Backoff Algorithm with Release Stages for Slotted ALOHA Systems

Katsumi Sakakibara, Mitsuru Sasaki, and Jiro Yamakita, Non-members

ABSTRACT

In this paper, we propose a decentralized backoff algorithm with release stages. In the proposed algorithm the contention window (CW) of a successful user is gradually decreased, while it is reset to zero after successful transmission in conventional backoff algorithms. The rate of decrement of the CW is controlled by the release delay, which is newly introduced by the use of the release stages. The performance of the proposed algorithm is analyzed by means of the equilibrium point analysis in terms of throughput, average and coefficient of variation of transmission delay. Numerical results show that the proposed algorithm can significantly improve the performance by properly choosing the parameters. It is also found that the proposed algorithm can achieve long-term fairness in the sense that it can offer small fluctuation of transmission delay, even for improper selection of the parameters such that the throughput is deteriorated.

Keywords: Backoff Algoritms, Slotted ALOHA, Equilibrium Point Analysis, Throughput, Delay

1. INTRODUCTION

Backoff algorithms are favorably employed in random access communication systems in order to resolve packet collisions. In particular, binary exponential backoff (BEB) algorithms are involved in various standards of medium access control (MAC) protocols such as IEEE 802.3 CSMA/CD wired LAN's [1] and IEEE 802.11 CSMA/CA wireless LAN's [2]. In backoff algorithms packet transmission of each user is controlled by its own contention window (CW), which is increased after packet transmission failure until it reaches to a predefined maximum. At a packet transmission attempt, a user randomly generates the backoff delay, whose range is given by its CW at that time. When packet transmission results in success, the CW is initialized. Backoff algorithms can basically operate in a decentralized manner, since no information on the system state such as the number of backlogged users or the values of the CW of other users are required.

Performance of backoff algorithms has been analyzed in the literature such as [2–4] in conjunction with the IEEE 802.11 standard, [5, 6] with slotted ALOHA, and [7] with frequency-hopping slotted ALOHA. Bianchi [2] has modeled backoff algorithms by a two-dimensional stochastic process representing the level of its CW and the corresponding backoff delay. Ziouva and Antonakopoulos [3] have refined the model in [2] by taking into consideration busy periods detected by the carrier sense mechanism in the IEEE 802.11. Wu et al. [4] proposed an effective backoff algorithm, in which the initialization of the CW is slowed down after successful packet transmission. However, the performance analyses in [2–4] are limited under saturated conditions. Aldous [5], Jeong et al. [6] and Liu [7] discuss backoff algorithms in the slotted ALOHA protocol, using a simplified system mode, in which the backoff delay is replaced by the retransmission probability, so that we can model the system by a simple Markov chain. Furthermore, it has been also reported that backoff algorithms suggest ultimate instability [5] and the unfairness problem [8, 9]. Unfairness implies that a small number of users can have higher possibility of successful transmission than the rest of users, which is emphasized under heavy traffic conditions.

Several attempts have been devised for performance improvement of backoff algorithms. In the protocol proposed in [6], users share the information on the past and the present state of the system and adaptively alter the retransmission probabilities in a synchronized manner among all the users. In [10] the protocol is proposed in which each user estimates the system state such as the number of active users and load configuration and tunes its CW. Bhargavan et al. [9] proposed a fair backoff algorithm by sharing the values of the CW among all the users. However, these attempts require some type of information sharing, which forces users to equip additional hardware or functions.

In this paper, a new backoff algorithm is proposed. The proposed algorithm can be differentiated from conventional ones in the sense that it incorporates release stages with conventional backoff stages. The release stages offer gradual initialization of the CW of successful users in order to improve the performance and mitigate unfairness. The proposed algorithm can operate completely in a distributed manner and it requires neither estimation on the system state in contrast with [6, 10] nor information sharing on the CW.
[9]. The performance of the proposed algorithm is analyzed by means of the equilibrium point analysis (EPA) [11] in terms of throughput, average transmission delay, and the coefficient of variation of transmission delay. The analysis can be applied not only to saturation conditions but also to unsaturation conditions.

The present paper is organized as follows: Section 2 describes the proposed backoff algorithm and the system model for the analysis is presented in Section 3. In Section 4, equations in equilibrium are derived. Expressions for the performance measures are given in Section 5. Then, Section 6 provides numerical examples and indicate the superiority of the proposed algorithm. In order to validate the analytical results, computer simulations are also conducted. Finally, Section 7 concludes the present paper.

2. BACKOFF ALGORITHM WITH RELEASE STAGES

In order to focus on the backoff algorithm itself, we consider simple slotted random access communication systems such as slotted ALOHA, so that no carrier sensing mechanism is considered. The time axis is slotted and the length of all packets is equal to the slot duration, by which time is measured. Suppose that each user can store a single packet in its buffer and that it can generate no new packet if the buffer is occupied (single buffer assumption). A user with empty buffer is said to be released and otherwise backlogged.

Let $N$ be the number of users in the system. Assume that a released user generates a new packet with probability $p$ in a slot. Similarly to the conventional backoff algorithm described in [2–4], operations of each user in the proposed algorithm are governed by its CW. Let $m$ be the level of the CW ($m \in \{0, 1, \ldots, M\}$) and $W_m$ be the corresponding CW, where $m$ indicates the number of transmission failures encountered by a packet in the buffer if $m < M$. By definition, $W_0 \leq W_m \leq W_M$. At a packet transmission attempt, a user with CW $W_m$ generates a $k$-slot backoff delay, where $k$ is randomly chosen among $\{0, 1, \ldots, W_m-1\}$. If $k = 0$, the packet is transmitted in that slot. Otherwise, $k$ is decreased on a slot-by-slot basis. Packet transmission failure results in an increment of $m$ by one up to $M$.

