Modeling and Online Adaptation of ALOHA for Low-Power Wide-Area Networks (LPWANs)

Jun-Bae Seo, *Member, IEEE*, Bang Chul Jung, *Senior Member, IEEE*, and Hu Jin, *Senior Member, IEEE*

***Abstract*—Unslotted ALOHA protocol has been adopted as a channel access mechanism in commercial low-power wide-area networks (LPWANs), such as Sigfox and long-range (LoRa) alliance. This work examines the throughput and random access (RA) delay distribution of unslotted ALOHA systems by considering exponential random backoff (ERB) or uniform random backoff (URB) algorithm. We further characterize the operating region of the systems as unsaturated stable, bistable, and saturated regions in terms of the new packet arrival and retransmission rates. To run the system stably with the maximum throughput, we propose a Bayesian online backoff algorithm that estimates the number of backlogged devices. Its performance is compared with other algorithms, such as particle filter (PF)-based algorithm, binary exponential backoff (BEB) algorithm, and the algorithm of exploiting exact backlog size information. Through extensive simulations, it is demonstrated that the performance of the proposed algorithm is very close to the upper bound and robust to time-varying traffic condition.**

***Index Terms*—Access delay, backoff algorithm, Bayesian estimation, online control, pure ALOHA.**

I. Introduction

A. Motivations

INTERNET OF THINGS (IoT) would be a key enabler of smart city, smart farming, smart manufacturing, smart home, etc., where sensing or monitoring devices collect data in order to make better reactive and even proactive decisions for efficiency, productivity, and safety. For instance, it is possible to take a long-term observation on a macroscopic scale, e.g., mobility of people in cities and trends in real estate so as to design a better urban environment for occupant well-being, and in smart city. In rural areas, it also enables farmers to track the movement of cattle and animals' migration and habitat, or obtain data on humidity, soil moisture, etc., of crop field in smart farming. In general, IoT sensors and monitoring devices in various areas, such as urban, suburban, and rural areas, are wirelessly connected with cloud servers and personal smart devices. Accordingly, various types of wireless networks are needed, e.g., short- and long-range one, licensed or unlicensed, minimum data rate required, and low power consumption of devices, etc., which are determined by the cost and lifetime of sensing devices, capital expenditure (CAPEX), and operational expenditure (OPEX) of networks. For example, WiFi is a cost-effective solution for wireless local area networks (WLANs) than long-term evolution (LTE) or the fifth-generation (5G) new radio (NR) system, since it exploits unlicensed band over small coverage area, e.g., 60 m. However, it is inadequate for large rural areas, e.g., around 10 km, where channel sensing is not feasible.

In wireless sensor networks especially for smart farming, it is required that *disposable* sensors can be wirelessly connected over a large coverage area by consuming extremely low power for a long life time. Furthermore, their CAPEX and OPEX should be so low that the price of the product by smart farming based on those networks should not be noticeably raised. To facilitate IoT over large coverage areas, Narrow-Band IoT (NB IoT) and long-range (LoRa) technology [2] for low-power wide-area networks (LPWANs) have been proposed. While both technologies target at a low rate of data collection with low cost, low power consumption, and long lifetime of sensors, NB-IoT basically comes from the third Generation Partnership Project (3GPP) and is compatible with licensed-band LTE such that the existing cellular infrastructure can be used. In contrast, LoRa is a cheaper solution than NB-IoT, since it makes use of the unlicensed band and is based on unslotted random access (RA) systems. Note that unslotted RA systems can make the hardware (especially sensors) much simple and cheap, since expensive hardware for tight time synchronization and signal processing algorithm for large coverage areas in the physical layer are not needed. It is also notable that these two systems indeed are not considered a competitor or a winner for LPWAN, because NB-IoT is more suitable for the urban environment including indoor while LoRa has advantages for rural IoT applications. Especially in LoRa technology, unslotted (pure) ALOHA protocol has been adopted for channel access mechanism at the medium access control layer. In order to make the most out of it, this work focuses on the performance characterization and optimization of unslotted ALOHA systems for rural IoT applications.

Manuscript received February 7, 2021; revised March 23, 2021; accepted April 9, 2021. Date of publication April 14, 2021; date of current version October 7, 2021. This work was supported in part by the "5G-Based IoT Core Technology Development Project" grant funded by the Korea Government (MSIT, Core Technologies for Enhancing Wireless Connectivity of Unlicensed Band Massive IoT in 5G+ Smart City Environment) under Grant 2020-0-00167, and in part by NRF through the Basic Science Research Program funded by the Ministry of Science and ICT under Grant NRF-2019R1A2B5B01070697. *(Corresponding author: Hu Jin.)*

Jun-Bae Seo is with the Department of Information and Communication Engineering, Gyeongsang National University, Tongyeong 53064, South Korea (e-mail: jbseo@gnu.ac.kr).

Bang Chul Jung is with the Department of Electronics Engineering, Chungnam National University, Daejeon 34134, South Korea (e-mail: bcjung@cnu.ac.kr).

