Performance Analysis of Exponential Backoff

Byung-Jae Kwak, Nah-Oak Song, *Member, IEEE*, and Leonard E. Miller, *Senior Member, IEEE*

***Abstract*—New analytical results are given for the performance of the exponential backoff (EB) algorithm. Most available studies on EB focus on the stability of the algorithm and little attention has been paid to the performance analysis of EB. In this paper, we analyze EB and obtain *saturation throughput* and *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; binary exponential backoff (BEB) algorithm is the special case with $r = 2$. We also derive the analytical performance of EB with maximum retry limit M (EB-M), a practical version of EB. The accuracy of the analysis is checked against simulation results.**

***Index Terms*—Backoff algorithm, BEB, exponential backoff, medium access delay, performance analysis, throughput.**

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 analyze the *exponential backoff* (EB) with backoff factor $r > 1$, where BEB is a special case of EB with $r = 2$. The analysis uses a model that closely resembles practical network transmission schemes and therefore is useful for system planning and analysis. We also derive the analytical performance of EB with maximum retry limit M (EB-M), a practical version of EB.

A. Previous Work on Exponential Backoff Algorithm

Many papers study exponential backoff algorithms including BEB, in terms of their effect on network performance as the offered load increases. Most of the works available, which are summarized below, focus on the stability of EB, and little work has been conducted on its performance analysis. Furthermore, these studies have produced contradictory results on the stability of EB instead of a consensus: some prove instability, others show stability under certain conditions. The mixed results are due to the differences in the analytical models, where simplified and/or modified models of the backoff algorithm are used, and in 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, but 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 nonzero 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 a bounded backlog of packets in the queue, or the recurrent property of Markov chains [6].

Most of the analytical and simulation studies on EB treat the backoff algorithm in the context of a specific network medium access control (MAC) protocol such as Ethernet [7]–[11] or WLAN [12]. The characteristics of the specific protocol seem to have as much effect on the network performance results as the intrinsic behavior of EB. Thus, the results depend heavily on which MAC protocol is used in the study and it is not possible to understand the behavior of EB from the results. Some of the analytical works that focus on EB itself are summarized as follows:

Kelly and MacPhee [13] 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 nonzero arrival rate, where $N(t)$ is the number of the successful transmissions made during the time $[0, t]$. This result tries to solve the open problem left in [13], 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

Manuscript received September 26, 2002; revised November 16, 2003; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor Z. Haas.

B.-J. Kwak and N.-O. Song are with the Electronics and Telecommunications Research Institute, Daejeon 305-350, Korea (e-mail: bjkwak@etri.re.kr; nsong@etri.re.kr).

L. E. Miller is with the National Institute of Standards and Technology, Gaithersburg, MD 20899-8920 USA (e-mail: lmiller@antd.nist.gov).

Digital Object Identifier 10.1109/TNET.2005.845533

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 [14], Al-Ammal *et al.* give a 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 [14] is further improved in [15], 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 the 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 [16], the infinite-node model used in [13] and [3] is a mathematical abstraction with limited practical application.

And except for [13], all of the studies cited actually analyze a modified BEB with p-persistent backoff protocol, where the probability p is exponentially decreased on collision instead of increasing the contention window. In BEB, after i consecutive packet transmission failures (collisions) a node selects *a single random slot* from the next 2^i slots (contention window) with equal probability for the next transmission. On the other hand, in the modified BEB, after i collisions, the probability of packet transmission in each slot is 2^{-i} until the transmission occurs, which can happen after 2^i slots. Clearly, it is easier to analyze the p-persistent like modified BEB because of its memoryless nature, but it is not guaranteed that it has the same characteristics as BEB.

B. Approach of This Paper

As mentioned earlier, most available studies on EB focus on the stability of the algorithm and little attention has been paid to the performance analysis of EB. However, the performance of EB—the characteristic behavior of the throughput and delay with respect to the network load—is of more importance in practical situations. In this paper, we analyze the performance of EB in steady state in terms of network load.

Network performance measures are usually given as a function of the offered load, which is the actual traffic demand presented to the network. The number of nodes that are contending for the access of the medium can be used as an offered load; this concept is what underlies BEB [7], which indirectly estimates the number of nodes contending by counting consecutive collisions. Another commonly used offered load is the total packet arrival rate of the system relative to the channel capacity. Since the purpose of BEB is to alleviate the effects of contention among the nodes and to adapt the system to the number of nodes, the number of nodes contending for medium access is a more appropriate definition of offered load for analyzing BEB, 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 conditions. Here *saturation condition* means that each node always has a packet to transmit. Thus, in our analysis, N represents the offered load of the network. We also assume an implicit ACK and no errors on the channel. The saturation condition assumption is also made by Bianchi in [12] and by Wu *et al.* in [17], where they used 2-D Markov chain models to analyze the throughput of the distributed coordination function (DCF) mode of the IEEE 802.11 wireless LAN standard. In our analysis, a 1-D Markov chain model is used instead of a 2-D Markov chain model, which enables us to analyze medium access delay in addition to saturation throughput.

Under this assumption, we analyze network *throughput* and *medium access delay* in steady state for a slotted system with exponential backoff and compare the analytical results with simulation. The analysis considers the general case of EB with backoff factor r; BEB is the special case with $r = 2$. Note that all the previously cited works do not evaluate delay, even though some of the papers define stability in terms of delay.

We consider two different versions of EB in our analysis. The first of the two is the original EB which is simply referred to as "EB" in this paper. In the analysis, we derive the analytical performance measures such as the throughput and the medium access delay of the network in term of the network load. From the analytical results, we also show the asymptotic behavior of EB as the number of nodes goes to infinity.

The second version considered is EB with maximum retry limit M, where a packet is dropped after M transmission retries. Variants of this truncated version of (binary) exponential backoff have been specified as part of the MAC protocol in several network standards, including the MAC layer of Ethernet local area networks (LANs) (IEEE 802.3) and the wireless LAN standard IEEE 802.11 [2], [18]. These truncated versions are referred to as EB in many papers [16]. We call it "EB-M" in this paper.

Even though the analyses of EB and EB-M share many common procedures, we deliberately separate the analyses to emphasize the differences in characteristics of the two versions of EB. The analysis of EB has been a theoretically very interesting problem as demonstrated by the numerous papers cited in Section I-A. Our interest in the characteristics of EB-M stems from the practical aspect of it as it is used in real network protocols successfully deployed and being used. By separate treatment, features unique to EB-M can be more clearly examined. Care is taken to minimize unnecessary repetition of the analytical steps.

This paper is organized as follows. In Section II, we analyze EB to obtain the analytical performance measures. The analysis is carried out in several steps, which consist of modeling of EB, analysis of saturation throughput and medium access delay, and analysis of asymptotic behavior. In Section III, EB with maximum retry limit is analyzed. Section IV presents simulation results with discussion of the applicability of the analysis model. Finally, Section V 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 $\{0, 1, \ldots, W_0 - 1\}$, where W_0 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 an integer random number D_i is generated within the contention window for the next transmission attempt, where on a packet's ith retransmission

$$\Pr\{D_i = k, k = 0, 1, \ldots, X_i - 1\} = \frac{X_i + 1 - Y_i}{X_i(X_i + 1)}$$
$$\Pr\{D_i = X_i\} = \frac{Y_i}{X_i + 1} \tag{1}$$

where X_i and Y_i are the integer and fractional parts of $r^i W_0$ defined by $X_i = \lfloor r^i W_0 \rfloor$ and $Y_i = r^i W_0 - X_i$, respectively. $i = 0$ represents the first transmission attempt. For integer r, this operation is equivalent to randomly selecting a number from $\{0, 1, \ldots, r^i W_0 - 1\}$ with equal probability $1/r^i W_0$. 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, which represent the largest possible load offered by the given number of nodes, a reasonable assumption for investigating EB.

We model the operation of EB at an individual node using the state diagram in Fig. 1(a), in which each node is in one of an infinite number of backoff states in steady state and p_c denotes the probability that a transmission experiences a collision. In backoff state i, $i = 0, 1, 2, \ldots$, 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(a) 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.