In the proposed algorithm, operations after successful transmission differ from those in the conventional ones. If the packet transmission succeeds, the user decreases its $m$ by one when $m > 0$. Then, if the resultant $m$ is greater than zero, the user generates the release delay among $\{0, 1, \ldots, R_m-1\}$ according to the predetermined probability distribution $\{q_{m,0}, q_{m,1}, \ldots, q_{m,R_m-1}\}$, where $q_{m,k}$ is the probability that the release delay is $k$ for given $m$ and satisfies

$$\sum_{k=0}^{R_m-1} q_{m,k} = 1 \quad \text{and} \quad q_{m,R_m-1} > 0, \quad (1)$$

for $m = 1, 2, \ldots, M - 1$. If a released user generates a new packet before its release delay expires, the packet is transmitted according to the transmission procedure with the CW $W_m$ at that time. Otherwise, if the release delay reaches to zero with no new packet generation, a user decreases its $m$ by one.

3. SYSTEM MODEL

We assume that a feedback channel is error-free and collision-free. The round trip delay is sufficiently small, compared to the slot duration, so that a user can obtain a positive or negative acknowledgment (ACK or NAK) immediately at the end of packet transmission. A packet transmission succeeds only if there are no other contending packets and that all the packets involved in a collision result in transmission failure.

In order to construct an appropriate model of the proposed algorithm, it is supposed that we sample a user state at the beginning of a time slot. Then, we can arrange probabilistic events in a slot, such as packet generation and transmission, as shown in Fig.1. Three events should be taken into account before state transition: 1) a packet generation process, 2) a packet transmission process, and 3) an outcome of packet transmission. Note that our sampling is different from that in [2–4], in which a user state is sampled in the middle of a time slot.

Let $b(t)$ and $d(t)$ denote stochastic processes representing the level of the CW and the backoff/release delay of a user at the beginning of slot $t$ [2]. As defined in the previous section, $b(t) \in \{0, 1, \ldots, M\}$ and $d(t) \in \{0, 1, \ldots, R_{b(t)}-1\}$, if a user is released, or $d(t) \in \{0, 1, \ldots, W(b(t))-1\}$, if a user is backlogged. Then, the user state can be described by a three-tuple $[R/B, b(t), d(t)]$, where the first element indicates whether a user is released (R) or backlogged (B). At this point, imposing an assumption of independent operation among users, we can construct a transition diagram of user states as shown in Fig.2. In Fig.2, $e$ denotes the probability of transmission failure. Suppose that $RS_m$ denotes a set of user states of $[R, m, k]$
for given $m$ and that BS$_m$ is a set of $[B, m, k]$, that is, $\text{RS}_m = \{[R, m, 0], [R, m, 1], \ldots, [R, m, R_m - 1]\}$ and $\text{BS}_m = \{[B, m, 0], [B, m, 1], \ldots, [B, m, W_m - 1]\}$. Compared to the corresponding diagram in [2-4], the followings should be noticed due to the difference of the sample timing of user states:

1. A state transition from $[R, 0, 0]$ to $[R, 0, 0]$ may occur, if a user generates a new packet and zero-backoff delay and if it has a successful transmission. This denies a user state $[B, 0, W_0 - 1]$.

2. Due to the same reason no state transitions from $[R, m, k]$ to $[B, m, W_m - 1]$ occur. A user state $[B, m, W_m - 1]$ can be reached only when the previous transmission results in failure and the backoff delay $k = W_m - 1$ is generated.

3. Our model can describe the immediate first transmission (IFT) and the delayed first transmission (DFT) protocols [12] in a unified manner. The model of the IFT protocol can be constructed by setting $W_0 = 1$, since BS$_0$ can be eliminated.

4. EQUATIONS AT EQUILIBRIUM POINTS

As shown in Fig.2, there are

$$L = 1 + \sum_{m=1}^{M-1} R_m + (W_0 - 1) + \sum_{m=0}^{M} W_m$$ (2)

for given $m$ and that BS$_m$ is a set of $[B, m, k]$, that is, $\text{RS}_m = \{[R, m, 0], [R, m, 1], \ldots, [R, m, R_m - 1]\}$ and $\text{BS}_m = \{[B, m, 0], [B, m, 1], \ldots, [B, m, W_m - 1]\}$. Compared to the corresponding diagram in [2-4], the followings should be noticed due to the difference of the sample timing of user states:

1. A state transition from $[R, 0, 0]$ to $[R, 0, 0]$ may occur, if a user generates a new packet and zero-backoff delay and if it has a successful transmission. This denies a user state $[B, 0, W_0 - 1]$.

2. Due to the same reason no state transitions from $[R, m, k]$ to $[B, m, W_m - 1]$ occur. A user state $[B, m, W_m - 1]$ can be reached only when the previous transmission results in failure and the backoff delay $k = W_m - 1$ is generated.

3. Our model can describe the immediate first transmission (IFT) and the delayed first transmission (DFT) protocols [12] in a unified manner. The model of the IFT protocol can be constructed by setting $W_0 = 1$, since BS$_0$ can be eliminated.

4. EQUATIONS AT EQUILIBRIUM POINTS

As shown in Fig.2, there are

$$L = 1 + \sum_{m=1}^{M-1} R_m + (W_0 - 1) + \sum_{m=0}^{M} W_m$$ (2)

user states. Hence, behavior of the proposed algorithm can be precisely described by an $L$-dimensional Markov chain. However, in order to avoid computational complexity to solve a Markov chain with such large dimension, we take advantage of the EPA [11]. Using the concept of the EPA, we can approximate the steady-state performance of protocols by that at equilibrium points. It is accomplished by solving balance equations derived from the assumption that the in-flow rate is equal to the out-flow rate for any user states.