Hu Jin is with the Division of Electrical Engineering, Hanyang University, Ansan 15588, South Korea (e-mail: hjin@hanyang.ac.kr).

Digital Object Identifier 10.1109/JIOT.2021.3073237

B. Previous Work and Contributions

Compared to slotted ALOHA and carrier sense multiple access (CSMA) systems, which are used as the RA systems for wireless cellular systems and WLANs, respectively, unslotted ALOHA systems have not received much research attention. As prior works, in order to improve the throughput of unslotted and slotted ALOHA systems, Metzner [3] considered two transmit power groups, such as high and low. When a packet belonging to the high transmit power group is received in the presence of packets from the low transmit power group, it can be successfully decoded. If packets from the same transmit power group are transmitted, none of them is decoded. Thus, the packet transmitted with the high transmit power enjoys a high priority and higher throughput than the ones with the low transmit power. Under the assumption that the packet length is variable, the throughput in [4]–[6] and the mean RA delay in [7] are investigated. In addition, RA delay distribution is examined for a fixed length of packets in [8]. Although [7] and [8] examined RA delay, they assumed that the retransmission rate is equal to the new packet arrival rate. However, the retransmission rate in practical systems can be controlled in order to prevent some congestion so that it can be different from the new packet arrival rate. Therefore, such an assumption cannot capture how randomized retransmission intervals or backoff algorithms affect RA delay distribution. In comparison with [3]–[8], in which new packet arrivals are assumed to be a Poisson process, arbitrary packet arrival times were considered in [9]. Diffusion approximation has been developed for performance analysis of unslotted ALOHA systems in [10], when backoff intervals follow an exponential distribution. The effects of imperfect radio channels on the performance have been investigated in [11] and [12]. However, RA delay distribution has not been analyzed in [9]–[12].

On the other hand, several backoff algorithms of controlling retransmission intervals (or rates) have been proposed in [13]–[15] in order to maximize the throughput of unslotted ALOHA systems. The algorithm developed in [13] utilizes the information on the number of backlogged devices as our proposed algorithm does. However, it needs some observation period to estimate the backlog information, while our proposed algorithm does not need it. Especially, in [14], the particle filter (PF) algorithm is applied to estimate the number of backlogged devices by observing the length of idle periods. In [16]–[18], multipacket reception (MPR) techniques are applied to unslotted ALOHA systems. In MPR-capable systems, more than one packet can be successfully decoded owing to some advanced signal processing techniques, e.g., successive interference cancellation (SIC). More specifically, even though some packets are transmitted during an ongoing packet transmission, such interference can be removed to some degree for successful decoding of the ongoing packet transmission. In [19] and [20] using stochastic geometry, spatial performance of the unslotted ALOHA system has been investigated, where signal-to-noise interference (SINR) played a key role.

Although various aspects of unslotted ALOHA systems with interesting assumptions have been studied so far, the basic performance characterization together with backoff algorithms and designing Bayesian online control algorithm are not yet much explored. Compared to [3]–[20], our work investigates throughput and RA access delay distribution for two backoff algorithms by constructing an approximate performance modeling of unslotted ALOHA systems based on an M/M/1 queueing model. In particular, two backoff algorithms that we consider for unslotted ALOHA systems are exponential random backoff (ERB) and uniform random backoff (URB) algorithms, respectively. Upon retransmission, devices with ERB (or URB) algorithm draw an exponential (or uniform) random variable and (re)schedule their retransmission according to it.

With the two backoff algorithms, our main contribution is *first* to find RA delay distribution when they are employed. It is important to see whether or not a packet delivery within a delay limit is guaranteed in unslotted ALOHA systems in order to employ time-sensitive IoT applications. *Second*, we show that the operating region of unslotted ALOHA systems is divided into three regions, such as unsaturated stable, bistable, and saturated one, in terms of the new packet arrival rate, retransmission rate, and the number of devices. It is notable that when a bistability breaks out in an RA system, the number of devices in retransmission swings from a high to a low number in a hysteretic manner over time, yielding a low throughput and long RA delay. It is desirable to avoid employing the system parameters that cause the bistability. As a result, it will be shown that the unslotted ALOHA system might not be free of bistability in time-varying environments, e.g., the number of devices changes, when a static backoff algorithm is used. *Finally*, in order to get rid of bistability and further achieve the maximum throughput in the time-varying environment, we propose a Bayesian online backoff algorithm. To demonstrate the performance of the proposed algorithm, it is compared to various known backoff algorithms, such as binary exponential backoff (BEB) and PF-based algorithms in [14]. Notice that the PF-based algorithm in [14] and the proposed algorithm in this work keep track of the mean of the number of backlogged devices by observing the length of idle periods. Compared to the PF-based algorithm that uses a set of samples (or particles) to represent the posterior distribution of the backlog process given the length of idle periods, the main contribution of the proposed algorithm lies in developing two simple update rules to estimate the *mean* of the backlog process *directly* and shows better performances than the PF-based algorithm.

