Design and Analysis of Backoff Algorithms for Random Access Channels in UMTS-LTE and IEEE 802.16 Systems

Jun-Bae Seo, *Member, IEEE*, and Victor C. M. Leung, *Fellow, IEEE*

***Abstract*—In this paper, we examine the performance of uniform backoff (UB) and binary exponential backoff (BEB) algorithms with retry limit, which can be used in the random-access channels of Universal Mobile Telecommunication System (UMTS)-Long Term Evolution (LTE) and IEEE 802.16 systems under the assumption of finite population under unsaturated traffic conditions. Additionally, we consider access prioritization schemes to provide differential performance by controlling various system parameters. We show that controlling the persistence value as specified in UMTS is effective in both backoff algorithms. The performances with and without access prioritization schemes are presented in terms of throughput, mean, and variance of packet retransmission delay, packet-dropping probability, and system stability. Finally, we consider a dynamic window assignment algorithm that is based on Bayesian broadcasting, in which the base station adaptively controls the window size of the UB algorithm under unsaturated traffic conditions. Results show that the proposed window assignment algorithm outperforms fixed window assignment in static and dynamic traffic conditions under the assumption of perfect orthogonality between random-access codes.**

***Index Terms*—Access protocol, algorithm design and analysis, cellular networks, communication system signaling, multiaccess communication.**

I. Introduction

A CANDIDATE for fourth-generation (4G) mobile networks, Universal Mobile Telecommunication System (UMTS)-Long Term Evolution (LTE) has been standardized to meet an ever-increasing demand on higher wireless access data rate. In parallel, IEEE 802.16 Worldwide Interoperability for Microwave Access (WiMAX) systems have emerged as a competing technology, offering a broadband wireless access alternative to "last-mile" wired access. Based on orthogonal frequency-division multiplexing (OFDM), both systems provide high-data-rate wireless services and have similar advanced technologies such as hybrid automatic-repeat request, adaptive modulation and coding, and multiple-input–multiple-output antennas. In particular, the random-access channels of these radio access networks employ similar protocols, with the exception that the power-ramping scheme[1] is used in UMTS-LTE. In IEEE 802.16 systems, terminals perform initial and periodic ranging operations using a random access code to enable the transmission power and timing of channel access to be adjusted and synchronized [1]. In UMTS-LTE, terminals perform this function using the power-ramping scheme in conjunction with contention-based random access [2]. To obtain an uplink data transmission channel, terminals in both systems transmit a random access code by contention in the random-access channel and back off their retransmissions if the contention results in a collision. After successful random access, they can transmit uplink bandwidth request messages using the uplink resource assigned by the base station (BS) in the random-access response message. Thus, an efficient collision resolution algorithm leads to fast system access.

The collision resolution algorithm standardized in both systems employ a backoff algorithm with retry limit: binary exponential backoff (BEB) in the case of the IEEE 802.16 system [1] and uniform backoff (UB) in the case of UMTS-LTE [2]. Aside from retry limits, each standard supports access-level priority to provide quality-of-service differentiation in the medium access control (MAC) layer among services such as security service, public utilities, and emergency service, e.g., rescue in a congested area [11]. In addition to prioritization in access control, access priority can be further enhanced by differentiating some parameters of those backoff algorithms. In this paper, we provide a comparative study on these backoff algorithms with retry limit and access priority differentiation.

Previous work on backoff algorithms with retry limit can be categorized based on the design of the backoff algorithm, i.e., whether window based or probability distribution based, and the modeling assumptions, i.e., finite or infinite population, and saturated or unsaturated traffic conditions. Kim [3] considered a frequency-hopped code-division multiple access (CDMA) system under an infinite population model, in which uniform retransmission intervals are assumed. It has been shown that system stability can be achieved by controlling the retry limit and the channel coding rate that affects the transmission success rate of a packet. Sakakibara *et al.* examined the exponential backoff (EB) algorithm with retry limit in slotted (S)-ALOHA

Manuscript received November 27, 2010; revised June 27, 2011; accepted August 18, 2011. Date of publication August 30, 2011; date of current version October 20, 2011. This work was supported in part by the Canadian Natural Sciences and Engineering Research Council through Grant RGPIN 44286 and a postgraduate scholarship. This paper was presented in part at the 2010 IEEE International Symposium on Personal, Indoor, and Mobile Radio Communications, Istanbul, Turkey, September 26–29. The review of this paper was coordinated by Prof. P. Langendoerfer.

The authors are with the Department of Electrical and Computer Engineering, University of British Columbia, Vancouver, BC V6T 1Z4, Canada (e-mail: jbseo@ece.ubc.ca; vleung@ece.ubc.ca).

Digital Object Identifier 10.1109/TVT.2011.2166569

[1]In UMTS-LTE random access, whenever retransmission is not successful, the transmission power is increased by a constant step, which is called power ramping.

systems [4] and frequency-hopped S-ALOHA systems [5], respectively. Additionally, Liu [6] examined an EB algorithm in frequency-hopped S-ALOHA systems under a finite population with unsaturated traffic (FP/UST) model. The EB algorithm in [4]–[6], which retransmits a packet with probability $r_i (\leq r_{i-1})$ after it has been unsuccessfully transmitted $i - 1$ times, is an approximation of the widely used window-based BEB backoff algorithm. While system bistability has been extensively examined according to the retry limit in [4]–[6], packet delay and packet-dropping probability have not been evaluated. In [7], Kwak *et al.* considered the retry limit of window-based BEB in S-ALOHA under a finite population with saturated traffic (FP/ST) model. Similarly, Falla *et al.* [8], Vu *et al.* [9], and Chuck *et al.* [10] examined window-based BEB with retry limit for an IEEE 802.16 system under FP/ST. It should be noted that the S-ALOHA system under the FP/ST model always operates in a stable but inefficient region [4], [5]. In contrast to [7]–[10] with the FP/ST model, in this paper, we examine the performance of UB and BEB algorithms with retry limit under a more realistic FP/UST model and discuss their stability.

There are a few studies on access prioritization schemes for random-access channels. Xiao [12] examined not only the effect of the retry limit of the BEB algorithm in IEEE 802.11 system[2] but considered prioritized access schemes by differentiating the contention window size of the BEB algorithm under FP/ST as well. Additionally, Rivero-Angeles *et al.* [13] considered a prioritization scheme in the S-ALOHA system by differentiating the retransmission probability of high- and low-priority terminals. In comparison with [12] and [13], we consider the performance of UB and BEB algorithms with access priority and their stability under an FP/UST model. Our comparative study of UB and BEB algorithms could provide a comprehensive understanding on the random access in UMTS-LTE and IEEE 802.16 systems.

The contributions of this paper are threefold. First, we examine UB and BEB algorithms with retry limit used in the random-access channels of UMTS-LTE and IEEE 802.16 system under the FP/UST model for nonprioritized access. Second, we consider some prioritized access schemes by differentiating the persistence value [11] and other parameters of the UB and BEB algorithms. Third, we extend an existing Bayesian broadcasting algorithm [14] into a dynamic window assignment (DWA) scheme in the UB algorithm, which is suitable for application in UMTS-LTE. In this paper, we ignore the physical characteristics of the OFDM-CDMA channel to focus on the backoff algorithms in the MAC layer. The main results are summarized as given here.

1) Under an FP/UST model, we obtain the stability conditions of the system with UB and BEB backoff algorithms with/without retry limit. In contrast to the previous work [4]–[6] showing that some approximate backoff algorithms exhibit bistability, our analysis shows that UB and BEB algorithms without retry limit do not have such a bistability characteristic. Instead, these backoff algorithms operate in either a stable or unstable manner. Thus, instead of controlling the retry limit to maintain system stability as in [4]–[6], we shall see that the retry limit provides a tradeoff between throughput and delay performance at the expense of the packet-dropping probability.
2) In contrast to the extensive studies on BEB algorithms in the previous work [7]–[12], we compare the performance of the UB algorithm standardized in UMTS-LTE [2] against that of the BEB algorithm standardized in WiMAX [1], in terms of throughput, mean, and variance of delay, and packet-dropping probability, given the retry limit. We show that UB can be designed to have a similar performance as BEB. It is also shown that UB achieves much smaller variances of the total access delay than BEB. Note that the variance of the access delay is an important indicator of fair use of the channel by all terminals over time i.e., a backoff algorithm showing a small variance in retransmission delay will support more even channel access by the terminals.
3) We examine some access prioritization schemes for the UB and BEB algorithms by differentiating the persistence probability as specified in [2] and the window size. It is shown that the persistence value effectively provides access prioritization, particularly when the system is overloaded.
4) In [2], UMTS-LTE allows the window size of the UB algorithm to be dynamically changed by the BS according to network congestion.[3] Our DWA scheme for the UB algorithm outperforms the existing mechanism in the random access channel of UMTS-LTE for static and time-varying traffic conditions, provided that orthogonality of random access preambles (RAPs) is preserved.

The rest of this paper is organized as follows: Section II introduces UB and BEB algorithms with retry limit and the corresponding access prioritization schemes, as used in UMTS-LTE and IEEE 802.16 systems, respectively. We analyze the performance of these backoff algorithms in terms of throughput, mean, and variance of delay, and packet-dropping probability; evaluate system stability; and access differentiation in Section III. Additionally, we propose a DWA algorithm for the UB algorithm in Section III. Numerical examples are presented in Section IV. Concluding remarks are given in Section V.

II. Backoff Algorithms

A. System Model

We first introduce the random-access channel and the random-access procedure in UMTS-LTE primarily, to which our system model has close resemblances. As previously mentioned, the physical random access channels of UMTS-LTE and IEEE 802.16 system are equivalent to an OFDM-CDMA channel, in which multipacket reception is possible in the presence of multiple-access interference (MAI). In UMTS-LTE,

[2]The BEB algorithm in IEEE 802.11 system decrements its backoff counter during idle channel, whereas the UB and BEB algorithms in wireless cellular networks [2]–[11] decrement their backoff counters, regardless of channel status.

[3]In the UB algorithm of UMTS-LTE [2], 13 different window sizes from 0 and 960 ms are specified, which can be dynamically selected.

there can be F random access channels in the frequency domain within one access period, among which an accessing terminal may randomly select one. At the same time, a terminal picks one of P RAPs and transmits it in the selected random access channel. If the same RAP is transmitted by more than one terminal over the same random access channel, a code collision occurs, or if the transmitted RAP is not acknowledged by the BS within a specified time window, e.g., the random access response (RAR) window in UMTS-LTE, it is also declared as a collision. This latter condition may occur due to the transmitted RAP not being successfully decoded under severe MAI. Whenever the terminals' RAP transmissions are not successful, they retransmit after a delay determined by the UB algorithm (in UMTS-LTE) or BEB algorithm (in an IEEE 802.16 system). These algorithms will be explained in the next sections. In both systems, if the number of retransmissions exceeds the maximum retry limit, the RAP is dropped in the MAC layer. This is reported to a higher layer. Once the RAP is successfully transmitted, the terminal can transmit a bandwidth request message for bandwidth reservation on the uplink channel that the BS has granted to it via the RAR message in UMTS-LTE or the uplink map (UL-MAP) in the IEEE 802.16 system. This procedure is triggered at the time when an end-user makes a phone call or uses an application that needs to access the Internet. Accordingly, fast resolution of collisions in RAP transmissions is expected to enable fast bandwidth reservations. Additionally, in UMTS-LTE, the window size of the UB algorithm, which is denoted by U hereafter and is assigned by the BS in the first RAR message, can be dynamically chosen according to network congestion. Thus, it is expected that appropriately controlling U can efficiently resolve collisions under some time-varying traffic condition.