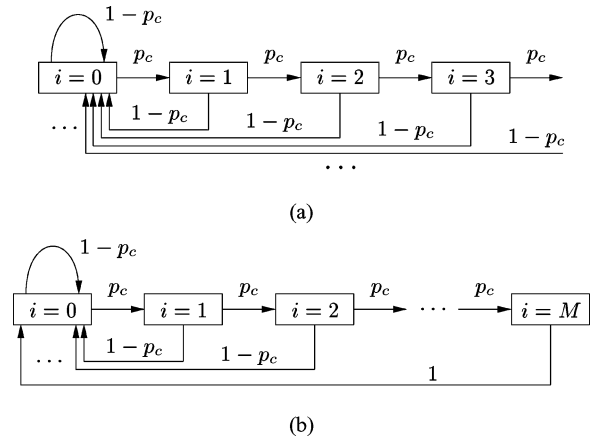


Fig. 1. State transition diagram of a node in steady state. (a) EB. (b) EB with maximum retry limit M.

Denote B_k as the kth state that a node enters in steady state. Then, B_k is a Markov chain with the transition probabilities $p_{i,j}$ given as follows:

$$p_{i,0} = \Pr\{B_{k+1} = 0 | B_k = i\} = 1 - p_c, \quad i = 0, 1, \ldots$$
$$p_{i,i+1} = \Pr\{B_{k+1} = i + 1 | B_k = i\} = p_c, \quad i = 0, 1, \ldots$$
$$p_{i,j} = 0, \quad j \neq 0, \quad j \neq i + 1.$$

Let P_i be a probability defined as

$$P_i = \Pr\{B_k = i\}, \quad i = 0, 1, \ldots \tag{2}$$

then P_i is the relative frequency that a node will enter state i in steady state. Since $\sum_{i=0}^{\infty} P_i = 1$, from Fig. 1(a), P_i can be obtained as follows:

$$P_i = (1 - p_c) p_c^i, \quad i = 0, 1, \ldots. \tag{3}$$

B. EB Throughput

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

The probability P_i given in (3) 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 distribution given by (1) 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

$$\overline{d}_i = \mathrm{E}[D_i + 1] = \frac{W_i + 1}{2} \tag{4}$$

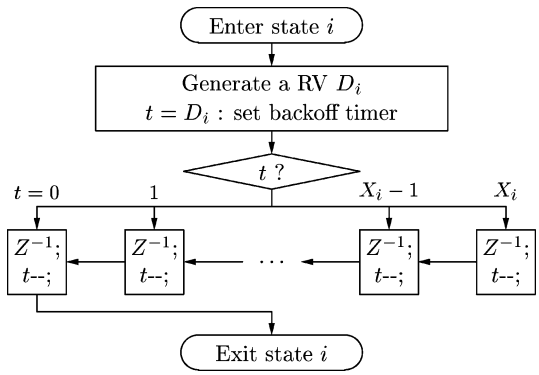


Fig. 2. Procedure taken by a node in state i. Z^{-1} means time delay by one time slot, and t-- is a decrement operation on the backoff timer.

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

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

where the summation

$$\sum_{i=0}^{\infty} P_i\overline{d}_i = \frac{W_0(1-p_c)+1-rp_c}{2(1-rp_c)} \tag{6}$$

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. As the number of nodes N grows, there are more transmission attempts and, as a result, the probability of collision p_c will also increase. Consequently, as shown in Section II-D, rp_c approaches 1 as the number of nodes goes to infinity, and for large N

$$\begin{aligned} S_i &= \frac{(1-p_c)(1-rp_c)}{W_0(1-p_c)+1-rp_c}\left(p_c^i + (rp_c)^i W_0\right) \\ &\approx (1-rp_c)(rp_c)^i \end{aligned} \tag{7}$$

where it is assumed that $1-rp_c \ll W_0(1-p_c)$ and $p_c^i \ll (rp_c)^i$. Equation (7) implies that, if the channel is crowded ($rp_c \approx 1$), the nodes are relatively evenly distributed over the states and the number of nodes in state i is similar to the number of nodes in state $i+1$ in steady state. This is because a node has $1/p_c \approx r$ times higher chance of entering state i than entering state $i+1$, but a node will stay r times longer in state $i+1$ than it does in state i on average.

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

$$S_i = \sum_{k=0}^{X_i} S_i \cdot \Pr\{t = k|i\} = \sum_{k=0}^{X_i} s_{i,k} \tag{8}$$

where $s_{i,k}$ is the probability that the node is in state i *and* the backoff timer has value k. Since the backoff counter is decreased by one with probability one at the elapse of each slot time, from the pmf (probability mass function) of D_i given in (1), there exist x and y such that

$$s_{i,k} = x(X_i - k) + y, \quad k = 0, 1, \ldots, X_i - 1 \tag{9}$$

$$s_{i,X_i} = y \tag{10}$$

where

$$x : y = \frac{X_i + 1 - Y_i}{X_i(X_i+1)} : \frac{Y_i}{X_i+1} \tag{11}$$

or equivalently

$$y = \frac{X_iY_i}{X_i + 1 - Y_i}x. \tag{12}$$

By substituting (9), (10), and (12) into (8), we obtain

$$x = \frac{X_i + 1 - Y_i}{(X_i+1)X_i}\frac{S_i}{\overline{d}_i} \quad y = \frac{Y_i}{X_i+1}\frac{S_i}{\overline{d}_i} \tag{13}$$

where $W_i = X_i + Y_i$ and (4) are used. If r is an integer, the equations in (13) are reduced to

$$x = s_{i,W_i-1} = \frac{S_i}{\overline{d}_i W_i} \quad y = 0.$$

From (13), (9) can be written as

$$s_{i,k} = \left(1 - \frac{X_i + 1 - Y_i}{(X_i+1)X_i}k\right)\frac{S_i}{\overline{d}_i}. \tag{14}$$

When $k = 0$, we have

$$s_{i,0} = \frac{S_i}{\overline{d}_i} = \frac{2(1-p_c)p_c^i(1-rp_c)}{W_0(1-p_c)+1-rp_c} \tag{15}$$

where $s_{i,0}$ is 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 $s_{i,0}$, $i = 0, 1, \ldots$, are the probabilities of mutually exclusive events,

$$p_t = \sum_{i=0}^{\infty} s_{i,0}. \tag{16}$$

From (15), $s_{i,0} = s_{i-1,0}p_c$, $i = 1, 2, \ldots$. Thus,

$$p_t = \frac{s_{0,0}}{1-p_c} = \frac{2(1-rp_c)}{W_0(1-p_c)+1-rp_c}. \tag{17}$$

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 discussion, 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.

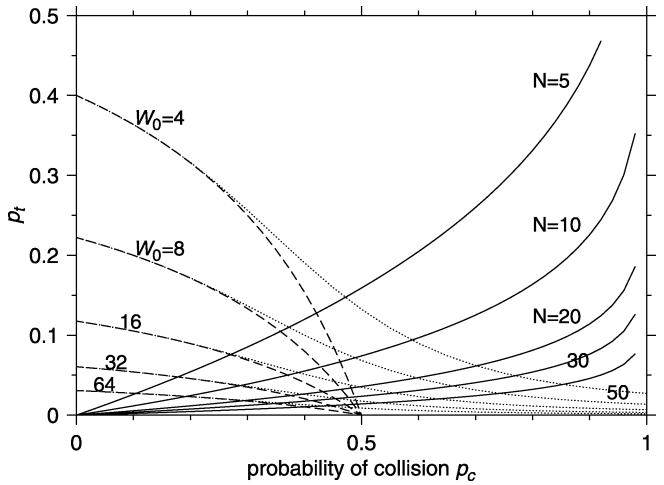


Fig. 3. Plots of p_t as a function of p_c; dashed lines: p_t in (17), solid lines: p_t in (19), dotted lines: p_t in (42) with $M = 6$. $r = 2$.