C. Organization

This article is organized as follows. In Section II, unslotted ALOHA systems are introduced. Section III presents our performance analysis for throughput and RA delay distribution, and bistability. It also derives the proposed online adaptive backoff algorithm. Section IV validates our analysis against simulations and discusses numerical results. In Section V, concluding remarks are given.

II. System Model

Suppose N devices randomly scattered under the coverage area of one serving access point (AP). We assume that a packet

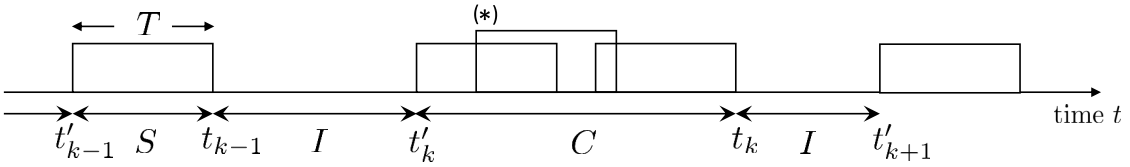


Fig. 1. Timing diagram of the unslotted ALOHA system.

is generated at each device according to a Poisson process with mean rate λ (packets/s) and each device can hold only one packet. We further assume that the length of a packet is a constant, say T (s) long. In unslotted ALOHA systems, a packet transmission is successful if no other packet is transmitted in the interval $(t-T, t+T)$. This period of $2T$ (s) is called the vulnerable period. In our system, the system parameters introduced later are *normalized* by the length of a packet, i.e., T, including the packet size itself. For example, due to this normalization, the vulnerable period is 2, whereas the length of a packet is equal to one. For the unnormalized system, the packet length T of interest can be multiplied back to all the parameters.

When a device has a packet to send, we call it backlogged. As mentioned before, we consider two backoff algorithms for the backlogged devices to perform. *First*, in the system with the ERB algorithm, the backlogged devices draw a random variable according to an exponential distribution with mean $1/\beta$ (s). Let us assume that device i draws a random variable τ_i. When this device finds its (re)transmission failure at time t, it will (re)transmit the packet to the AP at time $t+\tau_i$. *Second*, in the system with the URB algorithm, the backlogged devices pick up a random variable from a uniform distribution with the closed interval $[0, U]$. Note that regarding the unnormalized system, it is only necessary to multiply U and $1/\beta$ by the packet length T, respectively.

We consider time-division duplexing (TDD) so that the devices can switch from transmitting to receiving mode just after transmitting a packet, whereas the AP works in the opposite way, i.e., receiving to transmitting over the downlink after receiving incoming packets. Fig. 1 illustrates a time diagram of the channel, where S, I, and C denote the length of successful transmission, idle period, and collision period, respectively. Note that the length of the success period is equal to one, due to the normalization by T. At the end of success and collision periods, we assume that the RA response message is broadcast *instantaneously* such that the devices receive the outcome of their (re)transmission, i.e., success. Accordingly, the devices that do not find their success can recognize unsuccessful RA. Upon RA failure, the devices reschedule their retransmission according to ERB (or URB) algorithm in the system with ERB (or URB) until making a successful RA.

III. Analysis

Section III-A presents an approximate analysis of the system throughput and RA delay distribution. Section III-B examines the bistability of unslotted ALOHA systems based on catastrophe theory. Finally, Section III-C presents an online adaptive backoff algorithm based on Bayesian estimation. According to the proposed online adaptive backoff algorithm, the AP observes the duration of each idle period and estimates the number of backlogged users real time. Then, it calculates the optimal backoff rate β which maximizes the throughput and broadcasts the backoff rate. It is notable that the contents of Sections III-A and III-B provide the mathematical background for the design of the proposed algorithm.

A. Performance Analysis

Suppose an unslotted ALOHA system with a total of N devices, where the devices employ the ERB algorithm. In Fig. 1, t'_k and t_k for $t'_k, t_k \in \mathbb{R}$ denote the time epoch of the beginning and end of the kth event, respectively, which can be either success or collision.

Let X_{t_k} be the number of backlogged devices at time epoch t_k. We can observe that in the course of time, X_{t_k} changes as

$$X_{t_k} = X_{t_{k-1}} - 1 \cdot \mathbb{I}(\mathcal{S}) + \mathcal{A} \tag{1}$$

where $\mathbb{I}(\mathcal{S})$ takes one if a successful transmission $\mathcal{S}$ takes place; otherwise, zero. Additionally, $\mathcal{A}$ denotes the number of the devices that newly join backlog during $t_k - t_{k-1}$. Since the successful transmission period, idle period I, and collision period C do not follow an exponential distribution, it is true that the evolution of X_{t_k} in (1) is a semi-Markov process. However, in analyzing X_{t_k}, we approximate this system with a generalized M/M/1 queueing system with a population size N as follows. In this approximate model, we focus on the mean input rate to the system and the mean output rate from it. When the system has i backlogged devices, the packet arrival rate (the mean input rate) to the system can be expressed as

$$\lambda_i = (N-i)\lambda, \quad \text{for} \quad i \in \{0, 1, \ldots, N-1\}. \tag{2}$$