In what follows, we assume that the transmission power and timing of RAPs are perfectly known to terminals, which simplifies the power-ramping scheme in UMTS-LTE. This is equivalent to the assumption that initial and periodic ranging processes are sufficiently well performed in an IEEE 802.16 system. We further assume that perfect orthogonality among RAPs simultaneously transmitted by multiple terminals is preserved. Owing to this perfect orthogonality assumption, a system with P RAPs and F random access channels is equivalent to a system with $P \cdot F$ RAPs. Note that, in UMTS-LTE, each BS provides 64 RAPs, some of which can be reserved for the use of noncontention-based access. Additionally, in the frame structure of type-1 frequency-division duplex (FDD) mode, one random access channel in the frequency domain appears every two slots [15]. For analytical simplicity, we set $F = 1$ every slot and assume a slotted system with FDD, in which each slot corresponds to one RAP transmission (one packet) time. The number of RAPs will be given later in Section V. Hereafter, we interchangeably use the terms RAP and packet. In the system model, we assume a finite population of M terminals. Each terminal can hold only one packet for transmission or retransmission until it is successfully received. A terminal that has no packet to transmit is said to be idle. An idle terminal generates a packet in every slot based on a Bernoulli trial with probability ϵ, i.e., unsaturated traffic condition. When a terminal has a packet to send, it immediately becomes backlogged by performing the backoff algorithm, i.e., it employs a delayed-first-transmission protocol. We finally assume that terminals can get feedback information for the RAP transmitted, e.g., success or failure, from the BS right before the next slot with no error, i.e., no feedback delay is taken into account.

B. *Uniform Backoff Algorithm in UMTS-LTE*

Each backlogged terminal randomly draws a backoff counter between 0 and $U - 1$. The backoff counter is then decremented by one at every slot, until it reaches zero. When the backoff counter is zero, the terminal selects a random real number between 0 and 1. As in [11], if this number is less than the persistence probability (or persistence value) p, the terminal transmits its packet; otherwise, it postpones the packet transmission by repeating the preceding procedure in the next slot, until the packet is transmitted. When the packet is not successfully transmitted, i.e., the transmission encounters a collision, the terminal delays its retransmission again by a random time newly chosen between 0 and $U - 1$. This trial is repeated up to L times more, until the retransmission succeeds, or the terminal drops the packet and goes back to idle if the $L + 1$th retransmission remains unsuccessful.

C. *BEB Algorithm in IEEE 802.16*

Upon an unsuccessful packet transmission, the BEB algorithm exponentially increases the backoff window by doubling its size. Let W_i denote the backoff window size of the ith backoff stage, which is determined by

$$W_i = 2^i \mathcal{W}_0 \quad \text{for} \quad 0 \leq i \leq K \tag{1}$$

where $\mathcal{W}_0$ is the minimum backoff window size. At the ith backoff stage, the backlogged terminal randomly chooses a backoff counter in the range $[0, W_i - 1]$ for $0 \leq i \leq K$. Once chosen, the backoff counter is decremented every slot as in the UB algorithm. When it reaches zero, the terminal transmits its packet. Upon transmission failure, the terminal increments the backoff stage by one and chooses the next backoff counter again. Although the retry limit L is independent of K in [1], we assume here that the BEB algorithm retries L times additionally at the Kth stage. Thus, upon the retransmission failure after a total of $L + 1$ retransmissions at the Kth stage, it drops the packet. One can find more details of the random access procedure for IEEE 802.16 in [16].

D. *Access Prioritization by Persistence Value*

To provide access priority, UMTS includes eight access priority classes [11], each of which has a different persistence value $p = P_i(N) = s_i \cdot 2^{-(N-1)}$ for $0 \leq i \leq 7$. Note that the value of i can indicate a certain partition of the random access channels. The persistence scaling factor s_i is assumed to be a value between 0.2 and 0.9 by step of 0.1, and the dynamic persistence level N from 1 to 8 is regularly broadcast. If the persistence scaling factors are not broadcast, a default value of one is assumed. Once the backoff counter reaches zero, the

TABLE I
NOMENCLATURE FOR ANALYSIS

Notation	Definition
P	Number of orthogonal random access codes (preambles)
M	Total number of terminals
L	Retry limit of retransmissions
U	Window size of uniform backoff algorithm
K	Maximum stage of BEB algorithm, maximum window $W_K = 2^K - 1$
ϵ	Probability that a packet is generated at an idle terminal every slot, i.e. Bernoulli trial with probability ϵ
p_s	Transmission success probability of a packet
p_r	Retransmission probability of a packet
p_c	Code-collision probability of a packet, $p_c = 1 - p_s$
p	Persistence probability (value)

terminal draws a real random number between 0 and 1. If this number is less than $P_i(N)$, the terminal transmits a packet. Otherwise, it repeats this test until transmission. Note that two successive tests are separated by a certain time period called the T_2 timer. We apply this access prioritization scheme to both UB and BEB algorithms to facilitate comparisons.

III. ANALYSIS

Before proceeding with our analysis, we summarize in Table I the notations previously defined, which will be used in this section.

A. Channel Model

Throughout this paper, we denote the binomial probability distribution function with parameters b, L, and p by $\Omega(b, L, p) = \binom{L}{b}(p)^b(1-p)^{L-b}$. Define as $\Lambda(k|v, P)$ the probability that k among v RAPs simultaneously transmitted are uniquely chosen among a total of P RAPs, i.e., successful transmission probability of k out of v packets. Recall that backlogged terminals select one of P RAPs and transmit it. Then, we can recursively obtain $\Lambda(k|v, P)$ by [18]

$$\Lambda(k|v,P) = \sum_{m=0}^{v} \Omega(m, v, 1/P)\Lambda(k - \mathcal{I}(m-1)|v-m, P-1) \tag{2}$$

in which the indicator function $\mathcal{I}(x)$ equals 1 at $x = 0$ and 0 elsewhere. Equation (2) indicates that, by examining each of the v packets transmitted in order, if a unique RAP is transmitted with probability $\Omega(1, v, P)$, it is counted as one success, and we decrement k, v, and P by one. Otherwise, we subtract from v a group of m duplicated RAPs, which occurs with probability $\Omega(m, v, P)$ for $m \neq 1$. Equation (2) has the following initial conditions:

$$\Lambda(0|1,0) = 1, \quad \Lambda(0|1,P) = 0 \text{ for } P \geq 1, \quad \Lambda(1|1,0) = 0$$
$$\Lambda(1|1,P) = 1 \text{ for } P \geq 1, \quad \Lambda(1|v,1) = 0 \text{ for } v \neq 1$$
$$\Lambda(k|v,P) = 0 \text{ for } \min(v,P) < k$$
$$\Lambda(-1|v,P) = 0 \text{ for any } v \text{ and } P.$$

For $P = 1$, the channel is reduced to an S-ALOHA channel. Let $C_{k,n}$ denote the probability that k packets are successfully received among n simultaneous packet transmissions for $k \leq n$, i.e., there are $n - k$ code-collisions. Suppose that a terminal among M terminals (re)transmits a packet with probability one. Then, its transmission success probability p_s can be written as

$$p_s = \sum_{n=0}^{M-1} \frac{1}{n+1}\left[\sum_{k=0}^{n}(k+1)C_{k+1,n+1}\right]\Omega(n, M-1, p_r) \tag{3}$$

in which the term in the bracket indicates the conditional throughput, given $n+1$ packets transmitted (including the packet of our interest). By dividing this term by $n+1$, we get the conditional transmission success probability, given $n+1$ packets. For the code collision under perfect orthogonality condition among RAPs, we have $C_{k+1,n+1} = \Lambda(k+1|n+1, P)$. Other channel models can be analyzed by replacing $C_{k+1,n+1}$ accordingly.

B. Uniform Backoff Algorithm

Define π_E and $\pi_{i,j}$ for $0 \leq i \leq L$ and $0 \leq j \leq U-1$ as the stationary state probability that a terminal is idle and that a terminal is at the jth backoff counter value of the ith backoff stage, respectively. Then, we can write the state balance equation for π_E as

$$\pi_E = (1-\epsilon)\pi_E + p \cdot p_s \sum_{i=0}^{L-1} \pi_{i,0} + p\pi_{L,0}. \tag{4}$$

We can write $\pi_{0,j}$ for $0 \leq j \leq U-1$ as

$$\begin{aligned} \pi_{0,0} &= (1-p)\pi_{0,0} + \frac{\epsilon}{U}\pi_E + \pi_{0,1} \\ \pi_{0,j} &= \frac{\epsilon}{U}\pi_E + \pi_{0,j+1} \text{ for } 1 \leq j \leq U-2, \quad \text{and} \\ \pi_{0,U-1} &= \frac{\epsilon}{U}\pi_E. \end{aligned} \tag{5}$$

We can also write $\pi_{i,j}$ for $1 \leq i \leq L$ and $0 \leq j \leq U-1$ as

$$\begin{aligned} \pi_{i,0} &= (1-p)\pi_{i,0} + \frac{pp_c}{U}\pi_{i-1,0} + \pi_{i,j+1} \\ \pi_{i,j} &= \frac{pp_c}{U}\pi_{i-1,0} + \pi_{i,j+1} \text{ for } 1 \leq j \leq U-2, \quad \text{and} \\ \pi_{i,U-1} &= \frac{pp_c}{U}\pi_{i-1,0}. \end{aligned} \tag{6}$$

Using (5) and (6), we rewrite $\pi_{i,j}$ for $0 \leq i \leq L$ as

$$\begin{aligned} \pi_{i,0} &= p_c^i \frac{\epsilon}{p}\pi_E \quad \text{and} \\ \pi_{i,j} &= p_c^i \frac{U-j}{U} \cdot \epsilon\pi_E \text{ for } 1 \leq j \leq U-1. \end{aligned} \tag{7}$$

Using the normalization condition $\pi_E + \sum_{i=0}^{L}\sum_{j=0}^{U-1} \pi_{i,j} = 1$, we can get π_E as

$$\pi_E = \left[1 + \epsilon\left(\frac{1}{p} + \frac{U-1}{2}\right)\frac{1-p_c^{L+1}}{1-p_c}\right]^{-1}. \tag{8}$$

Then, retransmission probability p_r can be expressed as

$$p_r = p\sum_{i=0}^{L}\pi_{i,0} = \frac{1-p_c^{L+1}}{p_s}\epsilon\pi_E. \tag{9}$$

When $L \to \infty$ in (8) and (9), we have the case without retry limit.

To find the solution of p_r simultaneously satisfying (3) and (9), we construct a function $f_U(p_r)$ with (3) and (9) as

$$f_U(p_r) = p_r - \frac{1-(1-p_s)^{L+1}}{p_s}\epsilon\pi_E. \tag{10}$$

Note that (10) is a function of p_r, when we substitute (3) into (10) for p_s. The solution of p_r satisfying $f_U(p_r) = 0$ is numerically found for $0 < p_r < 1$. If a unique p_r is found, we have a stable system. Otherwise, the system is said to be either bistable or unstable. We shall discuss this in Section III-D.