We denote by $r_{m,k}$ and $b_{m,k}$ the ratio of users in $[R, m, k]$ and $[B, m, k]$ at the beginning of slot, respectively. Obviously, the relation

$$1 = \sum_{m=0}^{M-1} \sum_{k=0}^{R_m-1} r_{m,k} + \sum_{k=0}^{W_0-2} b_{0,k} + \sum_{m=1}^{M} \sum_{k=0}^{W_m-1} b_{m,k}$$ (3)

holds. We will express $r_{m,k}$ and $b_{m,k}$ with one unknown variable $\epsilon$, the probability of transmission failure in equilibrium, as follows:

1. Derive balance equations in RS$_m$, which are expressed as a function of $r_{m,R_m-1}$;

2. Derive balance equations among RS$_m$’s and BS$_m$’s and obtain $r_{m,R_m-1}$ as a function of $r_{0,0}$;

3. Derive balance equations in BS$_m$, which are expressed as a function of $r_{0,0}$;

4. Substitute the derived equations into (3), so that we can obtain $r_{0,0}$ as a function of $\epsilon$.

Note that it follows from our sample timing of user state in Fig.1 that any successful user should be in either of RS$_m$’s at the beginning of the next time slot. Hence, we have $r_{0,0} > 0$, as long as $\epsilon < 1$.

4.1 Balance Equations in RS$_m$

Let $x$ be the ratio of users moving into RS$_m$ in a slot. The flow balance at user state $[R, m, R_m - 1]$ provides

$$r_{m,R_m-1} = q_{m,R_m-1} x$$ (4)
for \( m = 1, 2, \ldots, M - 1 \). The flow balance at \([R, m, k]\) can be given by a recursive expression

\[
r_{m,k} = (1 - p)r_{m,k+1} + q_{m,k}x
\]

for \( k = 0, 1, \ldots, R_m - 2 \). It follows from (4) and (5) that

\[
 r_{m,k} = (1 - p)^2 r_{m,k+2} + \sum_{i=0}^{1} (1 - p)^i q_{m,k+i} x \]

\[
 = \cdots \\
 = (1 - p)^{R_m-1-k} r_{m,R_m-1} \\
+ \sum_{i=0}^{R_m-1-k} (1 - p)^i q_{m,k+i} x \\
= \sum_{i=0}^{R_m-1-k} (1 - p)^i q_{m,k+i} x \\
= \frac{r_{m,R_m-1}}{q_{m,R_m-1}} \sum_{i=0}^{R_m-1-k} (1 - p)^i q_{m,k+i}
\]

(6) for \( k = 0, 1, \ldots, R_m - 1 \). Then, we have

\[
 r_{m,0} = g_m r_{m,R_m-1}, \quad (7)
\]

\[
 \sum_{k=0}^{R_m-1} r_{m,k} = h_m r_{m,R_m-1}, \quad (8)
\]

where \( g_m \) and \( h_m \) are functions defined by

\[
g_m = \frac{1}{q_{m,R_m-1}} \sum_{i=0}^{R_m-1-k} (1 - p)^i q_{m,i}, \quad (9)
\]

\[
h_m = \frac{1}{q_{m,R_m-1}} \sum_{k=0}^{R_m-1-k} \sum_{i=0}^{R_m-1} (1 - p)^i q_{m,k+i}, \quad (10)
\]

for \( m = 1, 2, \ldots, M - 1 \), respectively.

**Example 1:** If a user with successful transmission randomly selects its release delay in \( R_{m} \), that is, if \( q_{m,k} = 1/R_m \) for any \( k = 0, 1, \ldots, R_m - 1 \), then

\[
g_m = \sum_{i=0}^{R_m-1} (1 - p)^i = \frac{1 - (1 - p)^{R_m}}{p}, \quad (11)
\]

\[
h_m = \sum_{k=0}^{R_m-1-k} \sum_{i=0}^{R_m-1} (1 - p)^i \\
= \frac{pR_m - (1 - p)\{1 - (1 - p)^{R_m}\}}{p^2}, \quad (12)
\]

for \( m = 1, 2, \ldots, M - 1 \). This distribution of \( \{q_{m,k}\} \) is referred to as RAND, hereafter.

**Example 2:** If a successful user sets its release delay to \( R_m - 1 \) in a deterministic manner, we have \( q_{m,0} = q_{m,1} = \cdots = q_{m,R_m-2} = 0 \) and \( q_{m,R_m-1} = 1 \). In this case,

\[
g_m = (1 - p)^{R_m-1}, \quad (13)
\]

\[
h_m = \sum_{i=0}^{R_m-1} (1 - p)^i = \frac{1 - (1 - p)^{R_m}}{p}. \quad (14)
\]

for \( m = 1, 2, \ldots, M - 1 \). This distribution of \( \{q_{m,k}\} \) is referred to as FIFO.

**Example 3:** If \( R_1 = R_2 = \cdots = R_{M-1} = 1 \), then

\[
g_m = h_m = 1 \quad (15)
\]

for \( m = 1, 2, \ldots, M - 1 \). This distribution of \( \{q_{m,k}\} \) is referred to as FIX1, which is a straightforward extension of the protocol in [4].

Note that the proposed algorithms with RAND, FIFO, and FIX1 are equivalent for \( p = 1.0 \) and reduced to the system model presented in [4].

### 4.2 Balance Equations among \( R_{m} \)'s and \( B_{m} \)'s

Next, consider the flow balances among \( R_{m} \)'s and \( B_{m} \)'s.