Let us find the mean output rate from the system by considering when one packet is successfully transmitted. To proceed further, let us tag one packet transmitted at time t from i backlogged devices. For the tagged packet to be successfully (re)transmitted, two conditions should hold. The first condition is that no prior transmission should be overlapped with this tagged packet transmission. Namely, no packet has been transmitted between $t-T$ and t. As shown in Fig. 1, let us take a look at a packet marked with $*$ and assume that its (re)transmission epoch is time t. Since there is a packet transmission between $t-T$ and t, its transmission fails. To find the probability of no interference between a packet previously transmitted and the tagged packet, let us denote by τ_j the backoff interval that is drawn by device j from an exponential random variable with mean $1/\beta$. Since the time epoch of a first

retransmission made by i backlogged devices is the minimum of i exponential random variables, we can write it as

$$I = \min(\tau_1, \tau_2, \ldots, \tau_i). \tag{3}$$

Let X denote the number of backlogged devices. In (3), we have $X = i$. Then, the complementary conditional cumulative distribution function (CDF) of I can be expressed as

$$\Pr[I > t | X = i] = e^{-i\beta t}. \tag{4}$$

For $I > T$, the first condition occurs with probability $e^{-i\beta T}$. The second condition of the successful transmission of the tagged packet is that once the tagged packet is transmitted, no transmission should be made by the remaining $i-1$ devices during its transmission. Let us consider the conditional probability density function (PDF) of (4)

$$f_I(t|X = i) = i\beta e^{-i\beta t} \tag{5}$$

which is exponential distribution with mean $1/(i\beta)$. We then expect that the remaining $i-1$ devices take exponentially distributed backoff times with mean $1/((i-1)\beta)$, which enables us to assume that retransmissions from $i-1$ remaining backlogged devices form a Poisson process with mean $(i-1)\beta t$, this occurs with probability $e^{-(i-1)\beta T}$. Let μ_i denote the (average) output rate from the system with $X_{t_k} = i$. This can be obtained as

$$\mu_i = i\beta e^{-i\beta T} e^{-(i-1)\beta T} = i\beta e^{-(2i-1)\beta T}. \tag{6}$$

Since we normalize the packet length T by itself, i.e., $T = 1$, we get $\mu_i = i\beta e^{-(2i-1)\beta}$.

To find the system throughput, let π_i denote the steady-state probability that the system has i backlogged devices in this approximation. If π_i is found, the system throughput τ can be obtained as

$$\tau = \sum_{k=1}^{N} \mu_k \pi_k. \tag{7}$$

To find π_i, let us write a flow balance equation at each state as

$$\lambda_{i-1}\pi_{i-1} = \mu_i \pi_i, \quad \text{for} \quad i \in \{1, 2, \ldots, N\}. \tag{8}$$

We can rewrite π_i in (8) for $i = \{1, 2, \ldots, N\}$ as

$$\pi_i = \pi_0 \prod_{k=1}^{i} \frac{\lambda_{k-1}}{\mu_k}. \tag{9}$$

Using $\sum_{i=0}^{N} \pi_i = 1$, we get π_0 as

$$\pi_0 = \left(1 + \sum_{k=1}^{N} \prod_{i=1}^{k} \frac{\lambda_{i-1}}{\mu_i}\right)^{-1}. \tag{10}$$

Let us examine RA delay distribution. To do this, we consider Laplace–Stieltjes transform (LST) of exponential random variables with mean $1/\beta$

$$\mathcal{L}_B(s) = \mathbb{E}\left[s^{-sB}\right] = \frac{\beta}{s+\beta}. \tag{11}$$

Let p_s be the RA success probability. It can be obtained as

$$p_s = \pi_0 + \sum_{i=1}^{N} e^{-(2i-1)\beta} \pi_i. \tag{12}$$

Since ERB is repeated upon each RA failure with probability $1 - p_s$, the LST of RA delay distribution can be found as

$$\begin{aligned}\mathbb{E}\left[e^{-sD}\right] &= \sum_{k=1}^{\infty} (\mathcal{L}_B(s))^k (1-p_s)^{k-1} p_s \\ &= \frac{\mathcal{L}_B(s) p_s}{1 - \mathcal{L}_B(s)(1-p_s)}.\end{aligned} \tag{13}$$