As a performance metric, the system throughput τ in packets per slot can be obtained as

$$\tau = \sum_{n=0}^{M}\sum_{k=0}^{n} k\Lambda(k|n,P)\Omega(n,M,p_r). \tag{11}$$

We then obtain the throughput for this channel, which was normalized with respect to the total number of codes, by $\tau_N = \tau/P$ (packet/slot/code). The packet-dropping probability P_d, due to reaching the retry limit, can be expressed as

$$P_d = p_c p\pi_{L,0} = p_c^{L+1}\epsilon\pi_E \tag{12}$$

where p_c is the code-collision probability. Note that P_d goes to zero as $L \to \infty$.

To obtain the mean packet transmission delay, i.e., access delay, some previous work [7]–[9] used the mean-value analysis. Instead, we use an absorbing Markov chain technique. First, we define the packet retransmission delay $D_{i,j}$ as the time for an absorbing process starting from the initial backoff counter value j at the ith backoff stage to reach (i.e., be absorbed into) the idle state. As usual, we denote the expectation of a random variable by an overline; thus, $\overline{D}_{i,j} = E[D_{i,j}]$. Note that, in the ith backoff stage for $0 \le i \le L$, the initial backoff counter value j is uniformly chosen from $0 \le j \le U-1$. Define the mean access delay $\overline{d}$ as the mean time taken for a terminal to successfully transmit a packet, after it is generated. We can write $\overline{d}$ as

$$\overline{d} = \frac{1}{U}\sum_{j=0}^{U-1}\overline{D}_{0,j} = \overline{D}_{0,0} + \frac{U-1}{2} \tag{13}$$

in which we have used the following relation:

$$\overline{D}_{i,j} = E[1+D_{i,j-1}] = j + \overline{D}_{i,0} \tag{14}$$

for $0 \le i \le L$ and $1 \le j \le U-1$. The first equality in (14) indicates that it always takes one slot for the absorbing process to go down by one to the next backoff counter value with probability one. Since the higher layer of a terminal will be informed upon packet dropping, we include the delay of the packets successfully transmitted and dropped in (13). We can write $\overline{D}_{i,0}$ for $0 \le i \le L-1$ as

$$\begin{aligned}\overline{D}_{i,0} &= pp_s + (1-p)E\left[1+D_{i,0}\right] + \frac{pp_c}{U}E\left[\sum_{j=0}^{U-1}(1+D_{I+1,j})\right]\\ &= \frac{1}{p} + p_c\left(\overline{D}_{i+1,0} + \frac{U-1}{2}\right)\\ &= \left(\frac{1}{p} + p_c\frac{U-1}{2}\right)\frac{1-p_c^{L-i}}{1-p_c} + p_c^{L-i}\end{aligned} \tag{15}$$

which indicates that the retransmission delay takes one slot with probability $p \times p_s$ when the backoff counter is zero. Furthermore, when the persistence test is not passed with probability $1-p$, it takes one more slot before the next persistence test. When a packet is retransmitted but unsuccessful, the retransmission delay takes one slot and additional delay, depending on the next backoff counter. We have

$$\overline{D}_{L,0} = 1\cdot p + (1-p)\left(1+\overline{D}_{L,0}\right) \tag{16}$$

which is simplified as $\overline{D}_{L,0} = 1/p$. Using (14) and (16), one can get $\overline{D}_{i,0}$ for $0 \le i \le L-1$ by

$$\overline{D}_{i,0} = \frac{1}{1-p_c}\left(\frac{1}{p}\left(1-p_c^{L-i+1}\right) + p_c\frac{U-1}{2}\left(1-p_c^{L-i}\right)\right). \tag{17}$$

To obtain the access delay variance, let $\overline{d^2}$ denote the second moment of the access delay, which is expressed by

$$\begin{aligned}\overline{d^2} &= \frac{1}{U}E\left[\sum_{j=0}^{U-1}D_{0,j}^2\right] = \frac{1}{U}\sum_{j=0}^{U-1}\overline{D^2}_{0,j}\\ &= \overline{D^2}_{0,0} + (U-1)\overline{D}_{0,0} + \frac{2U^2+3U-5}{6}\end{aligned} \tag{18}$$

where we have used

$$\begin{aligned}\overline{D^2}_{i,j} &= E\left[(1+D_{i,j-1})^2\right] = 1 + 2\overline{D}_{i,j-1} + \overline{D^2}_{i,j-1}\\ &= j + 2\sum_{k=0}^{j-1}\overline{D}_{i,k} + \overline{D^2}_{i,0}\\ &= 2j\overline{D}_{i,0} + j^2 + \overline{D^2}_{i,0}\end{aligned} \tag{19}$$

for $0 \le i \le L$ and $1 \le j \le U-1$. In the preceding equation, we can write $\overline{D^2}_{i,0}$ for $0 \le i \le L-1$ as

$$\begin{aligned}\overline{D^2}_{i,0} &= pp_s + (1-p)E\left[(1+D_{i,0})^2\right]\\ &\quad + \frac{pp_c}{U}E\left[\sum_{j=0}^{U-1}(1+D_{i+1,j})^2\right]\\ &= pp_s + (1-p)\left(1+2\overline{D}_{i,0}+\overline{D^2}_{i,0}\right)\\ &\quad + \frac{pp_c}{U}\left(U + 2\sum_{j=0}^{U-1}\overline{D}_{i+1,j} + \sum_{j=0}^{U-1}\overline{D^2}_{i+1,j}\right)\end{aligned} \tag{20}$$

which is the second moment of (15). By substituting (14) and (19) into (20), we can rewrite

$$\overline{D^2}_{i,0} = \frac{1}{p}\left[\left(1 + pp_c\frac{2U^2+3U-5}{6}\right) + 2(1-p)\overline{D}_{i,0} + pp_c\left((U+1)\overline{D}_{i+1,0} + \overline{D^2}_{i+1,0}\right)\right]. \quad (21)$$

Finally, we have

$$\overline{D^2}_{L,0} = 1^2 \cdot p + (1-p)E\left[(1+D_{L,0})^2\right] \quad (22)$$

which is simplified as $\overline{D^2}_{L,0} = 1/p(1+2(1-p)\overline{D}_{L,0})$. The variance of the packet access delay is obtained by $\overline{d}_v = \overline{D^2} - \overline{D}^2$.

C. BEB Algorithm

Let $\hat{\pi}_E$ and $\hat{\pi}_{i,j}$ denote the stationary state probability that a terminal has no packet and that a terminal is at the jth backoff time of the ith backoff stage, respectively, in the BEB algorithm. We can write a balance equation for $\hat{\pi}_E$ as

$$\hat{\pi}_E = (1-\epsilon)\hat{\pi}_E + pp_s\sum_{i=0}^{K-1}\hat{\pi}_{0,i} + \hat{\pi}_{0,K}. \quad (23)$$

Substituting W_0 for U in (5), we can get the balance equations for $i = 0$. We express $\hat{\pi}_{i,j}$ for $0 \le i \le K-1$ and $0 \le j \le W_i - 1$ as

$$\begin{aligned}\hat{\pi}_{i,0} &= (1-p)\hat{\pi}_{i,0} + \frac{pp_c}{W_i}\hat{\pi}_{i-1,0} + \hat{\pi}_{i,1}\\ \hat{\pi}_{i,j} &= \frac{pp_c}{W_i}\hat{\pi}_{i-1,0} + \hat{\pi}_{i,j+1} \qquad \text{for } 1 \le j \le W_i - 2\\ \hat{\pi}_{i,W_i-1} &= \frac{pp_c}{W_i}\hat{\pi}_{i-1,0}.\end{aligned} \quad (24)$$

Instead of W_i in (24), we use W_K for $\hat{\pi}_{i,j}$ for $K \le i \le K+L$. For $L = 0$, (24) is valid for $i = K$. We can simplify (24) for $0 \le i \le K-1$ as

$$\begin{aligned}\hat{\pi}_{i,0} &= p_c^i\frac{\epsilon}{p}\hat{\pi}_E \quad \text{and}\\ \hat{\pi}_{i,j} &= p_c^i\frac{W_i - j}{W_i}\epsilon\hat{\pi}_E \qquad \text{for } 1 \le j \le W_i - 1.\end{aligned} \quad (25)$$

We also have $\hat{\pi}_{i,j}$ for $K \le i \le K+L$

$$\begin{aligned}\hat{\pi}_{i,0} &= p_c^i\frac{\epsilon}{p}\hat{\pi}_E \quad \text{and}\\ \hat{\pi}_{i,j} &= \frac{p_c^i(W_K - j)}{W_K}\epsilon\hat{\pi}_E \qquad \text{for } 1 \le j \le W_K - 1.\end{aligned} \quad (26)$$

By using the following normalization condition:

$$\hat{\pi}_E + \sum_{i=0}^{K-1}\sum_{j=0}^{W_i-1}\hat{\pi}_{i,j} + \sum_{i=K}^{K+L}\sum_{j=0}^{W_K-1}\hat{\pi}_{i,j} = 1 \quad (27)$$

we can get

$$\hat{\pi}_E = \left[1 + \epsilon\left\{\frac{1-p_c^{K+L+1}}{p(1-p_c)} + \frac{p_c}{2}\left(\frac{2\mathcal{W}_0(1-(2p_c)^{K-1})}{1-2p_c} - \frac{1-p_c^{K-1}}{1-p_c}\right) + \frac{p_c^K(1-p_c^{L+1})}{1-p_c}\left(\frac{W_K-1}{2}\right)\right\}\right]^{-1}. \quad (28)$$

Denote by $\hat{p}_r$ the retransmission probability for the BEB algorithm, which can be expressed as

$$p_r = p\sum_{i=0}^{K+L}\hat{\pi}_{i,0} = \frac{1-p_c^{K+L+1}}{p_s}\epsilon\hat{\pi}_E. \quad (29)$$

Since (3) is also valid for the BEB algorithm, as in (10), we can construct $f_B(p_r)$ with (3) and (29) as

$$f_B(p_r) = p_r - \frac{1-p_c^{K+L+1}}{p_s}\epsilon\hat{\pi}_E \quad (30)$$

and then solve for $f_B(p_r) = 0$. The system throughput is readily obtained by (11). The packet-dropping probability can be obtained as

$$P_d = p_c p\hat{\pi}_{K+L,0} = p_c^{K+L+1}\epsilon\hat{\pi}_E. \quad (31)$$

As we have seen before in (12), either large K or L makes P_d vanish.

To analyze the delay performance of BEB algorithm, let $\mathcal{D}_{i,j}$ denote the random variable of the packet retransmission delay starting from the jth backoff counter of the ith backoff window of the BEB algorithm. Then, similarly to (13), the mean access delay $\overline{d}$ is expressed as

$$\overline{d} = \overline{\mathcal{D}}_{0,0} + \frac{\mathcal{W}_0 - 1}{2}. \quad (32)$$

As in (15), we can express $\overline{\mathcal{D}}_{i,0}$ for $0 \le i \le K+L-1$ by

$$\begin{aligned}\overline{\mathcal{D}}_{i,0} &= pp_s + (1-p)E[1+\mathcal{D}_{i,0}] + \frac{pp_c}{W_{\min(i+1,K)}}E\left[\sum_{j=0}^{W_{\min(i+1,K)}-1}(1+\mathcal{D}_{i+1,j})\right]\\ &= \frac{1}{p} + p_c\left(\overline{\mathcal{D}}_{i+1,0} + \frac{W_{\min(i+1,K)}-1}{2}\right)\end{aligned} \quad (33)$$

in which $\min(a,b)$ denotes the minimum of a and b. In (33), we have used $\overline{\mathcal{D}}_{i,j} = j + \overline{\mathcal{D}}_{i,0}$ and (1). Then, we have

$$\overline{\mathcal{D}}_{i,0} = \left(\frac{1}{p} - \frac{p_c}{2}\right)\frac{1-p_c^{K-1-i}}{1-p_c} + p_cW_0\frac{1-(2p_c)^{K-1-i}}{1-2p_c} + \overline{\mathcal{D}}_{K-1,0}p_c^{K-1-i} \quad (34)$$

for $0 \le i \le K-2$. We can write $\overline{\mathcal{D}}_{i,0}$ for $K-1 \le i \le K+L-1$ as

$$\overline{\mathcal{D}}_{i,0} = \frac{1}{1-p_c}\left(\frac{1}{p}\left(1-p_c^{K+L-i+1}\right) + p_c\frac{W_K-1}{2}\left(1-p_c^{K+L-i}\right)\right) \quad (35)$$

where we have used $\overline{\mathcal{D}}_{K+L,0} = 1/p$. For both the UB and BEB algorithms, as $L \to \infty$, the respective derivations correspond to the algorithm without retry limit.