As noted in [12], 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

$$\begin{aligned} p_c &= 1 - \Pr\{\text{none of the other } N-1 \text{ nodes will transmit}\} \\ &= 1 - (1-p_t)^{N-1}. \end{aligned} \tag{18}$$

Solving (18) for p_t, we have

$$p_t = 1 - (1-p_c)^{1/(N-1)}. \tag{19}$$

Equation (19) is a continuous monotonically increasing function of p_c in the range [0,1], and $p_t = 0$ at $p_c = 0$ and $p_t = 1$ at $p_c = 1$. Equation (17) is also a continuous function of p_c in the range $[0, 1/r]$. Furthermore, since

$$\frac{dp_t}{dp_c} = \frac{-2(r-1)W_0}{[W_0(1-p_c)+1-rp_c]^2} < 0, \quad \text{for } p_c \in \left[0, \frac{1}{r}\right] \tag{20}$$

(17) is a monotonically decreasing function in the interval with $p_t > 0$ at $p_c = 0$ and $p_t = 0$ at $p_c = 1/r$. Therefore, the curves in (17) and (19) have a unique intersection, which can be easily calculated numerically. Since (17) and (19) 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 (6). Fig. 3 shows plots of p_t as a function of p_c given in (17) (dashed lines) for $r = 2$ and various values of W_0, and in (19) (solid lines) for various values of N. The probability of collision p_c and the probability of transmission p_t, obtained numerically from (17) and (19) by calculating the intersection, are plotted in Fig. 4(a). The plot shows p_c and p_t converging to $1/r$ $(= 0.5)$ and zero, respectively, as the number of nodes increases. The symbols drawn along the curves are simulation results obtained for $W_0 = 16$, $W_0 = 32$, and $W_0 = 64$, respectively. Fig. 4(a) shows that the analytical and simulation results agree extremely well. (More discussion on the simulation is given in Section IV.)

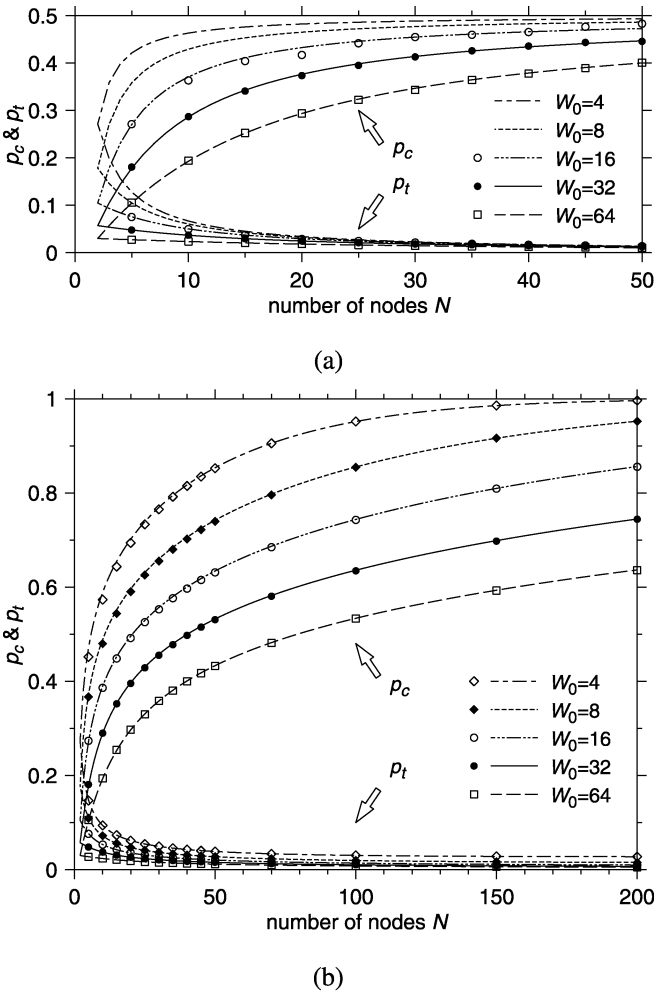


Fig. 4. Plots of the probability of collision p_c, and the probability of transmission p_t. $r = 2$. (a) EB. (b) EB with maximum retry limit $(M = 6)$.

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

$$P_{\text{busy}} = 1 - P_{\text{idle}} = 1 - (1-p_t)^N \tag{21}$$

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

$$P_{\text{succ}} = {}_NC_1 p_t (1-p_t)^{N-1} = N p_t (1-p_t)^{N-1} \tag{22}$$

where ${}_NC_1$ is the number of ways of choosing one 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}. Thus, if we ignore the packet overhead which may consist of MAC/PHY header and explicit ACK, the normalized throughput is simply P_{succ}. In the notation of [3], $P_{\text{succ}} = \lim_{t\to\infty} N(t)/t$. Fig. 5(a) shows plots of P_{succ}, the normalized throughput, for various values of W_0. Note that

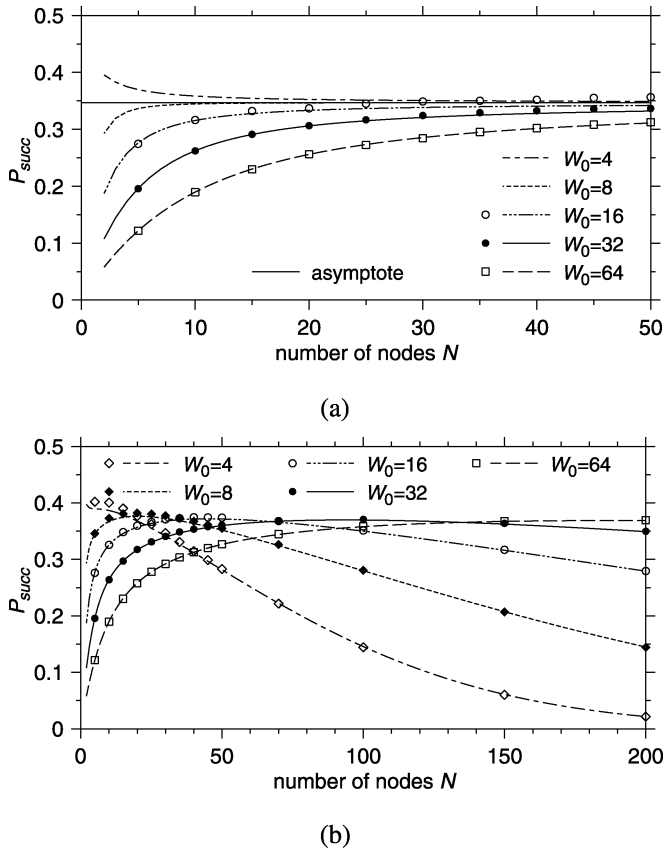


Fig. 5. The probability of successful transmission in a slot (normalized saturation throughput). $r = 2$. (a) EB. (b) EB with maximum retry limit $(M = 6)$.

$P_{\rm succ}$ converges to a *nonzero* constant ($1/2\ln 2$ to be precise when $r = 2$), which does not depend on W_0, as the number of nodes increases. Even when there are many nodes contending for the medium access, EB manages to control the transmission attempts in a slot to guarantee sustained probability of successful transmission. In fact, it is shown 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, $P_{\rm succ}$ 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 $P_{\rm succ}$.

C. EB 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 (3) 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, $i = 0, 1, \ldots$, in steady state. Then $Q_0 = 1$ because every packet starts at state $i = 0$, 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

$$Q_i = p_c^i, \quad i = 0, 1, 2, \ldots. \tag{23}$$

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 nth retransmission, then

$$T_n = Q_n - Q_{n+1} = (1 - p_c) p_c^n, \quad n = 0, 1, 2, \ldots \tag{24}$$

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 retransmissions per packet is given by

$$n_R = E[N_R] = \sum_{n=0}^{\infty} n T_n = \frac{p_c}{1 - p_c}. \tag{25}$$

On average, it requires

$$n_R + 1 = \frac{1}{1 - p_c} \tag{26}$$

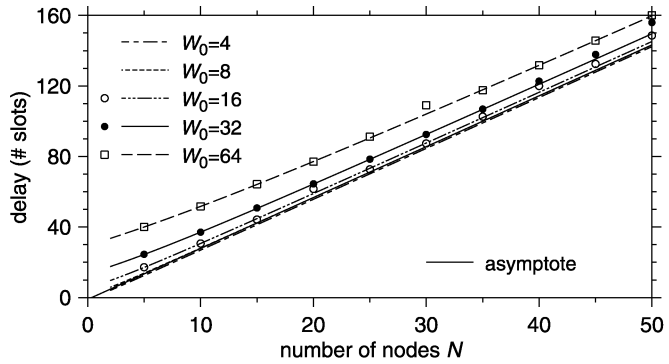
transmissions per packet for successful transmission. If a packet is retransmitted N_R times, then the packet will be delayed by $\sum_{i=0}^{N_R - 1}(D_i + 1) + D_{N_R}$ time slots, where D_i is an integer random variable defined in (1). Thus, the expected number of time slots of backoff per packet is given by

$$\begin{aligned} \overline{D}_{\Sigma} &= \mathrm{E}\left[\sum_{i=0}^{N_R - 1}(D_i + 1) + D_{N_R}\right] \\ &= \mathrm{E}_{N_R}\left[\frac{1}{2}\left(N_R + 1 + \frac{W_0(1 - r^{N_R+1})}{1 - r}\right)\right] - 1 \\ &= \frac{1}{2}\left(\frac{1}{1 - p_c} + \frac{W_0}{1 - r p_c}\right) - 1 = \frac{n_R + 1}{p_t} - 1 \end{aligned} \tag{27}$$