A substantial in-flow of \( R_0 \) (or equivalently \( r_{0,0} \)) consists of i) a flow from \( R_0 \) (ACK), ii) a flow from \( R_1 \) (no packet generation or ACK), iii) a flow from \( B_0 \) (ACK), and iv) a flow from \( B_1 \) (ACK). Thus, a balance equation for \( R_0 \) is expressed by

\[
pr_{0,0} = \frac{p(1 - e)}{W_0} r_{0,0} + (1 - p)r_{1,0} \quad (16)
\]

\[
+ \frac{p(1 - e)}{W_1} \sum_{k=0}^{R_1-1} r_{1,k} + (1 - e)b_{0,0} + (1 - e)b_{1,0}.
\]

A user with no packet generation at \([R, m, 0]\) and a user with a new packet at \([R, m, k]\) leave \( R_m \). An in-flow to \( R_m \) comes from \( R_{m-1} \) (no packet generation or ACK) and \( B_{m+1} \) (ACK). Then, we have a balance equation for \( R_m \)

\[
(1 - p)r_{m,0} + p \sum_{k=0}^{R_m-1} r_{m,k} \\
= (1 - p)r_{m+1,0} + \frac{p(1 - e)}{W_{m+1}} \sum_{k=0}^{R_{m+1}-1} r_{m+1,k} \\
+ (1 - e)b_{m+1,0}
\]

(17) for \( m = 1, 2, \ldots, M - 2 \). Similarly, we obtain a balance equation for \( R_{M-1} \)

\[
(1 - p)r_{M-1,0} + p \sum_{k=0}^{R_{M-1}-1} r_{M-1,k} = (1 - e)b_{M,0}. \quad (18)
\]

Consider a flow balance of \( B_{m} \) for \( m = 0, 1, \ldots, M \). The out-flow rate is \( b_{0,0} \), since the back-off delay expires. Users with transmission failure in \( B_{m-1} \) and users with delayed transmission in \( R_m \) enter \( B_m \). Hence, a balance equation for \( B_m \) can be given by

\[
b_{0,0} = \frac{p(W_0 - 1)}{W_0} r_{0,0}. \quad (19)
\]
\[ b_{m,0} = \frac{pe}{W_m-1} \sum_{k=0}^{R_{m-1}-1} r_{m-1,k} \]  
and (30) for \( m = 1, 2, \ldots, M - 1 \), shown at the top of the next page, where \( F_m(e) \) is defined by (23) (detailed derivations are given in Appendix B).

Therefore, substituting (8), (22), and (28)–(30) into (3), we obtain \( r_{0,0} \) as a function of one unknown variable \( e \); (31), shown at the top of the next page.

5. PERFORMANCE ANALYSIS

5.1 Probabilities of Packet Transmission and Transmission Failure

According to the analysis in [2, 4], we derive a non-linear equation which should be satisfied by \( e \) in equilibrium. Let \( \tau \) denote the probability that a user transmits a packet. Due to the difference of the sample timing for a user state, a slight modification is required in equating \( \tau \), compared to (7) in [2] and (19) in [4]. With the sample timing shown in Fig.1, a released user with \( R_m \) generates a new packet with probability \( p \) and it transmits the packet immediately with probability \( 1/W_m \). Hence, packet transmissions from released users should be taken in account as well as those from backlogged users. The ratio of users transmitting a packet is expressed by

\[ \tau = \frac{p}{W_0} r_{0,0} + \sum_{m=1}^{M-1} \frac{p}{W_m} \sum_{k=0}^{R_m-1} r_{m,k} + \sum_{m=0}^{M} b_{m,0}. \]  

At this point, the right hand side of (32) can be written as a function of \( r_{0,0} \) from (8) (with (22)), (19), (26), and (27).

From another point of view, a packet transmission succeeds if no other packets are simultaneously transmitted. Then, the probability of transmission failure \( e \) can be written as

\[ e = 1 - (1 - \tau)^{N-1}, \]  

where \( N \) is the number of users in the system. Equating (32) and (33), we can obtain a non-linear equation with respect to \( e \):

\[ 1 - e^{N-1} \sqrt{1-e} = \frac{p}{W_0} r_{0,0} + \sum_{m=1}^{M-1} \frac{p}{W_m} \sum_{k=0}^{R_m-1} r_{m,k} + \sum_{m=0}^{M} b_{m,0}, \]  

since \( r_{0,0} \) is a function of \( e \) given by (31). As discussed in Appendix C, there exists at least one root \( e \) in the range of \( (0, 1) \) for the non-linear equation (34). When (34) has two or more roots, we take the largest \( e \) according to the concept of the EPA [11], which provides a lower bound on throughput.
\[
W_m - 1 \sum_{k=0}^{W_m - 1} b_{m,k} = \frac{p e^m [(W_m + 1) ((1-p)W_mg_m + p(W_m - e)h_m) - p(W_m - 1)h_m] F_m(e) r_{0,0}}{2W_m ((1-p)g_m + p(1-e)h_m)}
\] (30)

\[
r_{0,0} = \left[ 1 + \sum_{m=1}^{M-1} \frac{p e^{m+1}h_m}{(1-p)g_m + p(1-e)h_m} F_m(e) + \frac{p(W_m-1)}{2} + \frac{p(MW_m+1)}{2(1-e)} F_M(e) + \sum_{m=1}^{M-1} \frac{p e^m [(W_m+1) ((1-p)W_mg_m + p(W_m - e)h_m) - p(W_m - 1)h_m] F_m(e)}{2W_m ((1-p)g_m + p(1-e)h_m)} \right]^{-1}
\] (31)

5.2 Throughput

We define throughput \( S \) as the average number of successful users in a slot. When the value of \( e \) is determined, we can calculate \( \tau \) from (33). Then, we have

\[
S = N\tau(1-\tau)^{N-1}.
\] (35)

5.3 Transmission Delay

We define transmission delay \( D \) as the time interval from a packet generation to its ACK. In order to derive the first- and second-order characteristics, we first formulate the probability generation function (PGF) of \( D \).