Now, the second moment of the access delay of the BEB algorithm can be expressed as

$$\overline{d^2} = \frac{1}{\mathcal{W}_0} \sum_{j=0}^{\mathcal{W}_0 - 1} \overline{\mathcal{D}^2}_{0,j} = \overline{\mathcal{D}^2}_{0,0} + \overline{\mathcal{D}}_{0,0} + 0.5. \tag{36}$$

As in (20), we can write $\overline{\mathcal{D}^2}_{i,0}$ for $0 \le i \le K + L - 1$ as

$$\overline{\mathcal{D}^2}_{i,0} = pp_s + (1-p)E\left[(1+\mathcal{D}_{i,0})^2\right] + \frac{pp_c}{W_{\min(i+1,K)}} E\left[\sum_{j=0}^{W_{\min(i+1,K)}-1} (1+\mathcal{D}_{i+1,j})^2\right]. \tag{37}$$

After some lengthy manipulation with $\overline{\mathcal{D}^2}_{i,j} = 2j\overline{\mathcal{D}}_{i,0} + j^2 + \overline{\mathcal{D}^2}_{i,0}$ for $0 \le i \le K + L$ and $1 \le j \le W_{\min(i+1,K)} - 1$, we can rearrange (37) as

$$\overline{\mathcal{D}^2}_{i,0} = \frac{1}{p}\Bigg[\left(1+pp_c\frac{2W^2_{\min(i+1,K)} + 3W_{\min(i+1,K)} - 5}{6}\right) + 2(1-p)\overline{\mathcal{D}}_{i,0} + pp_c\left(\left(W_{\min(i+1,K)}+1\right)\overline{\mathcal{D}}_{i+1,0} + \overline{\mathcal{D}^2}_{i+1,0}\right)\Bigg]. \tag{38}$$

Finally, we have $\overline{\mathcal{D}^2}_{K+L,0} = 1/p(1 + 2(1-p)\overline{\mathcal{D}}_{K+L,0})$.

D. Stability

We discuss now the stability of these backoff algorithms, particularly for the S-ALOHA channel, i.e., $P = 1$. Accordingly, in what follows, we use $p_s = (1 - p_r)^{M-1}$. For the system with $P > 1$, the following argument can be similarly applied: However, it might not seem obvious due to $C_{k,n}$ in (3). The stability of either the UB or BEB algorithm can be guaranteed, when a unique p_r that solves (10) or (30) exists. If multiple solutions are found, the system is said to be bistable, i.e., of multiple equilibrium points. Finally, when no solution is found, the system is said to be unstable, i.e., backlog grows without bound. It is not difficult to prove the existence of the solution by showing $f_U(0) < 0$ and $f_U(p_r) > 0$ for some $0 < p_r \le 1$ in the UB algorithm. At $p_r = 0$, we get

$$f_U(0) = -1/\left(1/\epsilon + (1/p + 0.5(U-1))\right). \tag{39}$$

As $p_r \to 1$, p_s exponentially vanishes ($p_c \to 1$) due to M in (3). Then, we approximately write

$$\lim_{p_r \to 1} f_U(p_r) = p_r - \frac{\delta_U}{1/\epsilon + \delta_U\left(1/p + 0.5(U-1)\right)} \tag{40}$$

with $\delta_U \equiv L + 1$. If (40) becomes positive for $0 < p_r \le 1$, it implies an odd number of solutions of $f_U(p_r)$. Note that whether (40) is positive is mainly determined by U. Even for $L \to \infty$ (no retry limit), the existence of the solution can be guaranteed when a large U relative to M is chosen in (40).

TABLE II
NOMENCLATURE FOR BAYESIAN WINDOW ASSIGNMENT ALGORITHM

Notation	Definition
$C_E(t)$	Number of RAPs not transmitted at time t
$C_S(t)$	Number of RAPs successfully received at time t
$\mathcal{B}(t)$	The estimated number of backlogged terminals at time t
λ_t	The estimated number of new packet arrivals at time t
U_t	The window size is assigned at time t
β	Smoothing factor

By the same token, for the BEB algorithm, we have, at $p_r = 0$

$$f_B(0) = -\epsilon p/(p+\epsilon) < 0. \tag{41}$$

As $p_r \to 1$, we can write

$$\lim_{p_r \to 1} f_B(p_r) = p_r - \frac{2\delta_B}{2/\epsilon + \mathcal{W}_0 2^K (L+1) + 2\delta_B/p - (K + L + 2\mathcal{W}_0)} \tag{42}$$

with $\delta_B \equiv K + L + 1$. As before, we can have an odd number of solutions of $f_B(p_r)$, if $f_B(p_r) > 0$ for $0 < p_r \le 1$. Similar to U in (40), whether (42) is positive depends on K. Even some small K can guarantee that (42) is positive due to the exponential term in the denominator of (42). Letting $\epsilon \to 1$ and $p \to 1$ and ignoring some negligible terms in (42), we have approximately

$$f_B(p_r) \approx p_r - \frac{\delta_B}{\mathcal{W}_0 2^{K-1}(L+1)}. \tag{43}$$

For the BEB algorithm ($\mathcal{W}_0 = 2$) under saturated traffic condition ($\epsilon \to 1$), the solution always exists for some K.

Until now, we have shown the existence of the solutions of $f_U(p_r)$ and $f_B(p_r)$. It is not easy to prove their uniqueness due to the nonlinearity of $f_U(p_r)$ and $f_B(p_r)$. It is shown in [3]–[6] that applying a retry limit, i.e., limiting feedback to the channel input, can stabilize the system with a probability-distribution-based backoff algorithm. If we show the stability of the system with window-based backoff algorithm without retry limit, i.e., $L \to \infty$, it is reasonable to expect that the system with retry limit is also stable. From (8) and (10), we can rewrite the solution of $f_U(p_r)$ as

$$p_r = \frac{\left(1 - (1-p_s)^{L+1}\right)\epsilon}{p_s\left(1 + \epsilon\left(\frac{1}{p} + \frac{U-1}{2}\right)\frac{1-p_c^{L+1}}{p_s}\right)} \tag{44}$$

in which the terms with L vanishes, as $L \to \infty$. We can simplify (44) as

$$p_r = \epsilon/\left[p_s + \epsilon\left(1/p + 0.5(U-1)\right)\right]. \tag{45}$$

Since $p_s = (1 - p_r)^{M-1}$ is negligible for $\epsilon \to 1$ (i.e., saturated traffic condition) and large M, a unique p_r can be determined, i.e., $p_r = 2/(U + 1)$ for $p = 1$. Although p_r can be affected by $(1 - p_r)^{M-1}$ for unsaturated traffic condition, the uniqueness of p_r can be found for large M. Similarly, letting $L \to \infty$, we

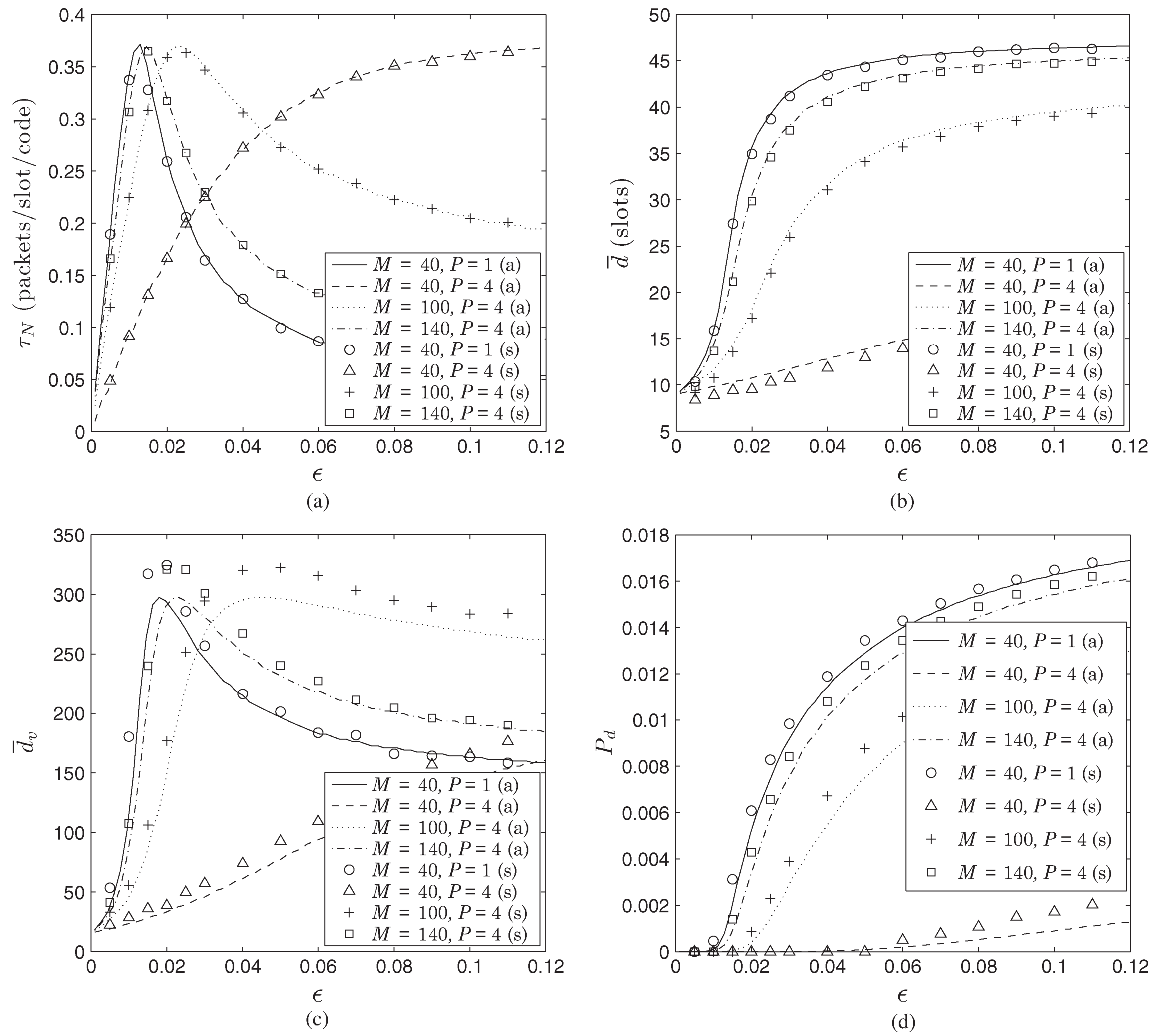


Fig. 1. UB algorithm with different M and P. (a) Normalized throughput. (b) Mean of retransmission delay. (c) Variance of retransmission delay. (d) Packet-dropping probability.

can approximately write the solution of $f_B(p_r)$ for the BEB algorithm in (30) as

$$p_r = \epsilon / \left[p_s + \epsilon \left(1/p + p_c p_s \mathcal{W}_0 (2^{K-1} - 1) + 0.5 p_c^K (W_K - 1) \right) \right]. \tag{46}$$

From (46), we get a unique $p_r = 2/(W_K - 1)$ since we have $p_s \to 0$ $(p_c \to 1)$ for saturated traffic and large M. Thus, window-based backoff algorithms, such as UB and BEB algorithms (even with large L, i.e., many retransmissions) show only either stable or unstable behavior in accordance with U (for the UB algorithm) or K (for the BEB algorithm), and ϵ. Note that the bistability observed in [3]–[6] results from an approximate model of some window-based backoff algorithms, i.e., based on a (truncated) geometric distribution, not from window-based backoff algorithms.