Plugging (11) into (13), we have

$$\mathbb{E}\left[e^{-sD}\right] = \frac{\beta p_s}{s + \beta p_s}. \tag{14}$$

Taking the inverse LST of (14) yields the CDF of RA delay as

$$\Pr[D \le t] = 1 - e^{-\beta p_s t}. \tag{15}$$

Let us consider RA delay distribution by URB. In this case, at the kth retransmission epoch, RA delay is the sum of k independent and identically distributed (i.i.d.) continuous random variables drawn from the interval $[0, U]$. Since the LST of uniform random variable for the interval $[0, U]$ is expressed as

$$\mathcal{L}_B(s) = \frac{1 - e^{-sU}}{sU} \tag{16}$$

we can get the CDF of RA delay by taking the inverse LST of (16) as done before. However, it is cumbersome. Notice that the CDF and PDF of the sum of n i.i.d. uniform random variables from the unit interval $[0, 1]$ has been known as Irwin–Hall distribution [21]. Its CDF and PDF are, respectively, given as

$$F_n(t) = \sum_{i=0}^{n} \left[(-1)^i \binom{n}{i} \frac{(t-i)^n}{n!} s_i(t) \right] \tag{17}$$

and

$$f_n(t) = \sum_{i=0}^{n} \left[(-1)^i \binom{n}{i} \frac{(t-i)^{n-1}}{(n-1)!} s_i(t) \right] \tag{18}$$

where $s_i(t)$ is the unit step function

$$s_i(t) = \begin{cases} 0, & \text{for } t < i \\ 1, & \text{for } t \ge i. \end{cases} \tag{19}$$

Now to find RA delay distribution by URB, let X and Y be the continuous uniform random variable in the unit interval and the interval $[0, U]$, respectively. Then, the CDF of the sum of Y's can be obtained by the linear transformation of one random variable into another, i.e., $\Pr[Y \le y] = \Pr[UX \le y] = \Pr[X \le y/U]$. Accordingly, when URB is employed, the CDF and PDF of RA delay can be obtained as

$$\Pr[D \le t] = \sum_{n=1}^{\infty} F_n\left(\frac{t}{U}\right)(1-p_s)^{n-1} p_s \tag{20}$$

and

$$f_D(t) = \frac{1}{U} \sum_{n=1}^{\infty} f_n\left(\frac{t}{U}\right)(1-p_s)^{n-1} p_s. \tag{21}$$

Since the mean of exponential and uniform random variables is $1/\beta$ and $U/2$, respectively, we will use p_s in (20) and (21) from (12) with $\beta = 2/U$.

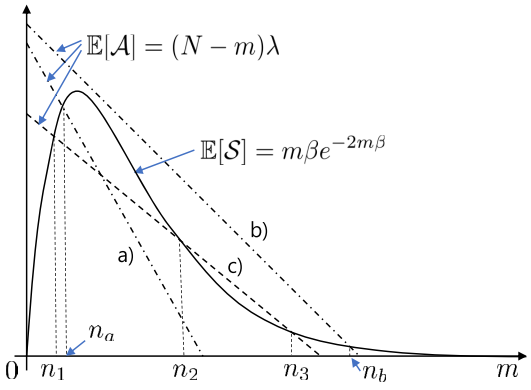


Fig. 2. Mean flow-out and flow-in of the system.

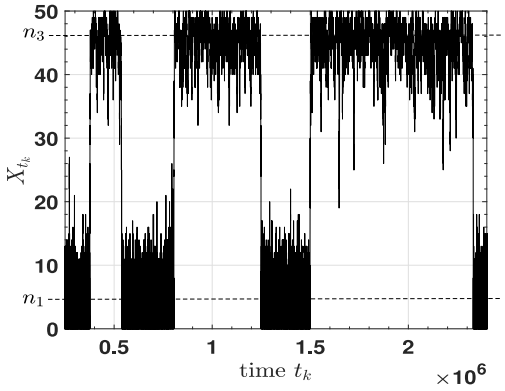


Fig. 3. Sample path of backlog size in the bistable system: $N = 50$.

B. Bistability

This section shows that the operating region of unslotted ALOHA systems consists of unsaturated stable, bistable, and saturated regions, whereas the unsaturated stable is the desirable operating region. The three operating regions are characterized by the new packet arrival rate λ, retransmission rate β, and the number of devices N. Before proceeding further, let us help understand the notion of bistability in unslotted ALOHA systems. Let us denote by $\mathbb{E}[\mathcal{S}]$ and $\mathbb{E}[\mathcal{A}]$ the mean rate of flow-out and flow-in of unslotted ALOHA systems, respectively. When G is the traffic load with m backlogged devices, i.e., $G = \beta m$, we have

$$\mathbb{E}[\mathcal{S}] \approx Ge^{-2G}. \tag{22}$$

In addition, referring to (2), we can write $\mathbb{E}[\mathcal{A}]$ as

$$\mathbb{E}[\mathcal{A}] = (N-m)\lambda = \left(N - \frac{G}{\beta}\right)\lambda. \tag{23}$$

Then, we can write the following flow balance equation[1]:

$$\widetilde{\mathcal{F}}(m) = \mathbb{E}[\mathcal{S}] - \mathbb{E}[\mathcal{A}]. \tag{24}$$