5.3.1 Initial Probability

It follows from the single-buffer assumption that no new packet can be generated by a backlogged user. Thus, a new packet is generated in either of \( RS_m \). Using \( r_{m,k} \), the conditional probability \( t_m \) that a new packet is generated in \( RS_m \) is given by

\[
t_m = \frac{R_m - 1}{\sum_{k=0}^{R_m - 1} r_{m,k}}.
\] (36)

5.3.2 Probability Generating Function

At a packet transmission attempt a user with \( W_m \) randomly generates its backoff delay among \( \{0, 1, \ldots, W_m - 1\} \). It is evident that the PGF with respect to the backoff delay is given by

\[
\psi_m(z) = (1 - e) \sum_{i=0}^{M} e^{i-m} \left\{ \prod_{k=m}^{i} \psi_k(z) \right\}
\] (38)

\[
+ (1 - e) \sum_{j=1}^{M-m} e^{M-m+j} \left\{ \prod_{k=m}^{M} \psi_k(z) \right\} \{\psi_M(z)\}^j.
\]

Averaging (38) on \( m \), the PGF of transmission delay \( D \) is induced as

\[
\Psi(z) = \sum_{m=0}^{M-1} t_m \Psi_m(z),
\] (39)

where \( t_m \) is given by (36).

5.3.3 Average

The average transmission delay is given by evaluating the first derivative of the PGF at \( z = 1 \);

\[
E[D] = \frac{d\Psi(z)}{dz} \bigg|_{z=1} = \sum_{m=0}^{M-1} t_m \frac{d\Psi_m(z)}{dz} \bigg|_{z=1},
\] (40)

where \( E[D] \) is the expectation of random variable \( D \).

5.3.4 Coefficient of Variation

The ratio of the standard deviation to the average of a random variable is referred to as the coefficient of variation [13]. We can compare the spread of two random variables with unequal average by means of the coefficient of variation. The coefficient of variation of transmission delay is defined by

\[
\rho = \frac{\sqrt{\text{Var}[D]}}{E[D]} = \frac{1}{E[D]} \left[ \frac{d^2\Psi(z)}{dz^2} \bigg|_{z=1} + E[D] - \{E[D]\}^2 \right].
\] (41)

In this context, we can utilize the coefficient of variation to measure the fairness of the algorithm. The algorithm is perceived to be fairer than others, if its coefficient of variation of transmission delay is smaller, since transmission delay is less fluctuated.
6. NUMERICAL EXAMPLES

The performance of the proposed algorithm is numerically compared to that of the conventional algorithm (the backoff algorithm without release stages). Equations in equilibrium for the conventional algorithm with the sample timing of user state shown in Fig.1 are briefly summarized in Appendix D.

We examine the derived expressions for $N = 100$ and $M = 5$. With respect to the backoff/release delay we consider simple and well-known BEB algorithm defined by

$$
\begin{align*}
W_m &= 2^m W_0, \quad \text{for } m = 0, 1, 2, 3, 4, 5, \\
R_m &= 2^m W_0, \quad \text{for } m = 1, 2, 3, 4,
\end{align*}
(42)
$$

with the initial backoff delay $W_0 = 4$ and $W_0 = 32$. Three types of the proposed algorithm are considered: RAND, FIFO, and FIX1, as presented in Examples 1–3. In order to verify the accuracy of the EPA, computer simulations were carried out. In the following graphs, the results obtained from the theory in the previous sections are presented by lines and those from computer simulations are plotted by points. In addition, the simulation results of MILD (multiplicative increase linear decrease) algorithm [9] are plotted as a representative of centralized backoff algorithms. The MILD algorithm considered here is specified by the following update of the CW [9]:

$$
\text{CW} = \begin{cases}
\min(1.5\text{CW}, W_0), & \text{packet collisions}, \\
\max(\text{CW} - 1, W_0), & \text{successful transmission}, \\
\text{CW}_{\text{succ}}, & \text{overhearing success},
\end{cases}
(43)
$$

where $\text{CW}_{\text{succ}}$ is the CW included in the overheard successful packet. This copy mechanism enables users to share the CW.

6.1 Probability of Transmission Failure

The nonlinear equation (34) is numerically solved\(^1\). Fortunately, it has been found that for the values discussed in this section, (34) has a unique root $e$ in the range of $(0, 1)$ for $0 < p < 1$.

The probability of transmission failure $e$ is shown in Fig.3. The results from the theory coincide with those from computer simulation, so that the use of approximation introduced by the EPA is validated. The proposed algorithm exhibits smaller $e$, compared to the conventional BEB algorithm. Therefore, the introduction of the release stages can favorably resolve packet collisions. In particular, the probability of transmission failure of FIFO is the smallest. Among RAND, FIFO, and FIX1, FIFO has the largest average of the release delay for each $m$. Hence, the largest backoff delay may be generated for a new packet, which can mitigate excessive packet collisions. It is also found that FIX1 offers its gain only for large $p$.

An iterative method with the maximum tolerable error of $10^{-10}$ is used in numerical evaluation of $e$.

Figure 3: Probability of transmission failure for $N = 100$, $M = 5$, $W_m = 2^m W_0$ ($m = 0, 1, \ldots, 5$), and $R_m = 2^m W_0$ ($m = 1, 2, 3, 4$): (a) $W_0 = 4$ (b) $W_0 = 32$.