E. Access Prioritization Scheme

Consider C access priority classes, each of which consists of $M^{[\nu]}$ terminals for $1 \le \nu \le C$. Hereafter, the superscript of a parameter indicates the access priority class. We can achieve access prioritization by differentiating p, U, K, L, or a combination of them. Let $p_s^{[\nu]}$ and $p_r^{[\nu]}$ denote the transmission success and retransmission probability of a terminal belonging to the νth access priority class. Then, we can obtain $p_s^{[\nu]}$ of both UB and BEB algorithms by

$$\begin{aligned} p_s^{[\nu]} = & \sum_{n^{[1]}=0}^{M^{[1]}} \cdots \sum_{n^{[\nu]}=0}^{M^{[\nu]}-1} \cdots \sum_{n^{[C]}}^{M^{[C]}} \frac{1}{\widetilde{n}+1} \sum_{k^{[1]}=0}^{n^{[1]}} \cdots \sum_{k^{[C]}=0}^{n^{[C]}} \\ & \times (\widetilde{k}+1) C_{\widetilde{k}+1, \widetilde{n}+1} \Omega \left(n^{[\nu]}, M^{[\nu]} - 1, p_r^{[\nu]} \right) \\ & \times \prod_{i=1, i \ne \nu}^{C} \Omega \left(n^{[i]}, M^{[i]}, p_r^{[i]} \right) \end{aligned} \tag{47}$$

where $\widetilde{k} = \sum_{\nu=1}^{C} k^{[\nu]}$, and $\widetilde{n} = \sum_{\nu=1}^{C} n^{[\nu]}$. For the S-ALOHA channel with $P = 1$, we readily get

$$p_s^{[\nu]} = \left(1 - p_r^{[\nu]} \right)^{M^{[\nu]}-1} \prod_{i=1, i \ne \nu}^{C} \left(1 - p_r^{[i]} \right)^{M^{[i]}}. \tag{48}$$

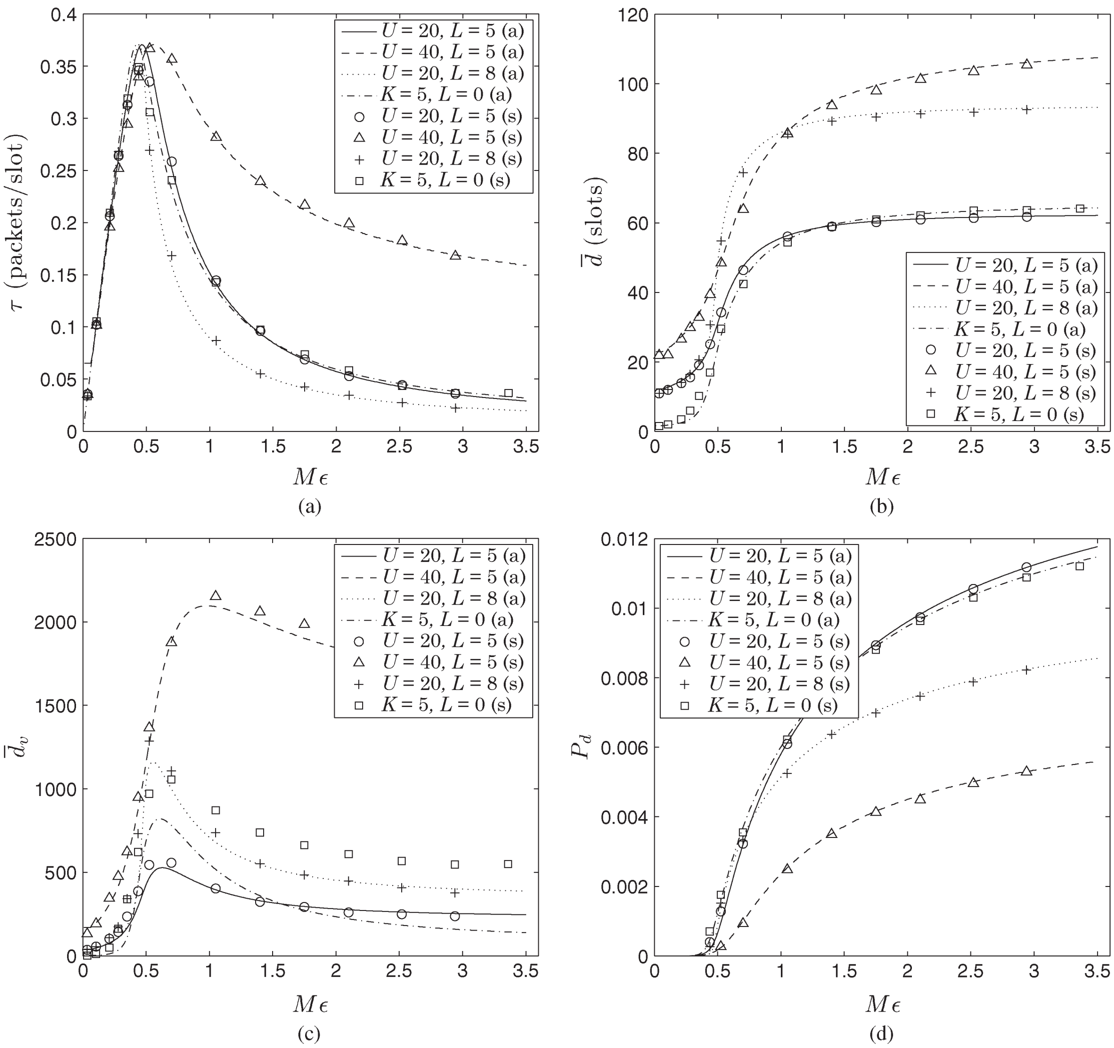


Fig. 2. Comparison of the UB and BEB algorithms with different U and L. (a) Throughput. (b) Mean of retransmission delay. (c) Variance of retransmission delay. (d) Packet-dropping probability.

We can obtain $p_r^{[\nu]}$ of both backoff algorithms for $1 \le \nu \le C$ by substituting (48) into either (9) or (29), respectively. Then, the system throughput can be obtained as

$$\tau = \sum_{n^{[1]}=0}^{M^{[1]}} \cdots \sum_{n^{[C]}=0}^{M^{[C]}} \sum_{k^{[1]}=0}^{n^{[1]}} \cdots \sum_{k^{[C]}=0}^{n^{[C]}} \widetilde{k} \times \prod_{\nu=1}^{C} \Lambda\left(k^{[\nu]}|, n^{[\nu]}, P\right) \Omega\left(n^{[\nu]}, M^{[\nu]}, p_r^{[\nu]}\right) \quad (49)$$

where the first summation holds for all possible combinations of $n^{[\nu]}$ for $1 \le \nu \le C$. For $P = 1$, we have $\tau = \sum_{\nu=1}^{C} \tau^{[\nu]}$, in which $\tau^{[\nu]}$ is expressed as

$$\tau^{[\nu]} = \Omega\left(1, M^{[\nu]}, p_r^{[\nu]}\right) \prod_{i=1, i\neq\nu}^{C} \Omega\left(0, M^{[i]}, p_r^{[i]}\right). \quad (50)$$

We can also obtain the mean and variance of a packet retransmission delay by using (13), (18), (32), and (36) for both backoff algorithms, respectively.

F. DWA Algorithm

The proposed DWA algorithm is given as Algorithm DWA, in which the variables are defined in Table II. We assume that a BS uses a bank of matched filters, each of which tracks a specific RAP, so that, in each time slot, it can determine the number of RAPs not transmitted $C_E(t)$ and the number of RAPs received $C_S(t)$. In the second line of Algorithm DWA, $P - C_E(t) - C_S(t)$ is the number of RAPs unsuccessfully transmitted. The underlying idea of this DWA algorithm is a simple extension of the existing Bayesian algorithm in multidimentional packet reservation multiple access [14], in which each terminal estimates $\mathcal{B}(t)$ first and then determines the retransmission probability by $p_r = \min(1, \mathcal{R}/\mathcal{B}(t))$. Here, $\mathcal{R}$ is the number of orthogonal resources, e.g., the time slots in [14]. In the UB algorithm, we expect $1/U_t$ to approximate p_r in the mean value. If there are P RAPs and $\mathcal{B}(t)$ backlogged terminals, we can determine $U_t = \max(1, \mathcal{B}(t)/P)$ under the assumption of perfectly orthogonal RAPs. Note that $\max(a, b)$ denotes the maximum of a and b. To this end, we estimate $\mathcal{B}(t)$ in the seventh line as the same way in [14]. For $P = 1$, the backlog estimation is identical to Rivest's algorithm [17]. In