Fig. 2 depicts (22) and (23). While lines a) and b) have only one intersection at $m = n_a$ and n_b with $\mathbb{E}[\mathcal{S}]$, respectively, line c) has three intersections, such as n_1, n_2, and n_3. In other words, $\widetilde{\mathcal{F}}(m) = 0$ has only one root in the previous two cases, while it has three roots in the last case. In the system with line a) [or line b)], n_a (or n_b) is approximately the mean number of backlogged devices. Thus, it can be said that the previous two systems correspond to unsaturated and saturated systems, respectively, since $n_a \ll n_b$. However, in the system with line c), the number of backlogged devices will swing from n_1 to n_3 and *vice versa*. More specifically, Fig. 3 shows a sample path of X_{t_k} for an unslotted ALOHA system with $N = 50$, $\lambda = 3 \times 10^{-3}$, and $\beta = 0.06$. In the system with these parameters, $\widetilde{\mathcal{F}}(m) = 0$ has three roots so that X_{t_k} moves back and forth between a low state, say n_1, and a high state n_3 over time in Fig. 3. In other words, the system state, i.e., the number of backlogged devices, swings from unsaturated to saturated one, and *vice versa* as indicated in Fig. 2. In such cases, it can be said that the system is in the *bistable* state. In bistable systems as shown in Fig. 3, the trajectory of the system state X_{t_k} shows a hysteretic (or ping-pong) behavior[2] between n_1 and n_3 via sudden jumps as addressed in [25], while the intermediate states between n_1 and n_3 are inaccessible. Accordingly, the bistable and saturated regions should be avoided.

Before asking how to avoid the bistability, we can raise the following research question: What is the set of λ and β that can drive the system bistable given N? In other words, we are interested in finding the set of λ and β, with which $\widetilde{\mathcal{F}}(m) = 0$ has more than one roots. To find this, we make use of the catastrophe theory [23]–[25]. Let $\mathcal{F}$ be a flow balance function of the system and $\mathcal{F} : \mathbb{R}^k \times \mathbb{R}^n \to \mathbb{R}$, where k and n are the number of control variables and system states, respectively. If $k = 2$, we have the fold and the cusp catastrophes and $n = 1$, i.e., the system state $x \in [0, 1]$, i.e., the number of backlogged devices normalized by a total number of devices N. Thus, $Nx = m$ is the number of backlogged devices. In [23] and [24], the cusp catastrophe of RA systems may exist if we can solve $\mathcal{F}(x)$ for $x \in [0, 1]$

$$\mathcal{F}(x) = \frac{\partial \mathcal{F}(x)}{\partial x} = \frac{\partial^2 \mathcal{F}(x)}{\partial x^2} = 0 \tag{25}$$

and $[(\partial^k \mathcal{F}(x))/(\partial x^k)] \neq 0$ for $k \geq 3$. It is noteworthy that the flow balance equation $\widetilde{\mathcal{F}}(m)$ in (24) is not differentiable with respect to the system state $m \in \mathbb{Z}^* = \{0\} \cup \mathbb{Z}^+$ as m is an integer. Using $Nx = m$, we can express $G = \beta m = \beta x N$. This enables us to transform the flow balance equation $\widetilde{\mathcal{F}}(m)$ into $\mathcal{F}(x)$, which is differentiable with respect to the normalized state variable x

$$\mathcal{F}(x) = \beta x N e^{-2\beta x N} - N(1-x)\lambda \tag{26}$$

where $\mathbb{E}[\mathcal{A}] = N(1-x)\lambda$.

[1]This flow balance equation is also obtained by first taking the expectation on both sides of (1), i.e., $\mathbb{E}[X_{t_k}] = \mathbb{E}[X_{t_{k-1}} - 1 \cdot \mathbb{I}(\mathcal{S}) + \mathcal{A}] = \mathbb{E}[X_{t_{k-1}}] - \mathbb{E}[1 \cdot \mathbb{I}(\mathcal{S})] + \mathbb{E}[\mathcal{A}]$. Then, letting $t_k \to \infty$ (i.e., steady state) we have $\mathbb{E}[X_{t_k}] = \mathbb{E}[X_{t_{k-1}}]$, which results in $\mathbb{E}[\mathcal{A}] - \mathbb{E}[\mathcal{S}] = 0$.

[2]Notice that at a low number of backlogged devices, the system may show a reasonable good throughput, while it yields extremely low throughput and excessive long RA delay at a high number of backlogged devices. It is expected that depending on the system parameters, X_{t_k} would stay at a high number of backlogged devices for an extremely long period of time.

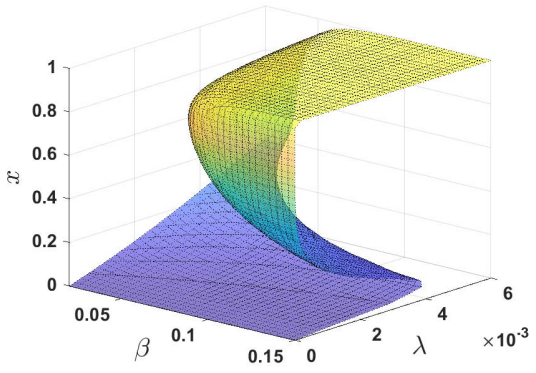


Fig. 4. Catastrophe manifold of the unslotted ALOHA system with $N = 50$.