Furthermore, the MILD algorithm can provide better performance, since the rate of decreasing the CW is much slower (decreased linearly as shown in (43)).

6.2 Throughput

The throughput $S$ is shown in Fig.4. It is clear from Fig.4(a) that the proposed algorithm can improve the throughput for $W_0 = 4$. However, from Fig.4(b), the proposed algorithm for $W_0 = 32$ rather degrades the throughput, although it yields smaller probability of transmission failure $e$ than the conventional algorithm, as shown in Fig.3(b). In the case of $W_0 = 4$, gradual initialization of the CW of a successful user can favorably resolve packet contentions. On the contrary, in the case of $W_0 = 32$, the rate at which a successful user reduces its CW is slower than...
Throughput: Packet Generation Probability: $p$

S

MILD

Theory   Sim.

Proposed (FIFO)

Conventional BEB

Proposed (RAND)

Proposed (FIX1)

For $W_0 = 4$ and $p > 0.01$ the ratio of idle slots of the proposed algorithm is less than 0.2 as well as the conventional one. Among them except for the MILD, the ratio for FIFO is largest. On the contrary, for $W_0 = 32$ and $p > 0.01$, approximately half of slots is wasted unused in FIFO and RAND, while the conventional algorithm yields the smallest ratio of idle slots. It explains the degradation of the throughput by the proposed algorithm for $W_0 = 32$ regardless of the fact that small probability of transmission failure is realized. Note that the ratio of idle slots for MILD exceeds 0.8 for $W_0 = 32$.

As a consequence, the use of the release stages is effective, when the values of the backoff/release delay are properly chosen. Besides, there is a possibility that large backoff/release delay affords excessive idle slots. It rather deteriorates the throughput perfor-
6.3 Transmission Delay

The average $\mathbb{E}[D]$ and the coefficient of variation of transmission delay $\rho$ are presented in Fig.6 and Fig.7, respectively. The same tendencies as Fig.7 can be observed in Fig.6. The proposed algorithm, in particular FIFO, presents improved performance for $W_0 = 4$. However, the average of transmission delay of the proposed algorithm for $W_0 = 32$ is slightly degraded, compared to that of the conventional BEB algorithm.

A noticeable feature of the proposed algorithm can be found from the coefficient of variation of transmission delay in Fig.7. As afore-mentioned, the coefficient of variation of $D$ can be used as a performance measure for fairness. Surprisingly, the proposed algorithm offers considerable superiority to the conventional algorithm even for $W_0 = 32$. For $W_0 = 4$, $\rho$ of the proposed algorithm is smaller than that of the conventional one, even though its average $\mathbb{E}[D]$, the denominator of $\rho$ in (41), is smaller, as shown in Fig.6(a). Furthermore, $\rho$ of the proposed algorithm for $W_0 = 32$ is also smaller despite the degraded throughput shown in Fig.4(b). It implies that long-term fairness can be achieved by the proposed algorithm. Here, notice that the MILD yields considerably small $\rho$, since its copy mechanism forces users to homogenize the same value of the CW.

Finally, we can conclude from the above numerical example that the proposed algorithm with release stages can significantly improve the performance by properly choosing the backoff/release delay, $W_m$. 

Fig.6: Average of transmission delay for $N = 100$, $M = 5$, $W_m = 2^m W_0$ ($m = 0, 1, \ldots, 5$), and $R_m = 2^m W_0$ ($m = 1, 2, 3, 4$) : (a) $W_0 = 4$ (b) $W_0 = 32$.

Fig.7: Coefficient of variation of transmission delay for $N = 100$, $M = 5$, $W_m = 2^m W_0$ ($m = 0, 1, \ldots, 5$), and $R_m = 2^m W_0$ ($m = 1, 2, 3, 4$) : (a) $W_0 = 4$ (b) $W_0 = 32$. 

rithm offers considerable superiority to the conventional algorithm even for $W_0 = 32$. For $W_0 = 4$, $\rho$ of the proposed algorithm is smaller than that of the conventional one, even though its average $\mathbb{E}[D]$, the denominator of $\rho$ in (41), is smaller, as shown in Fig.6(a). Furthermore, $\rho$ of the proposed algorithm for $W_0 = 32$ is also smaller despite the degraded throughput shown in Fig.4(b). It implies that long-term fairness can be achieved by the proposed algorithm. Here, notice that the MILD yields considerably small $\rho$, since its copy mechanism forces users to homogenize the same value of the CW.

Finally, we can conclude from the above numerical example that the proposed algorithm with release stages can significantly improve the performance by properly choosing the backoff/release delay, $W_m$. 

Fig.6: Average of transmission delay for $N = 100$, $M = 5$, $W_m = 2^m W_0$ ($m = 0, 1, \ldots, 5$), and $R_m = 2^m W_0$ ($m = 1, 2, 3, 4$) : (a) $W_0 = 4$ (b) $W_0 = 32$. 

Fig.7: Coefficient of variation of transmission delay for $N = 100$, $M = 5$, $W_m = 2^m W_0$ ($m = 0, 1, \ldots, 5$), and $R_m = 2^m W_0$ ($m = 1, 2, 3, 4$) : (a) $W_0 = 4$ (b) $W_0 = 32$. 

Finally, we can conclude from the above numerical example that the proposed algorithm with release stages can significantly improve the performance by properly choosing the backoff/release delay, $W_m$.
and \( R_m \). Furthermore, the proposed algorithm can achieve long-term fairness in the sense that it offers small fluctuation of transmission delay, even for improper selection of \( W_m \) and \( R_m \) such that the throughput is degraded.