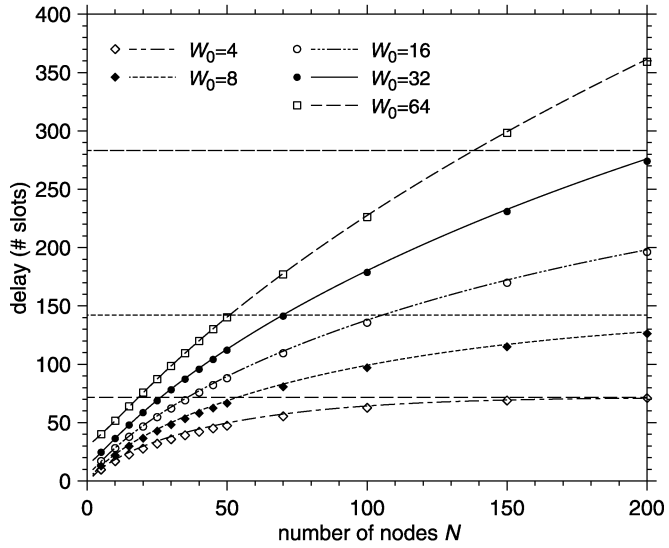
where $E_{N_R}[\cdot]$ is an ensemble average over the random variable N_R. The derivation of (27) can be found in Appendix II. Since it takes on average $\overline{D}_{\Sigma}$ time slots for a packet until successful transmission, $\overline{D}_{\Sigma}$ is the medium access delay in time slots. Fig. 6(a) illustrates 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. Asymptotic Behavior of EB

Now we investigate the asymptotic behavior of the analysis model for EB observed when the number of nodes N goes to infinity. As shown in Fig. 4(a), 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.



(a)



(b)

Fig. 6. Medium access delay in the number of time slots. $r = 2$. (a) EB. (b) EB with maximum retry limit $(M = 6)$.

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 nonzero value $\ln[r/(r-1)]$ as the number of nodes N goes to infinity.

Proof: We first show that p_c converges to $1/r$, and p_t converges to zero as N goes to infinity. From (17) and (19), we have

$$p_t = 1 - (1 - p_c)^{\frac{1}{N-1}} = \frac{2(1 - rp_c)}{W_0(1 - p_c) + 1 - rp_c}. \tag{28}$$

Taking an infinite limit of N, and noting that

$$\lim_{N\to\infty} p_t = \lim_{N\to\infty} \left(1 - (1 - p_c)^{\frac{1}{N-1}}\right) = 0$$

we have

$$0 = \lim_{N\to\infty} \frac{2(1 - rp_c)}{W_0(1 - p_c) + 1 - rp_c}$$

which implies $\lim_{N\to\infty} 2(1 - rp_c) = 0$. Thus,

$$\lim_{N\to\infty} p_c = \frac{1}{r}. \tag{29}$$

Now, rewrite (17) to yield p_c as a function of p_t as follows:

$$p_c = \frac{2 - (1 + W_0)p_t}{2r - (r + W_0)p_t}. \tag{30}$$

By equating (18) and (30), we have

$$\frac{r-1}{r} \cdot \frac{2 - p_t}{2 - \left(1 + \frac{W_0}{r}\right) p_t} = (1 - p_t)^{N-1}. \tag{31}$$

Since $\lim_{N\to\infty} p_t = 0$, by multiplying both sides by $1 - p_t$ and taking the limit as $N \to \infty$, we have

$$\frac{r-1}{r} = \lim_{N\to\infty} (1 - p_t)^N. \tag{32}$$

By taking the natural logarithm of both sides, (32) can be written as

$$\begin{aligned} \ln \frac{r-1}{r} &= \lim_{N\to\infty} N \ln(1 - p_t) \\ &= \lim_{N\to\infty} Np_t \lim_{N\to\infty} \frac{\ln(1 - p_t)}{p_t} \\ &= -\lim_{N\to\infty} Np_t \end{aligned}$$

which concludes the proof. ■

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 nonzero 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. These facts, however, do not hold for EB with maximum retry limit.

Fig. 7(a) shows the plots of Np_t versus 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(a).

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:

$$\lim_{N\to\infty} P_{\text{busy}} = \frac{1}{r}, \tag{33}$$

$$\lim_{N\to\infty} P_{\text{succ}} = \frac{r-1}{r} \ln \frac{r}{r-1}. \tag{34}$$

The proof of Corollary 2 is straightforward from Theorem 1. As noted in Section II-B, P_{succ} represents the normalized throughput. Thus, (34) in Corollary 2 shows that the normalized throughput of EB converges to a nonzero constant as the load of the network goes to infinity. The asymptote in (34) is drawn in Fig. 5(a) with thin solid line. Note that even with a large number of nodes, the channel is idle about 50% of the time (for $r = 2$), which guarantees sustained nonzero probability of

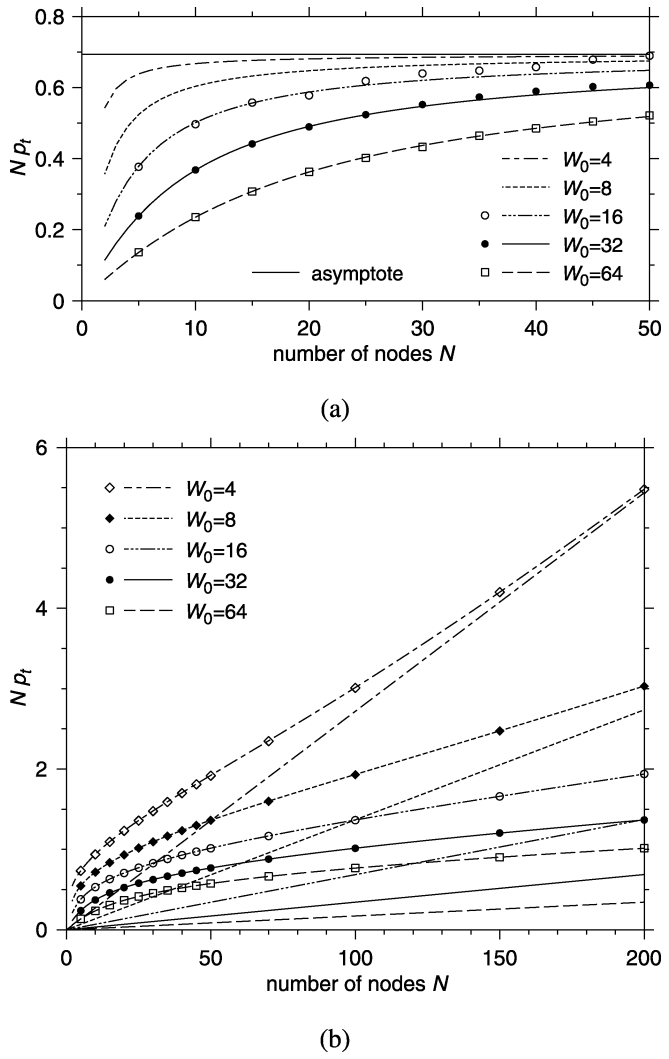


Fig. 7. The expected number of nodes Np_t that will transmit in an arbitrary time slot. $r = 2$. (a) EB. (b) EB with maximum retry limit ($M = 6$).