Now, assuming the existence of the cusp catastrophe, we explore it as follows [23], [24]. The catastrophe manifold Θ is a surface in three dimensions defined by $\Theta = \{(x, \lambda, \beta) | \mathcal{F}(x) = 0\}$. Let us further denote by Θ_B the fold line consisting of the points of Θ, where the manifold surface folds over. It is defined as $\Theta_B = \{(x, \lambda, \beta) | \mathcal{F}(x) = [(d\mathcal{F}(x))/dx] = 0\}$. Using $[(d^2\mathcal{F}(x))/dx^2]$, we can characterize Θ_B as three parts

$$\Theta^+ = \left\{(x, \lambda, \beta) | \mathcal{F}(x) = \frac{d\mathcal{F}(x)}{dx} = 0, \ \frac{d^2\mathcal{F}(x)}{dx^2} > 0\right\} \quad (27)$$

$$\Theta^- = \left\{(x, \lambda, \beta) | \mathcal{F}(x) = \frac{d\mathcal{F}(x)}{dx} = 0, \ \frac{d^2\mathcal{F}(x)}{dx^2} < 0\right\} \quad (28)$$

and the cusp point

$$\Theta_0 = \left\{(x, \lambda, \beta) | \mathcal{F}(x) = \frac{d\mathcal{F}(x)}{dx} = \frac{d^2\mathcal{F}(x)}{dx^2} = 0\right\}. \quad (29)$$

The bifurcation sets Θ^+ and Θ^- are obtained as the projection of the 3-D catastrophe manifold (fold line) Θ^+ and Θ^- into the control space (λ, β), respectively. Fig. 4 may help understanding this. For a given pair of λ and β in $\mathbb{R}^2$, it illustrates x's of satisfying $\mathcal{F}(x) = 0$. As discussed before, for some pairs of λ and β, it can be observed that there are three x's for $\mathcal{F}(x) = 0$. By projecting the 3-D surface onto the 2-D control space λ and β, we can characterize the region that the surface is folded over. The following lemmas find (27) and (28) based on (25).

Lemma 1: The λ corresponding to (27) and (28) can be expressed as

$$\lambda = -\beta e^{-2G}(1 - 2G). \quad (30)$$

Proof: Since λ in (27) and (28) should satisfy $[(d\mathcal{F}(x))/dx] = 0$, we have

$$\frac{\partial \mathcal{F}(x)}{\partial x} = (1 - 2G)e^{-2G} + \frac{\lambda}{\beta} = 0. \quad (31)$$

We can get (30) by rewriting (31) with respect to λ. ■

Lemma 2: Let $x_\pm$ be the normalized state corresponding to Θ^+ and Θ^- in (27) and (28), respectively. We can find $x_\pm$ as

$$x_\pm = \frac{G_\pm}{N\beta} \quad (32)$$

where $G_\pm$ is expressed as

$$G_\pm = \frac{\beta N \pm \sqrt{\beta N(\beta N - 2)}}{2}. \quad (33)$$

Proof: Plugging (30) into (26) we have

$$\mathcal{F}(G) = e^{-2G}\Big(2G^2 - 2\beta N G + \beta N\Big). \quad (34)$$

Setting $\mathcal{F}(G) = 0$, we obtain (33). ■

Notice that in (33), $G_\pm$ is a real number for $\beta > (2/N)$.

Lemma 3: The cusp point Θ_0 in (29) is obtained as

$$x = \frac{1}{\beta N} \xrightarrow{\beta N = 2} \frac{1}{2}. \quad (35)$$

Proof: From (29), we need to find $[(\partial^2\mathcal{F}(x))/(\partial x^2)] = 0$

$$\frac{\partial^2 \mathcal{F}(x)}{\partial x^2} = -4\beta N(1 - \beta x N)e^{-2G} = 0. \quad (36)$$

Therefore, we have $x = 1/(\beta N)$. In fact, the cusp point is the point, at which we have $x_+ = x_-$ or $G_+ = G_-$. From (33), we get $\beta N = 2$ for $x_+ = x_-$. This completes the proof. ■

Let us consider how bistability can be eliminated or avoided by controlling the backoff rate.

Proposition 1: When a backoff algorithm seeks for maximizing throughput by tracking the number of backlogged devices over time, the bistability does not show up.