7. CONCLUSION

In this paper, a new effective backoff algorithm has been proposed. The feature of the proposed algorithm is the use of the release stages. By introducing the release stage, a successful user can gradually decrease its CW. We can control the rate of the decrement of the CW by the release delay, which is generated at each release stage. The performance of the proposed algorithm has been analyzed by means of the EPA in terms of throughput, the average and the coefficient of variation of transmission delay. It has been numerically proved that the proposed algorithm can significantly improve the performance by properly choosing the backoff/release delay. The use of the release stages succeeds in facilitating resolution of packet contention. In particular, the proposed algorithm can achieve long-term fairness in the sense that it offers small fluctuation of transmission delay, even for improper selection of the backoff/release delay such that the throughput is degraded.

APPENDICES

A DERIVATION OF (22)

Consider the local flow balance between \( \text{RS}_{m-1}/\text{BS}_{m-1} \) and \( \text{RS}_m/\text{BS}_m \). We obtain

\[
\frac{pe}{W_m} \sum_{k=0}^{R_{m-1}-1} r_{m,k} + eb_{m-1,0} = (1-p)r_{m,0} + \frac{p(1-e)}{W_m} \sum_{k=0}^{R_{m-1}-1} r_{m,k} + (1-e)b_{m,0}
\]

for \( m = 1, 2, \ldots, M-1 \), where \( R_0 = 1 \). Equating (20) and (44), we have

\[
b_{m,0} = (1-p)r_{m,0} + \frac{p(1-e)}{W_m} \sum_{k=0}^{R_{m-1}-1} r_{m,k} + (1-e)b_{m,0}
\]

and

\[
(1-p)r_{m,0} + \frac{p(1-e)}{W_m} \sum_{k=0}^{R_{m-1}-1} r_{m,k} + (1-e)b_{m,0} + (1-e)b_{m,0} + \frac{p(W_m - 1)}{W_m} \sum_{k=0}^{R_{m-1}-1} r_{m,k},
\]

so that

\[
b_{m,0} = \frac{1-p}{e} r_{m,0} + \frac{p(W_m - e)}{eW_m} \sum_{k=0}^{R_{m-1}-1} r_{m,k}.
\]

Substitution of (19) and (46) into (16) and deletion of \( b_{0,0} \) and \( b_{1,0} \) result in

\[
pe r_{0,0} = \frac{1-p}{e} r_{1,0} + \frac{p(1-e)}{e} \sum_{k=0}^{R_{1-1}} r_{1,k}.
\]

Also, it follows from deletion of \( b_{m+1,0} \) in (17) using (46) that

\[
(1-p)r_{m,0} + p \sum_{k=0}^{R_{m-1}-1} r_{m,k} = (1-p)r_{m+1,0} + \frac{p(1-e)}{W_{m+1}} \sum_{k=0}^{R_{m+1}-1} r_{m+1,k}
\]

for \( m = 1, 2, \ldots, M-2 \). Rewriting (47) and (48) with (7) and (8) leads to the following equations:

\[
r_{1,R_1-1} = \frac{pe^2}{(1-p)g_1 + p(1-e)h_1} r_{0,0}
\]

and

\[
(1-p)g_{m-1} + ph_{m-1})r_{m-1,R_{m-1}-1} = (1-p)g_m + p(1-e)h_m r_{m-1,R_{m-1}-1}
\]

for \( m = 2, 3, \ldots, M-1 \). From the recursive relation (50), we have

\[
r_{m,R_{m-1}-1} = \frac{(1-p)g_{m-1} + ph_{m-1} e^{R_{m-1,R_{m-1}-1}}}{(1-p)g_m + p(1-e)h_m}
\]

In consequence, combination of (49) and (51) yields

\[
r_{m,R_{m-1}-1} = \prod_{i=2}^{m} \frac{(1-p)g_{i-1} + ph_{i-1}}{(1-p)g_i + p(1-e)h_i} e^{r_0,0}
\]

where \( F_m(e) \) is defined by (23), which completes the derivation. Note that it follows from (49) that (22) is true for \( m = 1 \) despite that the recursion of (50) holds for \( m = 2, 3, \ldots, M-1 \),

B DERIVATION OF (28)–(30)

First, consider balance equations in BS\(_0\). Let \( x \) be an in-flow to BS\(_0\), which is uniformly distributed among \( W_0 - 1 \) user states, \([B, 0, k] \) for \( k = 0, 1, \ldots, W_0 - 2 \), as shown in Fig.2. Then,

\[
b_{0,W_0-2} = \frac{x}{W_0 - 1}
\]
and 
\[
b_{0,k} = b_{0,k+1} + \frac{x}{W_0 - 1} b_{0,W_0-2} + \frac{(W_0 - 2 - k)x}{W_0 - 1} = \ldots = b_{0,W_0-2} + \frac{(W_0 - 1 - k)x}{W_0 - 1} \]
for \( k = 0, 1, \ldots, W_0 - 2 \). Since \( b_{0,0} = x \) (\( b_{0,0} \) is an out-flow from \( BS_0 \)),

\[
\sum_{k=0}^{W_0-2} b_{0,k} = \sum_{k=0}^{W_0-2} \frac{(W_0 - 1 - k)x}{W_0 - 1} = \frac{W_0}{2} b_{0,0} = \frac{p(W_0 - 1)}{2} r_{0,0} \] (55)

which provides (28). Note that the last modification is induced from (19).