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 (27) can be written as

$$\overline{D}_\Sigma = \frac{N}{(1-p_c)Np_t} - 1. \tag{35}$$

From (18) and (22), (35) can be written as

$$N = P_{\text{succ}}(\overline{D}_\Sigma + 1)$$

which is a variation of Little's result. 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

$$\lim_{N\to\infty} \overline{D}_\Sigma = \frac{N}{\frac{r-1}{r}\ln\frac{r}{r-1}} - 1 = \frac{1}{\lim_{N\to\infty} P_{\text{succ}}} N - 1. \tag{36}$$

Note that (36) is not a function of W_0. The thin solid line in Fig. 6(a) shows a plot of the asymptote (36).

As shown in (34), the throughput in the limit (the asymptotic throughput) is a function of the backoff factor r. The optimum backoff factor r_{opt} that maximizes the asymptotic throughput, and the asymptotic throughput when $r = r_{\text{opt}}$ are as follows:

$$r_{\text{opt}} = \frac{1}{1-e^{-1}}$$
$$\left(\lim_{N\to\infty} P_{\text{succ}}\right)_{\text{opt}} = \frac{1}{e}.$$

III. Analysis of EB With Maximum Retry Limit

EB works well as far as the throughput of the network is concerned. Assume, however, that a node is trying to send a packet to another node and the destination node is not reachable. Then the source node will continue to retransmit the packet, and other packets in the queue destined to other reachable nodes will not be able to get through even when the channel is idle. Furthermore, without a transmission retry limit, the packet transmission delay can grow beyond reasonable range. To avoid this kind of problem, practical network protocols use truncated BEB, where the contention window W_i is given by

$$W_i = \begin{cases} r^i W_0, & \text{for } i = 0, 1, \ldots, m \\ r^m W_0, & \text{for } i = m, m+1, \ldots, M \end{cases}$$

when $M > m$, and

$$W_i = r^i W_0, \quad i = 1, 2, \ldots, M$$

when $M \le m$, where M is the maximum retry limit and m determines maximum contention window size. We analyze the latter case where $M \le m$, which we call EB with maximum retry limit (EB-M) in this paper. The case of $M > m$ can be readily analyzed using the technique in this section.

A. Analytical Model of EB With Maximum Retry Limit

Fig. 1(b) illustrates the state transition diagram of EB-M. As shown in the figure, there is a limit, M, to the maximum retries per packet and after reaching state M, a node goes to state 0 with a new packet even if the transmission was unsuccessful.

Denote B_k as the kth state that a node enters. Then, B_k is a Markov chain with the transition probabilities $p_{i,j}$ given as follows:

$$\begin{aligned} p_{i,0} &= \Pr\{B_{k+1}=0 | B_k=i\} = 1-p_c, \quad i=0,1,\ldots,M-1 \\ p_{i,i+1} &= \Pr\{B_{k+1}=i+1 | B_k=i\} = p_c, \quad i=0,1,\ldots,M-1 \\ p_{M,0} &= 1 \\ p_{i,j} &= 0, \quad i=1,2,\ldots,M-1, \text{ and } (j\neq 0,\ j\neq i+1). \end{aligned}$$

Let P_i be a probability defined as

$$P_i = \lim_{k\to\infty} \Pr\{B_k = i\}, \quad i = 0, 1, \ldots, M \tag{37}$$

then P_i is the relative frequency that a node will enter state i in the steady state. Since $\sum_{i=0}^{M} P_i = 1$, from Fig. 1(b), P_i can be obtained as follows:

$$P_i = \frac{1-p_c}{1-p_c^{M+1}} p_c^i, \quad i = 0, 1, \ldots, M. \tag{38}$$

B. Throughput for EB With Maximum Retry Limit

For EB-M, the expected time that a node will stay in state i is the same as that of EB shown in (4). Thus, the probability that an arbitrary node is in state i is

$$S_i = \frac{P_i \overline{d}_i}{\sum_{j=0}^{M} P_j \overline{d}_j} = \frac{p_c^i (W_i + 1)}{W_0 \sum_{j=0}^{M} (rp_c)^j + \sum_{j=0}^{M} p_c^j}. \tag{39}$$

Note that $rp_c < 1$ is not required for the existence of the summation in (39). In fact, p_c has a value greater than $1/r$ when the number of nodes is large enough.

For large N, (39) can be approximated as

$$S_i \approx \frac{(rp_c)^i}{\sum_{j=0}^{M} (rp_c)^j} \tag{40}$$

where it is assumed that $\sum_{j=0}^{M} p_c^j \ll W_0 \sum_{j=0}^{M} (rp_c)^j$ and $1 \ll W_i$. Since $p_c > (1/r)$ is possible, (40) implies that there can be more nodes in state $i+1$ than in state i, while it was always $S_i > S_{i+1}$ for EB in Section II.

Substituting $k = 0$ and (39) into (14),[1] we have

$$s_{i,0} = \frac{2p_c^i}{W_0 \sum_{j=0}^{M} (rp_c)^j + \sum_{j=0}^{M} p_c^j}. \tag{41}$$

Since $p_t = \sum_{i=0}^{M} s_{i,0}$ (note that the summation is from 0 through M), we obtain

$$p_t = \frac{2}{W_0 \sum_{i=0}^{M} (rp_c)^i / \sum_{i=0}^{M} p_c^i + 1}. \tag{42}$$

The dotted lines in Fig. 3 are the plots of (42) for various values of W_0. Note that when p_c is small, the curves for (42) and (17) (dashed lines; p_t as a function of p_c for EB) follow almost the same trajectories. This is because most of the packets are successfully transmitted within a small number of retries and thus affected little by the maximum retry limit. As p_c gets larger, however, packets experience more collisions and the characteristics of the system are determined by the maximum retry limit, and the curves for (42) and (17) follow different trajectories. Furthermore, while (17) is defined only for $0 \le p_c < 1/r$, (42) is defined for $0 \le p_c \le 1$. In EB, the mechanism of the backoff algorithm and the unbounded contention window size make p_c upper bounded by $1/r$.[2] On the other hand, in EB-M, p_t in (42) is always greater than zero due to the upper bounded contention window size, and thus as N goes to infinity, Np_t goes to infinity and p_c converges to unity. Note that it is easily shown that (42) converges to (17) as M goes to infinity when $rp_c < 1$.

p_t in (42) is also a continuous monotonically decreasing function of p_c in the range of [0,1] as shown in Appendix I. Thus (19) and (42) have a unique intersection, and the intersection determines the probability of collision p_c, and the probability of transmission p_t for given values of N and W_0. Fig. 4(b) shows the probability of collision p_c and the probability of transmission p_t, obtained numerically from (42) and (19) by calculating the intersection, along with the simulation results represented by circles and bullets. Unlike for EB, p_c converges to 1 as the number of nodes increases. Furthermore, p_t does not converge to zero. This is due to the maximum retry limit, which limits the backoff.

For EB-M, P_{busy} and P_{succ} (the normalized throughput) are the same as for EB as defined in (21) and (22), respectively. However, since p_t for EB-M is different from that of EB, the behavior of P_{busy} and P_{succ} show big differences. See Fig. 5(b) for plots of P_{succ} for various values of W_0. The most noticeable difference from EB is that P_{succ} diminishes to zero as the number of nodes goes to infinity as shown in Section III-D. This is because of the excessive collisions caused by the limited backoff. Also note that when the number of nodes is small, P_{succ} for EB-M is slightly larger than the value for EB.

C. Expected Medium Access Delay for EB With Maximum Retry Limit

Let Q_i be the probability that a *packet* enters state i, $i = 0, 1, \ldots, M$, in steady state. Then, as in the case of EB, $Q_i = p_c^i$, $i = 0, 1, 2, \ldots, M$. Define T_n as the probability that a node will successfully transmit on nth retransmission, and let P_{drop} be the probability that a packet will be dropped after maximum allowed transmission retries, then $T_n = Q_n - Q_{n+1}$, $n = 0, 1, \ldots, M-1$, $T_M = Q_M(1-p_c)$, and $P_{\text{drop}} = Q_M p_c$, which results in

$$T_n = (1-p_c)p_c^n, \quad n = 0, 1, \ldots, M \tag{43}$$

$$P_{\text{drop}} = p_c^{M+1} \tag{44}$$

where T_0 is the probability that a packet will be successfully transmitted without retransmission. Since a packet is either successfully transmitted within $(M+1)$ tries or dropped, $\sum_{n=0}^{M} T_n + P_{\text{drop}} = 1$.