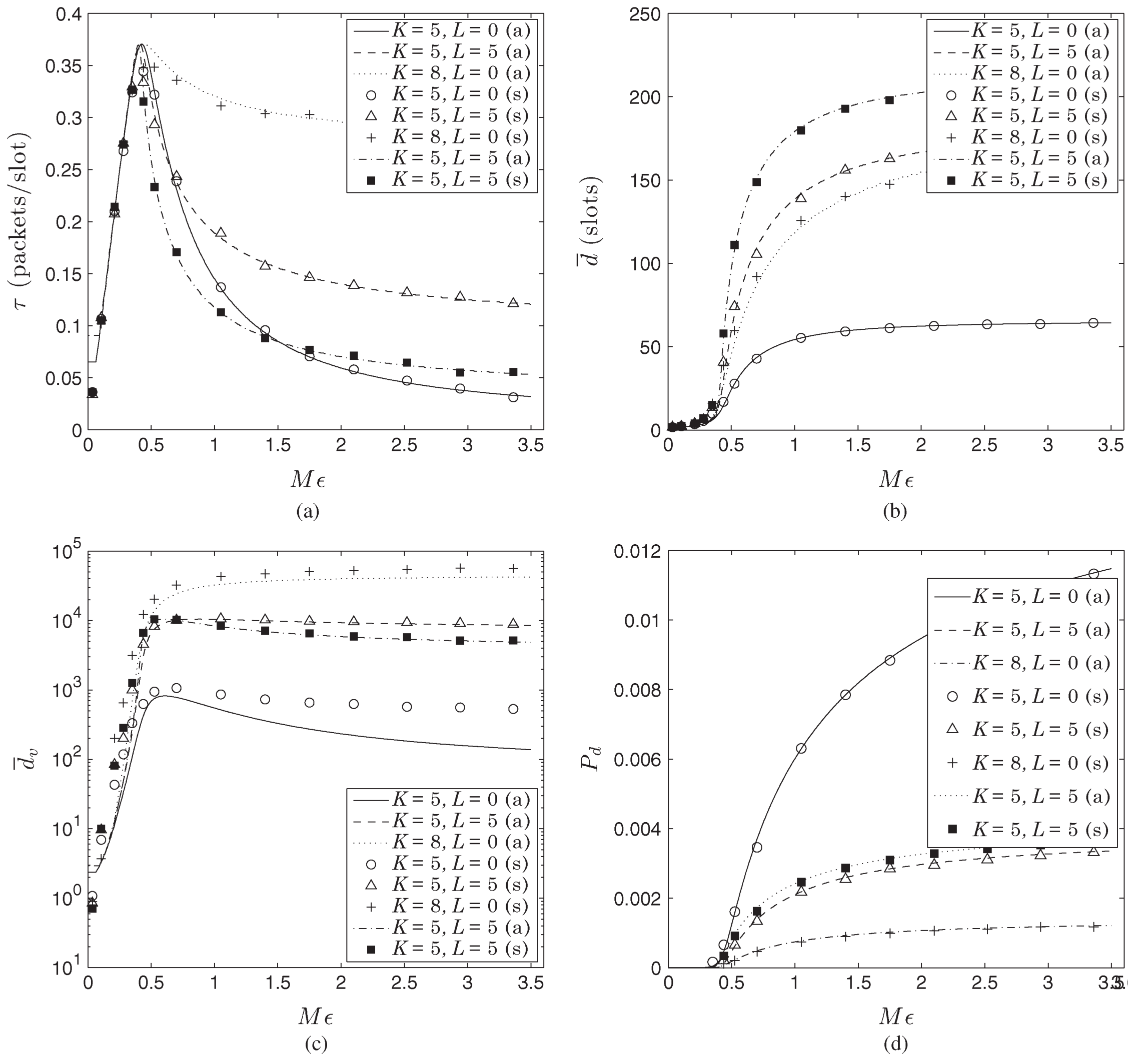


Fig. 3. Performance of the BEB algorithm with different K, L, and M. (a) Throughput. (b) Mean of retransmission delay. (c) Variance of retransmission delay. (d) Packet-dropping probability.

addition to backlog size estimation, we assign the averaged U_t, as shown in the eight line, in which $\lceil x \rceil$ denotes the smallest integer not less than x. In practical implementation, U_t can be broadcast to terminals right before the next slot, or it is given via the first RAR message in UMTS-LTE. Note that, if P_d is not sufficiently small, the estimation on the number of backlogged terminals becomes inexact due to packet dropping.

Algorithm DWA

1: Initialize $\lambda_0 = 0$, $U_0 = 20$, and $\mathcal{B}(0) = 0$, and do the following at $t = 1, 2, \ldots$.
2: **if** $P - C_E(t) - C_S(t) = P$ **then**
3: $\quad \lambda_t = Pe^{-1}$.
4: **else**
5: $\quad \lambda_t = 0.9\lambda_t + 0.1C_S(t)$.
6: **end if**
7: $\mathcal{B}(t) = \max(0, \mathcal{B}(t-1) + (e-2)^{-1}(P - C_E(t) - C_S(t)) - (C_E(t) + C_S(t))) + \lambda_t$.
8: $U_t = \lceil \beta U_{t-1} + (1-\beta)\mathcal{B}(t)/P \rceil$.

IV. Numerical Studies

To validate our analysis, we built a simulation program with MATLAB. Each simulation is carried out for 20 000 slots. In the following, we use $\mathcal{W}_0 = 2$ in the EB algorithm. In Figs. 1–6, the curves are numerically obtained from the equations given in the preceding analysis, whereas the symbols indicated the corresponding simulation results.

1) Effect of Multichannel Code Collisions: We first examine the effect of code collision in the UB algorithm by varying P in Fig. 1(a)–(d), where we have $p = 1$, $U = 15$, and $L = 5$. As expected, we can see that large P provides higher τ_N for large ϵ, given M in Fig. 1(a), and better performances, e.g., $\overline{d}$ in Fig. 1(b). However, we check that the performances of $M = 160$ and $P = 4$ are identical to those of $M = 40$ and $P = 1$. Therefore, it can be concluded that, although a large P can increase the random-access channel capacity, the intrinsic properties of the system performances do not change under perfect orthogonality assumption.

2) Effect of U, K, and L in the UB and EB Algorithms: We examine the UB and EB algorithms without priority classes, i.e., $p = 1$, $C = 1$ in the S-ALOHA channel $(P = 1)$ to see

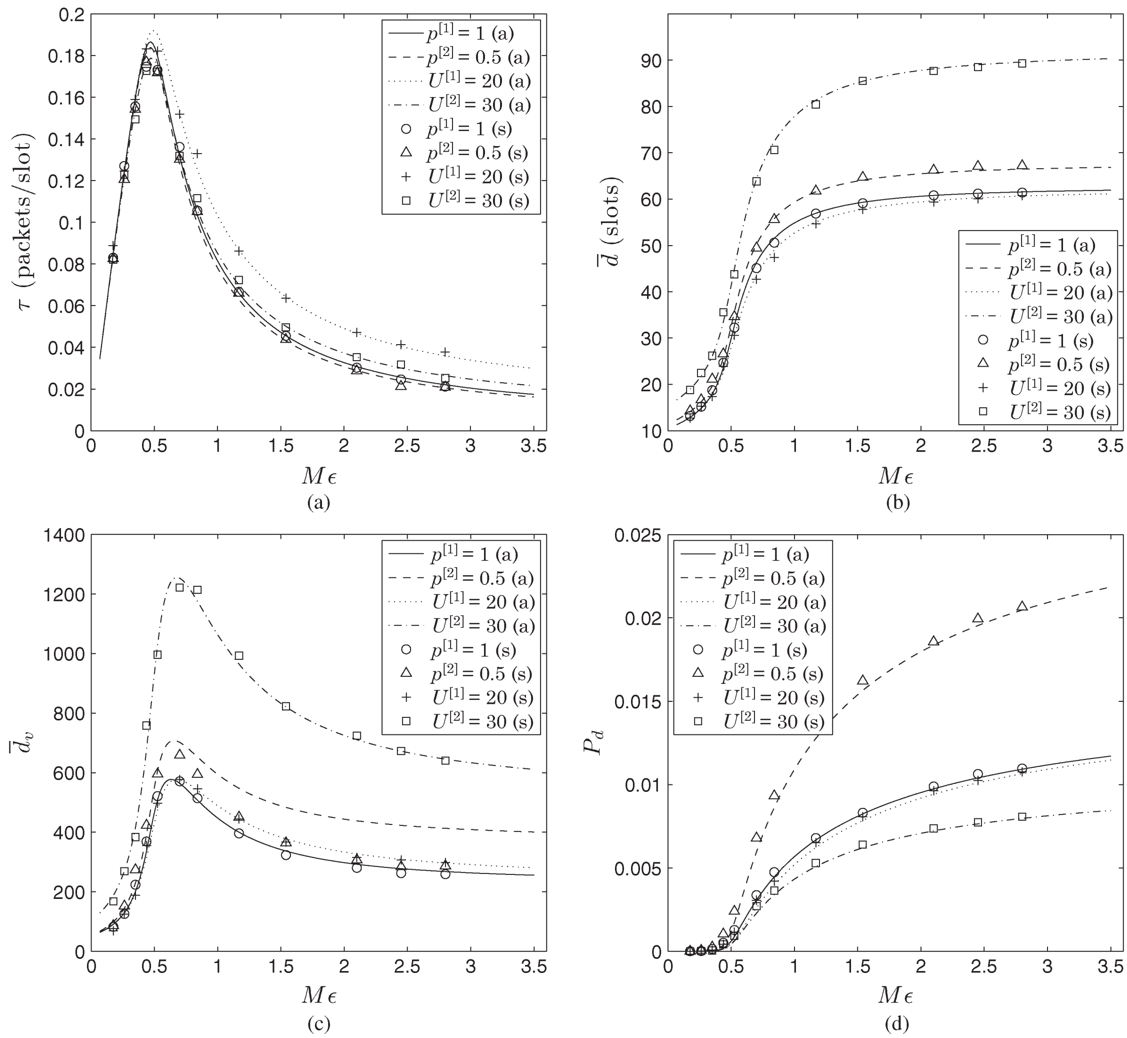


Fig. 4. Access prioritization scheme of the UB algorithm with differential p and U. (a) Throughput. (b) Mean of retransmission delay. (c) Variance of retransmission delay. (d) Packet-dropping probability.

their own characteristics in accordance with window size U, K, and with retry limit L in Figs. 2(a) and 3(d). The horizontal axis indicates the traffic load $(M\epsilon)$ with $M = 70$. In general, our analysis agrees well with simulations. However, we have found that analytical and simulation results show some difference in the delay variance of BEB with small $Lvalues(\leq 3)$, whereas other performance metrics show good agreements between analysis and simulation, given any set of parameters. When a larger L is used $(L \geq 4)$ in the BEB algorithm, we can find good agreement between analytical and simulation results in the delay variance. Our conjecture is that, when L becomes large, the process of our interest becomes memoryless, i.e., less dependent on lower stages with different window sizes.

Comparing the UB and EB algorithms, we observe that UB with $U = 20$ $(L = 5)$ and BEB with $K = 5$ $(L = 0)$ show almost identical throughput τ, the mean access delay $\overline{d}$, and packet-dropping probability P_d in Fig. 2(a), (b), and (d), respectively. Accordingly, it can be concluded that, by properly tuning parameters U, K, and L, the UB algorithm can give similar τ, $\overline{d}$, and P_d as the BEB algorithm. However, we can find two distinct differences between the UB and BEB algorithms. First, $\overline{d}$ of BEB in the underloaded region $(M\epsilon < 0.45)$ is much smaller than that of UB, because lower backoff stages are more frequently utilized in BEB. It is noticeable that, in the underloaded region, τ and P_d of the two backoff algorithms do not show large differences. Second, the delay variance of BEB $(\overline{d}_v)$ is much worse than that of UB in Fig. 2(c), which might be caused by the exponentially increasing nature of backoff intervals in the BEB algorithm. In both algorithms, we can observe that the peak of $\overline{d}_v$ occurs at the traffic load that gives the maximum τ. We can also find that, when the system becomes overloaded, i.e., $p_c \to 1$, $\overline{d}$, and $\overline{d}_v$ of UB are approximated as $\overline{d} = 0.5(L+1)U$ and $\overline{d}_v = (L+1)(U^2-1)/12$, respectively. For BEB, we have $\overline{d} = 0.5\sum_{i=0}^{K} W_i + 0.5LW_K$ and $\overline{d}_v = (1/12)\sum_{i=0}^{K}(W_i^2-1) + L(W_K^2-1)/12$. For the effect of window size, i.e., U in the UB algorithm and K in the BEB algorithm, one can find an important tradeoff between the throughput-packet-dropping probability pair and the mean and variance of delay, particularly in the overloaded region $(M\epsilon > 0.45)$. In this region, larger U or K increases τ and reduces P_d at the expense of increasing the delay. Note that large U or K can stabilize the system for

