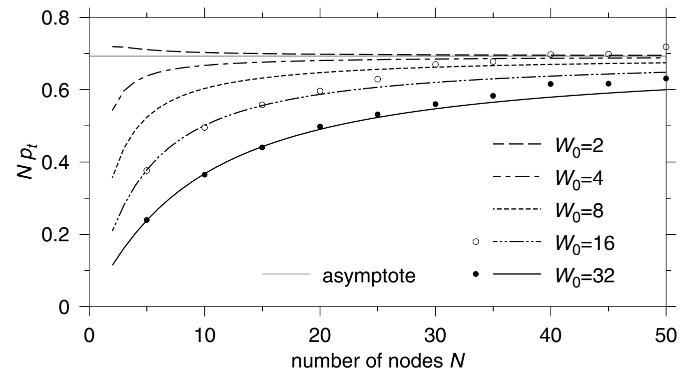

Fig. 7. The expected number of nodes Np_t that will transmit in an arbitrary time slot. $r = 2$.

while most nodes tried between 7,000 and 8,000 packet transmissions during 500,000 time slots, there was a node with only 5 transmission tries. Also, there were several nodes with less than 3,000 transmission attempts. For smaller W_0, however, a capture effect [17] was observed instead of starvation, whose result is not included in the figures. The reason for these effects is the necessarily finite observation time of the simulation.

IV. Conclusion

We analyzed the performance of EB to obtain the saturation throughput and the medium access delay. The stability of EB was also established by showing the asymptotic behavior of EB when the number of nodes goes to infinity. From the analysis results, we showed that EB guarantees a certain amount of throughput no matter how many nodes are present in the network. Also shown was that the medium access delay has a linear asymptotic behavior with respect to the number of nodes in the network. The performance of EB with maximum retry limit (known as truncated EB) is also analyzed in [18], where it is shown that EB with maximum retry limit is unstable due to the transmission retry limit.

In this paper, a new and efficient analytical method was used to analyze the characteristics of EB. This analytical method can also be applied to analyze network protocols using EB, such as IEEE 802.11 DCF.

References

[1] R. M. Metcalfe and D. R. Boggs, "Ethernet: Distributed packet switching for local computer networks," *Communications of the ACM*, vol. 19, no. 7, July 1976.

[2] "P802.11, IEEE standard for wireless lan medium access control (MAC) and physical layer (PHY) specifications," November 1997.

[3] D. J. Aldous, "Ultimate instability of exponential back-off protocol for acknowledgment-based transmission control of random access communication channels," *IEEE Trans. Information Theory*, vol. 33, no. 2, 1987.

[4] J. Goodman, A. G. Greenberg, N. Madras, and P. March, "Stability of binary exponential backoff," *Journal of the ACM*, vol. 35, no. 3, 1988.

[5] J. R. Shoch and J. A. Hupp, "Measured performance of an ethernet local network," *Commun. ACM*, vol. 23, no. 12, pp. 711–721, Dec. 1980.

[6] J. Håstad, T. Leighton, and B. Rogoff, "Analysis of backoff protocols for multiple access channels," *SIAM J. Computing*, vol. 25, no. 4, 1996.

[7] F.P. Kelly and I.M. MacPhee, "The number of packets transmitted by collision detect random access schemes," *Annals of Probability*, vol. 15, pp. 1557–1568, 1987.

[8] H. Al-Ammal, L. A. Goldberg, and P. MacKenzie, "Binary exponential backoff is stable for high arrival rates," in *Proc. 17th Int. Symp. on Theoretical Aspects of Computer Science*, Lille, France, Feb. 2000.

[9] H. Al-Ammal, L. A. Goldberg, and P. MacKenzie, "An improved stability bound for binary exponential backoff," *Theory of Computing Systems*, vol. 30, 2001.

[10] D. R. Boggs, J. C. Mogul, and C. A. Kent, "Measured capacity of an ethernet: Myths and reality," in *Proc. ACM Symp. on Commun. Architecture and Protocols (SIGCOMM 88)*, 1988, pp. 222–234.

[11] K. K. Ramakrishnan and H. Yang, "The ethernet capture effect: Analysis and solution," in *Proc. of 19th Conf. on Local Computer Networks*, 1994.

[12] Dong Geun Jeong and Wha Sook Jeon, "Performance of an exponential backoff scheme for slotted-aloha protocol in local wireless environment," *IEEE Transactions on Vehicular Technology*, vol. 44, no. 3, August 1995.

[13] K. Sakakibara, H. Muta, and Y. Yuba, "The effect of limiting the number of retransmission trials on the stability of slotted aloha systems," *IEEE Trans. Vehicular Technology*, vol. 49, no. 4, July 2000.

[14] K. Sakakibara, T. Seto, D. Yoshimura, and J. Yamakita, "On the stability of slotted aloha systems with exponential backoff and retransmission cutoff in slow-frequency-hopping channels," in *Proc. 4th Int. Symp. on Wireless Personal Multimedia Communications*, Aalborg, Denmark, Sep. 2001.

[15] L. A. Goldberg and P. D. MacKenzie, "Analysis of practical backoff protocols for contention resolution with multiple servers," *Journal of Computer and System Sciences*, 1999.

[16] Giuseppe Bianchi, "Performance analysis of the ieee 802.11 distributed coordination function," *JSAC Wireless series*, vol. 18, no. 3, March 2000.

[17] Mart L. Molle, "A new binary logarithmic arbitration method for ethernet," Tech. Rep. CSRI-298, Computer Systens Research Institute, University of Toronto, April 1994, Revised July 1994.

[18] Byung-Jae Kwak, Nah-Oak Song, and Leonard E. Miller, "Stability and performance analysis of exponential backoff," Submitted for review to IEEE Trans. Networking.