Let N_R be a random variable representing the number of retransmissions per packet, then the average number of retransmissions per packet for the successfully transmitted packets is given by

$$\begin{aligned} n_R &= E[N_R | \text{no drop}] \\ &= \sum_{n=0}^{M} n \frac{T_n}{1 - P_{\text{drop}}} = \frac{p_c}{1-p_c} - \frac{(M+1)p_c^{M+1}}{1-p_c^{M+1}}. \end{aligned} \tag{45}$$

Thus, it requires

$$n_R + 1 = \frac{1}{1-p_c} - \frac{(M+1)p_c^{M+1}}{1-p_c^{M+1}} \tag{46}$$

transmissions per packet on average for successful transmission. Note that (46) converges to (26) as M goes to infinity. At first

[1] Equation (14) is valid for both EB and EB-M.

[2] If we assume $rp_c \ge 1$, then (42) converges to zero as M goes to infinity. Since there are only finite number of nodes in the network, this means there is no transmission in the network. Thus the conditional probability of collision p_c converges to zero as M goes to infinity, which contradicts with the assumption. As a result, (17) for EB is defined only for $0 \le p_c < 1/r$.

glance, n_R in (46) seems to have a smaller value than n_R in (26). However, since p_c for EB-M is greater than p_c for EB, (46) has a larger value under the same conditions.

For EB-M, as the number of nodes increases, p_c approaches unity and most packets will be dropped after M transmission retries, and only a few packets will be successfully transmitted within $M+1$ transmission tries. Since (46) represents the average number of required transmission attempts of the successfully transmitted packets, the value is bounded by $M+1$. In fact, it can be shown that

$$\lim_{N\to\infty} n_R + 1 = \frac{M}{2} + 1. \tag{47}$$

The expected medium access delay for EB-M is obtained similarly to EB as

$$\begin{aligned}\overline{D}_\Sigma &= \mathrm{E}\left[\sum_{i=0}^{N_R-1}(D_i+1) + D_{N_R}\middle|\,\text{no drop}\right] \\ &= \mathrm{E}_{N_R}\left[\frac{1}{2}\left(N_R+1+\frac{W_0(1-r^{N_R+1})}{1-r}\right)\middle|\,\text{no drop}\right]-1 \\ &= \frac{1}{2}\left\{\frac{1}{1-p_c} - \frac{(M+1)p_c^{M+1}}{1-p_c^{M+1}} \right. \\ &\qquad \left. + \frac{W_0}{1-r}\left(1-\frac{r(1-p_c)}{1-p_c^{M+1}}\sum_{n=0}^{M}(rp_c)^n\right)\right\}-1 \end{aligned} \tag{48}$$

where $\overline{D}_\Sigma$ is the expected time delay per packet in slots. Note that $\overline{D}_\Sigma$ in (48) and n_R in (45) are defined only for the packets transmitted successfully. But, p_t in (42) is relevant to all packets including the dropped ones. So, a relationship similar to (27) does not exist for EB-M. Equations (47) and (48) are derived in Appendix II.

As discussed above, for EB-M the number of retries is bounded by M, and thus the medium access delay also has an upper bound. This property is well illustrated in Fig. 6(b), where the expected medium access delay is plotted for various values of W_0 along with the corresponding asymptotes, which are calculated in Section III-D. However, to guarantee the upper bound of the medium access delay, EB-M drops packets that fail M retries, and eventually the probability of dropping a packet will converge to unity as N goes to infinity.

D. Asymptotic Behavior of EB With Maximum Retry Limit

From (42), we can see that p_t is always greater than zero for all $p_c \in [0,1]$. As a result, instead of converging to a constant, Np_t goes to infinity as N goes to infinity. Since p_c converges to unity as N goes to infinity [see (18)], from (42),

$$\lim_{p_c\to 1} p_t = \frac{2(M+1)}{W_0\sum_{j=0}^{M} r^j + M + 1} \tag{49}$$

and thus Np_t approaches the following linear function of N as N goes to infinity

$$\frac{2(M+1)}{W_0\sum_{j=0}^{M} r^j + M + 1}N. \tag{50}$$

In Fig. 7(b), the asymptotes are drawn with thin lines in the same dash patterns as the corresponding plots of Np_t.

The divergence of Np_t that we have noted implies that there will be infinitely many nodes transmitting in a slot as N goes to infinity. In fact, since $\lim_{N\to\infty}\alpha^N = 0$ and $\lim_{N\to\infty} N\alpha^N = 0$ for $0<\alpha<1$, P_{busy} and P_{succ} converge to unity and zero, respectively, as N goes to infinity. That is, because of collisions, the channel is always busy and a successful transmission will be achieved with probability zero.

Finally, the delay asymptotes in Fig. 6(b) can be obtained by taking the limit of (48) as follows:

$$\lim_{N\to\infty}\overline{D}_\Sigma = \frac{1}{2}\left\{\frac{M}{2}+1+\frac{rW_0(r^{M+1}-1)}{(M+1)(r-1)^2} - \frac{W_0}{r-1}\right\}-1.$$

IV. Simulation

In Figs. 4–7, simulation results, which are represented by symbols, are added to the curves of analytical results. The simulator is written in the C++ programming language, and simulation results were obtained by running 5 000 000 time slots after 1 000 000 time slots of warming up. The backoff factor $r=2$ was used for the simulations included in the paper.

In the simulation of EB, minimum contention window sizes W_0 =16, 32, and 64 were used. Note that, in the IEEE 802.11 specification, 16, 32, and 64 are used as the minimum contention window sizes for frequency hopping spread spectrum (FHSS), direct sequence spread spectrum (DSSS), and infrared (IR) physical layers, respectively. The simulations were run for $N = 5, 10, \ldots, 50$. The simulation results in Figs. 4(a)–(7a) agree with those obtained from our analysis. However, when the steady state assumption does not hold true, the analysis and simulation produce different results. See Section IV-A for the discussion of the applicability of the analysis model.

In the simulation of EB-M, $M=6$ was used as in the IEEE 802.11 specification. Additional cases with a larger number of nodes (N = 70, 100, 150, 200) were considered for a better comparison of the simulation results with the asymptotes. Furthermore, minimum contention window sizes $W_0=4$ and 8 were also considered in addition to W_0 = 16, 32, and 64, to display that a capture effect does not occur even for small W_0 in the case of EB-M (see Section IV-A). Consequently, the simulation results for EB-M give much better match with the analytical results for a wide range of W_0 and N, as shown in Fig. 4(b)–(7b).

A. Applicability of the Analysis Model

Our analysis model is based on the assumption that the system is in *steady state*. Even well-designed analysis models cease to represent the real system correctly under inordinate operating conditions. Our simulation study shows that our analysis model represents the exponential backoff algorithm accurately over a wide range of operating conditions. However, when W_0 is too small, a capture effect was observed. A capture effect makes only a few nodes consume the whole transmission channel, and results in a higher throughput by making most of the nodes starve (no transmission for a prolonged time period). When $W_0=1$, an absolute capture occurs, that is, a single node captures the channel with probability one, which is observed

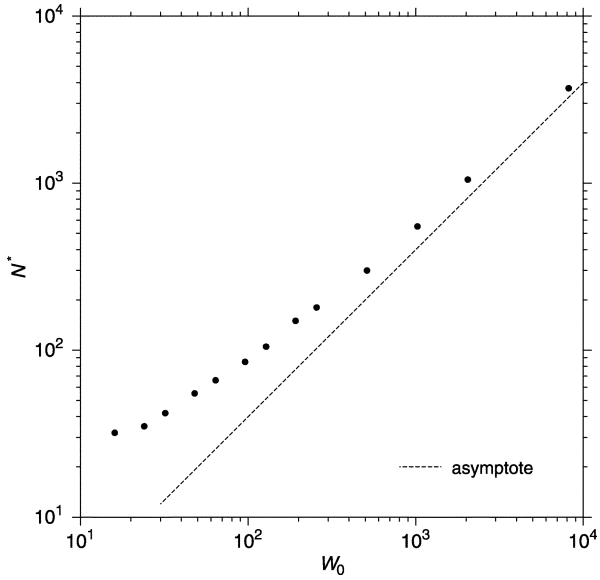


Fig. 8. N^* versus W_0, where N^* is the number of nodes for which the simulation results begin to depart from the analytical results.