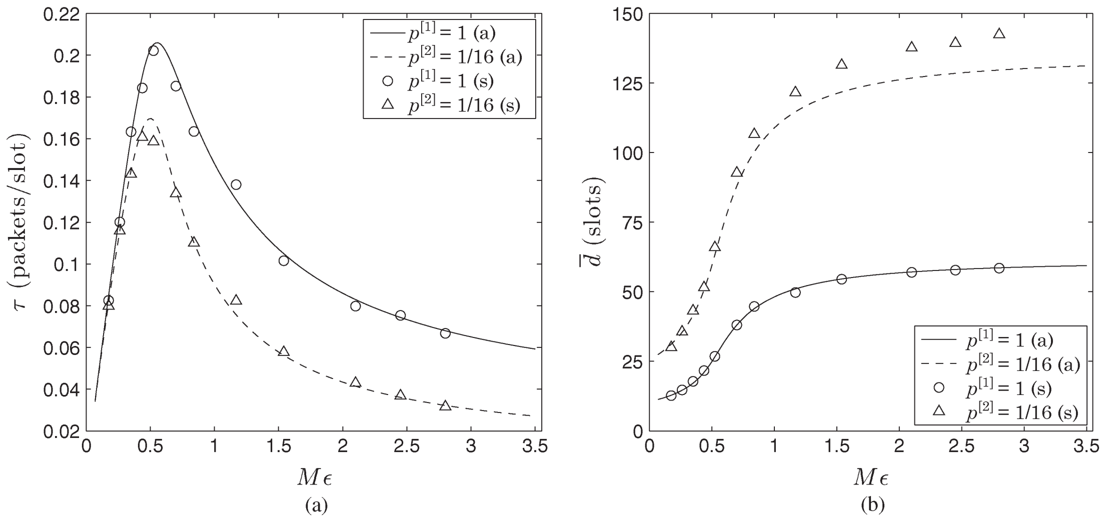


Fig. 5. Access prioritization scheme of the UB algorithm with differential p. (a) Throughput. (b) Mean of retransmission delay.

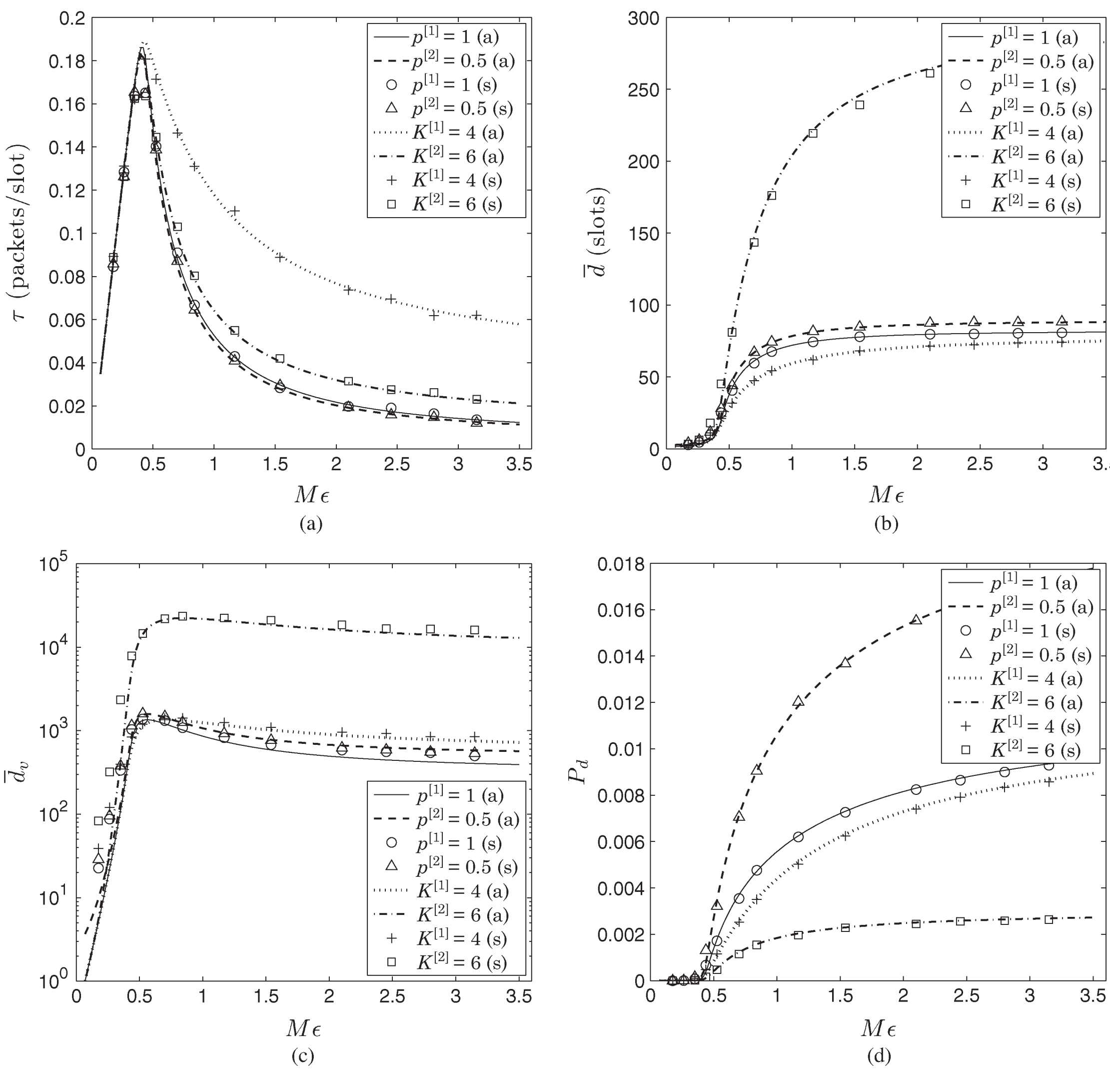


Fig. 6. Access prioritization scheme of the BEB algorithm with differential p and K. (a) Throughput. (b) Mean of retransmission delay. (c) Variance of retransmission delay. (d) Packet-dropping probability.

large M. Therefore, it might be desirable that U and K would be adjusted according to traffic load to minimize delay performance.

For the effect of retry limit, larger L reduces P_d in the UB and BEB algorithms, whereas other performance metrics get worse in the UB algorithm. In other words, UB with a smaller

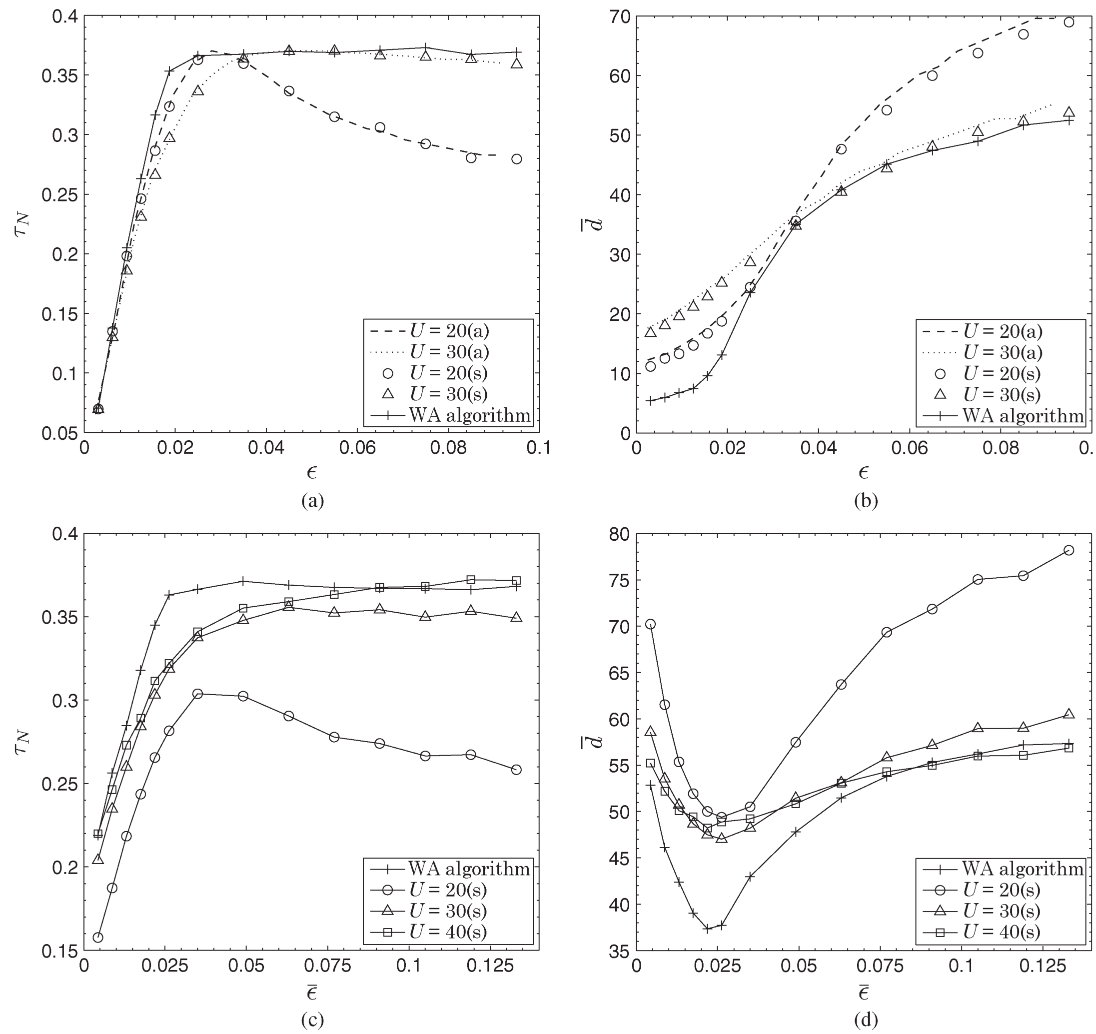


Fig. 7. Performance comparison of the UB algorithm with or without window assignment algorithm. (a) Throughput. (b) Mean of retransmission delay. (c) Throughput. (d) Mean of retransmission delay.

retry limit can provide better throughput and delay performance at the expense of increased packet-dropping probability, in essence removing collisions by packet dropping.

In the UB algorithm, L represents the number of retransmissions, as well as the number of backoff stages. In contrast, we can allow more retransmissions in BEB without increasing the backoff stage, i.e., $K + L$. Thus, we additionally examine the effect of L in the BEB algorithm in Fig. 3(a)–(d). In these figures, the line with black squares indicates the case with $M = 100$, whereas other lines and marks indicate the cases with $M = 70$. It is shown that increasing L, given K in BEB, improves τ and reduces P_d, whereas it increases the mean and variance of delay. The case with $K = 8$ and $L = 0$ improves τ much more in the overloaded region than the case with $K = 5$ and $L = 5$, whereas the former shows huge variance of delay in Fig. 3(c). Note that, at the same traffic load, τ and $\overline{d}$ with larger population size $M = 100$, which is depicted by the dash-dotted line with black squares, are worse than those with $M = 70$.