Second, consider balance equations in \( BS_m \) for \( m = 1, 2, \ldots, M - 1 \). In this case, we can divide an inflow into two categories, depending on its distribution in \( BS_m \). One from \( RS_{m-1} \) and \( BS_{m-1} \) is randomly distributed in \( BS_m \), which is denoted by \( x \). The other from \( RS_m \) is also randomly distributed in \( BS_m \) except for the user state \( [B, m, W_m - 1] \), which is denoted by \( y \). Then, we have \( b_{m,0} = x + y \), \( b_{m,W_m-1} = x/W_m \), and

\[
b_{m,k} = b_{m,k+1} + \frac{x}{W_m} b_{m,W_m-1} + \frac{y}{W_m - 1} \]

\[
= \frac{(W_m - k)x}{W_m} + \frac{(W_m - 1 - k)y}{W_m - 1}. \] (56)

Summing (56) on \( k \) results in

\[
\sum_{k=0}^{W_m-1} b_{m,k} = \sum_{k=0}^{W_m-1} \left\{ \frac{(W_m - k)x}{W_m} + \frac{(W_m - 1 - k)y}{W_m - 1} \right\} = \frac{W_m}{2} x + \frac{W_m}{2} y = \frac{W_m}{2} (x + y) - \frac{1}{2} y = \frac{W_m}{2} b_{m,0} - \frac{p(W_m - 1)}{2W_m} \sum_{k=0}^{R_m-1} r_{m,k}, \] (57)

since \( y \) is the in-flow from \( RS_m \), so that the relation

\[
y = \frac{p(W_m - 1)}{W_m} \sum_{k=0}^{R_m-1} r_{m,k} \] (58)

holds. As a result, (30) can be derived by substituting (26), (8), and (22) into (57).

Finally, consider balance equations in \( BS_M \). By setting \( m = M \) and \( y = 0 \) in (57), we have

\[
\sum_{k=0}^{w_M-1} b_{M,k} = \frac{W_M + 1}{2} b_{M,0}. \] (59)

Substitution of (27) into (59) provides (29).

C EXISTENCE OF EQUILIBRIUM PROBABILITY OF TRANSMISSION FAILURE

Let \( \lambda(e) \) and \( \gamma(e) \) denote the left-hand and right-hand side functions in (34), respectively. It is clear that \( \lambda(e) = 1 - \sqrt{1 - e} \) is a continuous and narrow-sense increasing function, where \( \lambda(0) = 0 \) and \( \lambda(1) = 1 \). Alternatively, let us consider \( \gamma(e) \). From (31), \( r_{0,0}|_{e=0} = 2/[2 + p(W_0 - 1)] \). Then, for \( p > 0 \), we have

\[
\gamma(0) = \frac{p}{W_0} r_{0,0} + \frac{p(W_0 - 1)}{W_0} r_{0,0} \bigg|_{e=0} = \frac{2p}{2 + p(W_0 - 1)} > 0. \] (60)

On the other hand, since the fact that \( e = 1 \) implies that all the users are in \( BS_M \) in steady-state, (32) is reduced to \( \tau = b_{M,0} \) for \( e = 1 \) and from (59) we have

\[
\sum_{k=0}^{w_M-1} b_{M,k} = \frac{W_M + 1}{2} b_{M,0} = 1. \] (61)

Then, for \( W_M > 1 \),

\[
\gamma(1) = b_{M,0} = \frac{2}{W_M + 1} < 1. \] (62)

Furthermore, \( \gamma(e) \) is continuous in \((0, 1)\). Finally, the relations \( \gamma(0) > \lambda(0) = 0 \) and \( \gamma(1) < \lambda(1) = 1 \) and the continuity of \( \lambda(e) \) and \( \gamma(e) \) guarantee the existence of one or more root \( e \) for (34) in the range of \((0, 1)\).

D EQUATIONS FOR CONVENTIONAL ALGORITHM

In the conventional algorithm, all the user states \([R, m, k]\) degenerate into one state \([R, 0, 0]\). We can formulate balance equations in a similar manner presented in Section 4. Then, we can derive expressions of \( b_{m,k} \) as a function of \( r_{0,0} \);

\[
b_{0,k} = \frac{p(W_0 - 1 - k)}{W_0} r_{0,0}, \] (63)

for \( k = 0, 1, \ldots, W_0 - 2 \),

\[
b_{m,k} = \frac{p e^m (W_m - k)}{W_m} r_{0,0}, \] (64)

for \( m = 1, 2, \ldots, M - 1 \) and \( k = 0, 1, \ldots, W_m - 1 \), and

\[
b_{M,k} = \frac{p e^M (W_M - k)}{(1 - e)W_M} r_{0,0}, \] (65)
for $k = 0, 1, \ldots, W_M - 1$. Corresponding to (31) and (32), the following equations can be respectively obtained;

$$r_{0,0} = \left[ 1 + C \sum_{k=0}^{W_1 - 1} b_{0,k} + \sum_{m=1}^{M} \sum_{k=0}^{W_1 - 1} b_{m,k} \right]^{-1}$$

$$= \left[ 1 + \frac{p(W_0 - 1)}{2} + \sum_{m=1}^{M} \frac{e^m(W + 1)}{2} \right]^{-1} (66)$$

$$\tau = \frac{p}{W_0 r_{0,0}} + \sum_{m=0}^{M} b_{m,0} = \frac{p}{1 - \epsilon} r_{0,0}$$

Thus, a nonlinear equation which should be satisfied by $\epsilon$ in equilibrium is simply

$$1 - \frac{1}{\sqrt{1 - \epsilon}} = \frac{p}{1 - \epsilon} r_{0,0}$$

where $r_{0,0}$ is given by (66).

Once $\epsilon$ is calculated, the throughput is given by (35). The PGF of transmission delay is $W(z) = \Psi_0(z)$, where $\Psi_0(z)$ is obtained from (38).

References