both in EB and EB-M. In EB, when W_0 is greater than one but too small, the nodes randomly take turns to capture the channel temporarily. As a result, the network cannot reach steady state and our analysis model does not apply. Also for a large number of nodes N, a system may not reach steady state, depending on the size of the minimum contention window size W_0. For given W_0, when the number of nodes N is too large, a significant number of the nodes stay in the states of extremely large contention window sizes, where the exact distribution of nodes over the states is given by S_i defined in (5) for EB, and in (39) for EB-M. While the nodes in the states of extremely large contention window sizes backoff for prolonged periods, other nodes may have many chances of transmission, causing the network to adapt locally in time excluding the nodes with prolonged backoff. Eventually, the nodes with prolonged backoff will transmit packets and these transmissions will disturb the network, preventing it from reaching a steady state. An investigation of the transmission history of an individual node shows that the nodes experience temporary starvation and capture throughout the simulation when there are discrepancies between the analysis and the simulation. In the case of EB-M, the maximum contention window size is limited by $W_0 r^M$, and thus this problem is less likely to occur, and the steady state assumption holds true for much wider operating conditions. In our simulation, $M = 6$ is used and no noticeable discrepancy between the analysis and simulation was observed. For larger M, however, it is expected that EB-M will also exhibit the problem described above.

Let N^* be the number of nodes for which the simulation results begin to depart from the analytical results. In the case of EB, simulation results with $N > N^*$ show larger p_c compared to the analytical results for the same given W_0. Fig. 8 shows a plot of N^* with respect to W_0 for $r = 2$, where N^* for various W_0 are manually obtained from extensive simulation results. As shown in the figure, N^* increases with W_0, and appears to converge to an asymptote $N^* = C \cdot W_0$, where C is a constant. Finding the exact C remains an open problem, but the simulation study shows that a larger value of r produces a smaller value of C.

V. Conclusion

The contribution of this paper is that we provide a new and efficient analytical means to evaluate the performance of a network with an exponential backoff algorithm. Using the proposed analytical model, we analyze the performance of EB and EB-M to obtain the saturation throughput and the medium access delay. The asymptotic behaviors of EB and EB-M are also shown. To validate the analytical results, the simulation results are provided. The results indicate that EB-M provides the nodes fairer service even for very small values of W_0. EB-M also bounds the medium access delay. These benefits are accomplished by limiting the number of transmission tries for a packet and thus giving a chance of transmission to the next packet waiting. But the maximum retry limit also causes the throughput to diminish as N increases.

The analysis presented in this paper is an analysis of EB and EB-M in steady state (equilibrium). A sudden change of the offered load (the number of contending nodes) will cause change of equilibrium and there will be a transition to the new equilibrium. The dynamic behavior of the backoff algorithm when there is a sudden change of offered load is another interesting problem. Note that the result of the steady state analysis and the dynamic behavior are very closely related because the equilibrium provides the limit that the transient response converges to. In that regard, the steady state analysis may be considered as a prerequisite for an analysis of the dynamic behavior.

Appendix I
Proof of Continuity and Monotonicity of (42)

To show the continuity of (42) in the range [0,1], we need to establish the continuity of p_t at $p_c = 0$. If $p_c = 0$, the nodes will always stay in state $i = 0$, and it can be shown that $S_0 = 1$, from which we obtain

$$p_t|_{p_c=0} = \frac{2}{W_0 + 1}. \tag{51}$$

To establish continuity, we also need to show that p_t converges to (51) as p_c goes to zero

$$\lim_{p_c \to 0} p_t = \lim_{p_c \to 0} \sum_{i=0}^{M} s_{i,0} \tag{52}$$

$$= \lim_{p_c \to 0} \sum_{i=0}^{M} \frac{2p_c^i}{W_0 \sum_{j=0}^{M} (rp_c)^j + \sum_{j=0}^{M} p_c^j} \tag{53}$$

$$= \frac{2}{W_0 + 1}. \tag{54}$$

Thus, p_t is continuous in the range $p_c \in [0, 1]$.

Now, we show that (42) is a monotonically decreasing function by showing that

$$D = \frac{\sum_{j=0}^{M} (rp_c)^j}{\sum_{j=0}^{M} p_c^j} \tag{55}$$

is a monotonically increasing function of p_c in [0,1], which is equivalent to showing $(dD/dp_c) > 0$. In order for the inequality to hold, the following equation needs to be true:

$$\frac{1}{p_c}\sum_{j=1}^{M} j(rp_c)^j \sum_{i=0}^{M} p_c^i > \frac{1}{p_c}\sum_{j=0}^{M}(rp_c)^j \sum_{i=1}^{M} i p_c^i. \tag{56}$$

But

$$\begin{aligned}
&\sum_{j=1}^{M} j(rp_c)^j \sum_{i=0}^{M} p_c^i - \sum_{j=0}^{M}(rp_c)^j \sum_{i=1}^{M} i p_c^i \\
&\quad = \sum_{i=0}^{M}\sum_{j=0}^{M} j r^j p_c^{i+j} - \sum_{i=0}^{M}\sum_{j=0}^{M} i r^j p_c^{i+j} \\
&\quad = \sum_{i=0}^{M}\sum_{j=i+1}^{M} (j-i) r^j p_c^{i+j} - \sum_{i=0}^{M}\sum_{j=i+1}^{M} (j-i) r^i p_c^{i+j} \\
&\quad = \sum_{i=0}^{M}\sum_{j=i+1}^{M} (j-i)(r^j - r^i) p_c^{i+j}.
\end{aligned} \tag{57}$$

Since $j > i$ inside the summation, (57) is always positive. This concludes that p_t in (42) is a continuous and monotonically decreasing function in the range [0,1]. □

APPENDIX II
DERIVATION OF EQUATIONS (27), (47), AND (48)

A. Derivation of Equation (27)

Noting that N_R and D_i, $i = 0, 1, \ldots, N_R$, are both random variables

$$\begin{aligned}
\overline{D}_\Sigma &= E\left[\sum_{i=0}^{N_R-1}(D_i+1) + D_{N_R}\right] \\
&= E_{N_R}\left[E\left[\sum_{i=0}^{N_R-1}(D_i+1) + D_{N_R}\,\middle|\, N_R\right]\right].
\end{aligned}$$

From (4)

$$\begin{aligned}
\overline{D}_\Sigma &= E_{N_R}\left[\sum_{i=0}^{N_R-1}\frac{W_i+1}{2} + \frac{W_{N_R}+1}{2} - 1\right] \\
&= E_{N_R}\left[\sum_{i=0}^{N_R}\frac{W_i+1}{2}\right] - 1 \\
&= \frac{1}{2}E_{N_R}\left[\frac{W_0(1-r^{N_R+1})}{1-r} + (N_R+1)\right] - 1
\end{aligned} \tag{58}$$

where $W_i = W_0 r^i$ is used. Since T_n in (24) is the probability mass function of N_R

$$\begin{aligned}
E[r^{N_R+1}] &= \sum_{n=0}^{\infty} r^{n+1} T_n = \sum_{n=0}^{\infty} r^{n+1}(1-p_c)p_c^n \\
&= r(1-p_c)\frac{1}{1-rp_c}.
\end{aligned} \tag{59}$$

Substituting (59) and $E[N_R] = p_c/(1-p_c)$ [(25)] into (58), and using (17), we have

$$\overline{D}_\Sigma = \frac{1}{2}\left(\frac{1}{1-p_c} + \frac{W_0}{1-rp_c}\right) - 1 = \frac{n_R+1}{p_t} - 1.$$