3) Access Prioritization in the UB Algorithm: In Fig. 4(a)–(d), we present the performance of the prioritized access scheme with two priority classes for $P = 1$ as an example. We experiment with two prioritization schemes: One is to control the persistence probability of the UB algorithm specified in [2], and the other one differentiates U. In the first scheme, we set $p^{[1]} = 1$, $p^{[2]} = 0.5$, and $U^{[1]} = U^{[2]} = 20$, whereas we have $p^{[1]} = p^{[2]} = 1$, $U^{[1]} = 20$, and $U^{[2]} = 30$ for the case of controlling U. In both cases, we have $M^{[1]} = M^{[2]} = 35$ and $L^{[1]} = L^{[2]} = 5$. The horizontal axis of each figure denotes the traffic load of the sum of each priority class, i.e., $M\epsilon = \sum_i M^{[i]}\epsilon^{[i]}$. First, we note that the performance differentiation does not seem to be outstanding in the underloaded region. We can see that access priority can be effectively provided in the overloaded region for both schemes. In Fig. 4(a) and (b), where $p^{[1]} = 1$ and $p^{[2]} = 0.5$ are used, the performance differentiation in τ is not as distinguishable as $\overline{d}$. To obtain a significant differentiated in τ, we decrease $p^{[2]}$ to 2^{-4} while keeping $p^{[1]}$ fixed at 1. The results are shown in Fig. 5(a) and (b). The smaller p is used for the lower priority class, and the larger performance differentiation is achieved in τ and $\overline{d}$. Note that $1/p$ corresponds to the initial delay at each retransmission opportunity. Similarly, controlling U affects both throughput and delay performances. It should be noted that terminals with high priority show a lower P_d, when prioritization is achieved by controlling p in comparison with controlling U, as shown in Fig. 4(c) and (d). In other words, controlling p provides terminals of high priority with better (more desirable) performance in all aspects.

4) Access Prioritization in the BEB Algorithm: In Fig. 6(a)–(d), we examine access prioritization in BEB with p and K for $P = 1$. As in Fig. 4(a)–(d) for the UB algorithm, controlling p provides terminals of higher priority with better performance in all aspects. However, controlling K provides terminals of lower priority with lower P_d. Additionally, terminals with smaller K in BEB get both higher throughput and lower delay, whereas they have higher packet-dropping probability. In other simulations (not shown here) in which $\mathcal{W}_0$ is controlled, the same performance characteristics as controlling K are observed. Therefore, controlling p (initial access delay differentiation at each retransmission opportunity) gives higher priority terminals better performance in all aspects for both backoff algorithms.

5) DWA in UB Algorithm: In Fig. 7(a) and (b), we first test the DWA algorithm in static traffic condition in which ϵ does not change during the entire simulation. As an example, we set $P = 3$. It is observed that the algorithm achieves maximum throughput and minimum delay. In Fig. 7(c) and (d), we set initial ϵ from 0.0031 to 0.095. Then, we increase it by 1.8 times at the 10 000th slot. The results are obtained only from simulations and are time averaged. Here, the horizontal axis indicates the average value of ϵ, e.g., $\overline{\epsilon} = 0.0031(1 + 1.8)/2 = 0.0043$. In such a time-varying traffic condition, the algorithm achieves the maximum throughput and the minimum delay as well. Thus, under the perfect orthogonality assumption and low P_d, it is concluded that $1/U_t$ can approximate the optimum retransmission probability mentioned in [14] and [17].

V. Conclusion

In this paper, we have examined UB and BEB algorithms with retry limit for random-access channels in UMTS-LTE and IEEE 802.16 systems, respectively. We have analyzed their performance in terms of throughput, mean, and variance of packet retransmission delay and packet-dropping probability, and examined the stability properties of both algorithms. We have shown that stability mainly depends on the uniform window size and the maximum backoff stage and also that systems employing these algorithms are either stable or unstable, but do not show bistability behavior. Most of our analytical results have shown good agreement with simulations, except for the delay variance of systems with small retry limits. By properly adjusting the parameters of each backoff algorithm, the two algorithms show virtually identical throughput but different delay variances. We have also shown that controlling the persistence probability provides effective access priority control with respect to various performance metrics. Finally, we have proposed a DWA algorithm, which is a simple extension of an existing Bayesian broadcasting algorithm. The DWA algorithm outperforms the existing backoff algorithms in any traffic condition, as long as perfect orthogonality assumption among random-access code holds.

References

[1] *IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Broadband Wireless Access Systems*, IEEE Std. 802.16™, May 2009.

[2] 3GPP TS 36.321 V9.1.0, *Third-Generation Partnership Project; Evolved Universal Terrestrial Radio Access (E-UTRA) Medium Access Control (MAC) protocol specification*, Sophia-Antipolis, France, Dec. 2009.

[3] S.-W. Kim, "Frequency-hopped spread-spectrum random access with retransmission cutoff and code rate adjustment," *IEEE J. Sel. Areas Commun.*, vol. 10, no. 2, pp. 344–349, Feb. 1992.

[4] 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.

[5] K. Sakakibara, T. Seto, D. Yoshimura, and J. Yamakita, "Effect of exponential backoff scheme and retransmission cutoff on the stability of frequency-hopping slotted ALOHA systems," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 714–722, Jul. 2003.

[6] Y.-S. Liu, "Performance analysis of frequency-hop packet radio networks with generalized retransmission backoff," *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 703–711, Oct. 2002.

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

[8] Y.-P. Fallah, F. Agharebparast, M. R. Minhas, H. M. Alnuweiri, and V. C. M. Leung, "Analytical modeling of contention-based bandwidth request mechanism in IEEE 802.16 wireless networks," *IEEE Trans. Veh. Technol.*, vol. 57, no. 5, pp. 3094–3107, Sep. 2008.

[9] H. L. Vu, S. Chan, and L. L. H. Andrew, "Performance analysis of best-effort service in saturated IEEE 802.16 networks," *IEEE Trans. Veh. Technol.*, vol. 59, no. 1, pp. 460–472, Jan. 2010.

[10] D. Chuck, K.-Y. Chen, and J. M. Chang, "A comprehensive analysis of bandwidth request mechanisms in IEEE 802.16 networks," *IEEE Trans. Veh. Technol.*, vol. 59, no. 4, pp. 2046–2056, May. 2010.

[11] 3GPP TS 25.321 V5.14.0, *Third-Generation Partnership Project; Medium Access Control (MAC) Protocol Specification*, Sophia-Antipolis, France, Sep. 2008.

[12] Y. Xiao, "Performance analysis of priority scheme for IEEE 802.11 and IEEE 802.11e wireless LANs," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1506–1515, Mar. 2000.

[13] M. E. Rivero-Ángeles, D. L. Rodríguez, and F. A. C. Pérez, "Differentiated backoff strategies for prioritized random access delay in multiservice cellular networks," *IEEE Trans. Veh. Technol.*, vol. 58, no. 1, pp. 381–397, Jan. 2009.

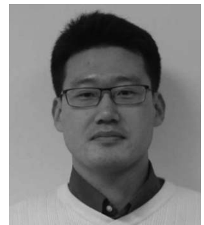
[14] B. E. Brand and A. H. Aghvami, "Multidimensional PRMA with prioritized Bayesian broadcast—A MAC strategy for multiservice traffic over UMTS," *IEEE Trans. Veh. Technol.*, vol. 47, no. 4, pp. 1148–1161, Nov. 1998.

[15] 3GPP TS 36.211 V9.1.0, *Third-Generation Partnership Project; Evolved Universal Terrestrial Radio Access (E-UTRA) Physical channel and modulation specification*, Sophia-Antipolis, France, Apr. 2010.

[16] L. Nuaymi, *WiMAX: Technology for Broadband Wireless Access*. New York: Wiley, 2007.

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

[18] Z. Zhang and Y.-J. Liu, "Comments on 'The effect of capture on performance on multichannel slotted ALOHA system,'" *IEEE Trans. Commun.*, vol. 41, no. 10, pp. 1433–1435, Oct. 1993.



Jun-Bae Seo (M'11) received the B.S. and M.Sc. degrees in electrical engineering from Korea University, Korea, 2000 and 2003, respectively. He is currently working toward the Ph.D. degree in electrical and computer engineering with the Department of Electrical and Computer Engineering, University of British Columbia, Canada.

From 2003 to 2006, he was with the Electronics and Telecommunications Research Institute, Korea, as a Member of the Engineering Staff, carrying out research on IEEE 802.16 systems.

Mr. Seo was the recipient of a Natural Sciences and Engineering Research Council Postgraduate Scholarship from September 2009 to August 2011.

Victor C. M. Leung (S'75–M'89–SM'97–F'03) received the B.A.Sc. (Hons.) and Ph.D. degrees in electrical engineering from the University of British Columbia (UBC), Vancouver, BC, Canada, in 1977 and 1981, respectively.

From 1981 to 1987, he was a Senior Member of Technical Staff with MPR Teltech Ltd., Burnaby, BC, specializing in the planning, design, and analysis of satellite communication systems. In 1988, he started his academic career with the Chinese University of Hong Kong, Shatin, Hong Kong, where he was a Lecturer with the Department of Electronics. In 1989, he returned to the UBC as a Faculty Member and is currently a Professor and the TELUS Mobility Research Chair in Advanced Telecommunications Engineering with the Department of Electrical and Computer Engineering. He is a member of the Institute for Computing, Information, and Cognitive Systems, UBC. He also holds adjunct/guest faculty appointments at Jilin University, Changchun, China; Beijing Jiaotong University, Beijing, China; South China University of Technology, Guangzhou, China; Hong Kong Polytechnic University, Kowloon, Hong Kong; and Beijing University of Posts and Telecommunications. He has coauthored more than 500 technical papers in international journals and conference proceeding papers. His research interests are architectural and protocol design, management algorithms, and performance analysis for computer and telecommunication networks, with a current focus on wireless networks and mobile systems.

Dr. Leung is a Fellow of the Engineering Institute of Canada and the Canadian Academy of Engineering. He is a Registered Professional Engineer in the Province of British Columbia. He is a Distinguished Lecturer of the IEEE Communications Society. He has served on the Editorial Boards of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS WIRELESS COMMUNICATIONS SERIES, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, and is serving on the Editorial Boards of the IEEE TRANSACTIONS ON COMPUTERS, *Computer Communications*, the *Journal of Communications and Networks*, as well as several other journals. He has guest edited several journal special issues and served on the Technical Program Committees (TPCs) of numerous international conferences. He is a General Co-Chair of CSA 2011, Chinacom 2011, and MobiWorld and GCN Workshops at the IEEE Infocom 2011. He is a TPC Co-Chair of the Medium Access Control and Cross-layer Design track of the 2012 IEEE Wireless Communications and Networking Conference (WCNC). He chaired the TPC of the wireless networking and cognitive radio track at the Fall 2008 IEEE Vehicular Technology Conference. He was the General Chair of AdhocNets 2010, QShine 2007, and Symposium Chair for Next-Generation Mobile Networks at the International Wireless Communications and Mobile Computing Conference from 2006 to 2008. He was a General Co-Chair of BodyNets 2010, the Cognitive Wireless Communications and Networking Workshop at Infocom 2010, Advanced Sensor Integration Technology Workshop at IEEE Globecom 2010, MobiWorld Workshop at IEEE CCNC 2010, the 2009 IEEE International Conference on Embedded and Ubiquitous Computing, and the 2006 Association on Computing Machinery MSWiM, and a TPC Vice-Chair of the IEEE WCNC 2005. He was the recipient of the APEBC Gold Medal as head of the graduating class with the Faculty of Applied Science, a Natural Sciences and Engineering Research Council Postgraduate Scholarship, and several Best Paper Awards.