B. Derivation of Equation (47)

Since $\lim_{N\to\infty} p_c = 1$ (see Section III-D), from (46), we have

$$\begin{aligned}
\lim_{N\to\infty} n_R + 1 &= \lim_{p_c\to 1}\left(\frac{1}{1-p_c} - \frac{(M+1)p_c^{M+1}}{1-p_c^{M+1}}\right) \\
&= \lim_{p_c\to 1}\left(\frac{1-p_c^{M+1}-(M+1)(1-p_c)p_c^{M+1}}{(1-p_c)\left(1-p_c^{M+1}\right)}\right) \\
&= \lim_{p_c\to 1}\left(\frac{1+p_c+\ldots+p_c^M-(M+1)p_c^{M+1}}{1-p_c^{M+1}}\right).
\end{aligned}$$

Using L'Hospital's rule

$$\begin{aligned}
\lim_{N\to\infty} n_R+1 &= \lim_{p_c\to 1}\left(\frac{1+2p_c+\ldots+Mp_c^{M-1}-(M+1)^2p_c^M}{-(M+1)p_c^M}\right) \\
&= \frac{(M+1)^2-(1+2+\ldots+M)}{M+1} \\
&= \frac{M}{2}+1.
\end{aligned}$$

C. Derivation of Equation (48)

The derivation of (48) is similar to the derivation of (27) in Appendix II-A. Since N_R and D_i, $i = 0, 1, \ldots, N_R$, are both random variables, as in the case of (27)

$$\begin{aligned}
\overline{D}_\Sigma &= \mathrm{E}\left[\sum_{i=0}^{N_R-1}(D_i+1) + D_{N_R}\,\middle|\,\text{no drop}\right] \\
&= \frac{1}{2}\mathrm{E}_{N_R}\left[\frac{W_0(1-r^{N_R+1})}{1-r} + (N_R+1)\,\middle|\,\text{no drop}\right] - 1.
\end{aligned} \tag{60}$$

From (43) and (44)

$$\begin{aligned}
E[r^{N_R+1}|\text{no drop}] &= \sum_{n=0}^{M} r^{n+1}\frac{T_n}{1-P_{\text{drop}}} \\
&= \sum_{n=0}^{M} r^{n+1}\frac{(1-p_c)p_c^n}{1-p_c^{M+1}} \\
&= \frac{(1-p_c)r}{1-p_c^{M+1}}\sum_{n=0}^{M}(rp_c)^n.
\end{aligned} \tag{61}$$

Substituting (61) and (45) into (60), we have

$$\begin{aligned}
\overline{D}_\Sigma = \frac{1}{2}\Bigg\{&\frac{1}{1-p_c} - \frac{(M+1)p_c^{M+1}}{1-p_c^{M+1}} \\
&+ \frac{W_0}{1-r}\left(1 - \frac{r(1-p_c)}{1-p_c^{M+1}}\sum_{n=0}^{M}(rp_c)^n\right)\Bigg\} - 1.
\end{aligned}$$

REFERENCES

[1] R. M. Metcalfe and D. R. Boggs, "Ethernet: distributed packet switching for local computer networks," *Commun. ACM*, vol. 19, no. 7, pp. 395–404, Jul. 1976.

[2] *P802.11, IEEE Standard for Wireless Lan Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, Nov. 1997.

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

[4] 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.

[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. Comput.*, vol. 25, no. 4, pp. 740–744, 1996.

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

[8] K. K. Ramakrishnan and H. Yang, "The ethernet capture effect: analysis and solution," in *Proc. 19th Conf. Local Computer Networks*, 1994, pp. 228–240.

[9] D. G. Jeong and W. S. Jeon, "Performance of an exponential backoff scheme for slotted-ALOHA protocol in local wireless environment," *IEEE Trans. Veh. Technol.*, vol. 44, no. 3, pp. 470–479, Aug. 1995.

[10] 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. Veh. Technol.*, vol. 49, no. 4, pp. 1449–1453, Jul. 2000.

[11] 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. Wireless Personal Multimedia Communications*, Aalborg, Denmark, Sep. 2001.

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

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

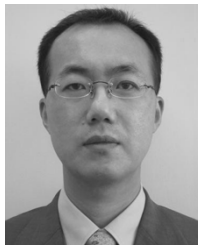
[14] H. Al-Ammal, L. A. Goldberg, and P. MacKenzie, "Binary exponential backoff is stable for high arrival rates," in *Proc. 17th Int. Symp. Theoretical Aspects of Computer Science*, Lille, France, Feb. 2000.

[15] ——, "An improved stability bound for binary exponential backoff," *Theory Comput. Syst.*, vol. 30, pp. 229–244, 2001.

[16] L. A. Goldberg and P. D. MacKenzie, "Analysis of practical backoff protocols for contention resolution with multiple servers," *J. Comput. Syst. Sci.*, vol. 58, no. 1, pp. 232–258, 1999.

[17] H. Wu, Y. Pang, K. Long, S. Cheng, and J. Ma, "Performance of reliable transport protocol over IEEE 802.11 wireless lan: analysis and enhancement," in *Proc. IEEE INFOCOM*, vol. 2, Jun. 2002, pp. 599–607.

[18] *P802.3, Carrier Sense Multiple Access With Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications*, Mar. 1991.

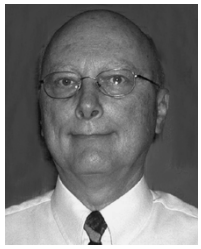


Byung-Jae Kwak received the B.S and M.S. degrees in electronic engineering from Yonsei University, Seoul, Korea, in 1989 and 1991, respectively, and the Ph.D. degree in electrical engineering and computer science from the University of Michigan, Ann Arbor, in 2000.

From 1991 to 1992, he was with the Engineering Research Institute, Yonsei University, where he was involved in the development of sonar systems. In 2000, he joined the Telecommunication R&D Center, Samsung Electronics, Korea, where he participated in the standardization effort of third-generation mobile communication systems (3GPP). From 2001 to 2004, he was with the Advanced Network Technologies Division at the National Institute of Standards and Technology, Gaitherburg, MD, as a visiting scholar. Since 2004, he has been with the Digital Home Research Division at the Electronics and Telecommunications Research Institute, Korea. His research interests include wireless communications, mobile ad hoc networks, distributed MAC protocols, and adaptive signal processing.

Nah-Oak Song (M'03) received the B.S. and M.S. degrees from Yonsei University, Seoul, Korea, in 1989 and 1993, respectively, and the Ph.D. degree from the University of Michigan, Ann Arbor, in 1999.

From 1989 to 1991, she was with the Application Specific Integrated Circuit Research Institute, Yonsei University, where she was involved in a project on echo cancellation systems. From 1999 to 2001, she was with the Telecommunication R&D Center, Samsung Electronics, Korea, and participated in the development of CDMA 2000 system. From 2001 to 2004, she was with the Advanced Network Technologies Division at the National Institute of Standards and Technology, Gaitherburg, MD, as a visiting scholar. Since 2004, she has been with the Digital Home Research Division at the Electronics and Telecommunications Research Institute, Korea. Her main interest is in wireless networks with emphasis on mobile ad hoc networks and wireless LAN. Her other interests include medium access control protocol, quality of service, stochastic scheduling, and MPLS.



Leonard E. Miller (S'63–M'64–SM'92) received the B.E.E. degree from Rensselaer Polytechnic Institute, Troy, NY, in 1964, the M.S.E.E. degree from Purdue University, West Lafayette, IN, in 1966, and the Ph.D. degree from The Catholic University of America, Washington, DC, in 1973.

From 1964 to 1978, he was with the Naval Surface Warfare Center, Silver Spring, MD, where he was a member of the Signal Processing Branch. From 1978 to 2000, he was with J. S. Lee Associates, Inc., Rockville, MD, as Vice President for Research, and was involved initially in analyzing the survivability and performance of military communications and electronic support systems, and later in modeling of propagation in the mobile environment and the design and analysis of cellular and personal communication systems. Since 2000, he has been a member of the Wireless Communication Technologies Group, National Institute of Standards and Technology, Gaithersburg, MD, where he is responsible for analysis and simulation of wireless ad hoc networks, wireless standards, and public safety wireless applications. He is a coauthor (with Dr. J. S. Lee) of *CDMA Systems Engineering Handbook* (Boston, MA: Artech House, 1998).