Proof: Let us maximize the system throughput $\mathbb{E}[\mathcal{S}]$ using $[(\partial\mathbb{E}[\mathcal{S}])/(\partial\beta)] = 0$. This yields

$$\beta = \frac{1}{2xN}. \quad (37)$$

Notice that xN can be a time-varying value, i.e., X_{t_k}. When (37) is employed, we get $\mathbb{E}[\mathcal{S}] = e^{-2}$, i.e., the maximum throughput of the unslotted ALOHA system. Plugging (37) into (26) yields

$$\mathcal{F}(x) = 0.5e^{-1} - N(1 - x)\lambda. \quad (38)$$

This is a linear function of x with $\mathcal{F}(0) = 0.5e^{-1} - N\lambda$ and $\mathcal{F}(1) = 0.5e^{-1}$. If $\mathcal{F}(0) < 0$, it has always a unique solution. ■

Finally, note that we have higher order derivatives of $\mathcal{F}(x)$ for $k \geq 3$ as

$$\frac{\partial^k \mathcal{F}(x)}{\partial x^k} = (-1)^k 4(N\beta)^k e^{-2N\beta x}\Big(2^{k-2}N\beta x - c_k\Big) \quad (39)$$

where c_k is expressed as $c_k = 2c_{k-1} + 2^{k-3}$ for $k = 3, 4, \ldots,$ and $c_2 = 1$.

C. Bayesian Online Backoff Algorithm

The design goal of a backoff algorithm is to control backoff rate $\beta = 1/(2i)$ in real time to realize Proposition 1, where i is the number of backlogged devices at an arbitrary instance of our interest. To this end, it is necessary to estimate i over time. Once the AP broadcasts $\beta = 1/(2i)$ at the end of the success or collision period, the backlogged devices that do not schedule their retransmission yet, or need to reschedule will draw a backoff interval from an exponential random variable with mean $1/\beta$ just broadcast.

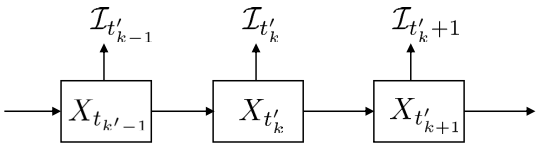


Fig. 5. Observation $\mathcal{I}_{t'_k}$ and the hidden Markov process $X_{t'_k}$.

To show how the proposed Algorithm 1 works, let us look at the timing diagram in Fig. 1. Referring to (5), it is important to notice that the idle period I is a function of $X_{t'_k}$. To be more specific, let $\mathcal{I}_{t'_k}$ denote the idle period caused by $X_{t'_k}$ in Fig. 5. When the AP tries to estimate $X_{t'_k}$, which is a *hidden Markov process*, it should make use of *observation* $\mathcal{I}_{t'_k}$, i.e., a function of $X_{t'_k}$. On the other hand, the time epoch that the AP broadcasts a backoff rate β is the end of the event denoted by t_k, not t'_k. Therefore, the AP also takes into account how many devices would join during the period of success or collision at time t_k in Fig. 1. However, it is not possible for the AP to know $X_{t'_k}$ exactly based on $\mathcal{I}_{t_k}$, we estimate its mean, i.e., $\mathbb{E}[X_{t'_k}]$ such that the AP broadcasts $\beta = 1/(2\mathbb{E}[X_{t_k}])$ by including the number of devices newly joining up to t_k. Accordingly, the main idea of the proposed algorithm is summarized as follows. Once an event occurs, the AP is to estimate the *mean* number of backlogged devices $\mathbb{E}[X_{t'_k}]$ at time t'_k according to observation $\mathcal{I}_{t'_k}$. By adding the number of devices that would join the backlog during the period of success or collision, the AP broadcasts $\beta = 1/(2\mathbb{E}[X_{t_k}])$.

Let us derive each line of the proposed algorithm. When an event occurs at t'_{k-1} (or t'_k), i.e., the beginning of success or collision, an idle period is over as shown in Fig. 1. Thus, the length of an idle period is $I = t'_k - t_{k-1}$ (or $t'_{k+1} - t_k$). This is given in line 3 in Algorithm 1. Next, in line 4, λ indicates the average (estimated) rate of packet arrivals. We use a first autoregressive model with weighting factor $\theta \in (0, 1)$. It is based on the assumption that in the steady state, the mean rate of packet arrivals (flow-in) is equal to the mean rate of successful packet transmission (flow-out).

The update equations in lines 6 and 8 are related to estimate $E[X_{t'_k}|I = t]$ based on observation I in line 3. To be specific, suppose that the system has $X_{t'_k} = i$ backlogged devices. Just after observing I, we are interested in finding $E[X_{t'_k}|I = t]$. For the sake of simplicity, let us drop the subscript t'_k of $X_{t'_k}$

$$E[X|I = t] = \frac{\sum_{m=0}^{\infty} m \Pr[X = m, I = t]}{f_I(t)} = \frac{\sum_{m=0}^{\infty} m f_I(t|X = m) \Pr(X = m)}{f_I(t)} \quad (40)$$

where $f_I(t|X = m)$ is given in (5), and $f_I(t)$ denotes the unconditional PDF of I. In (40), let us assume that the number of backlogged devices be a Poisson process with mean α (devices), i.e.,

$$\Pr[X = m] = \frac{\alpha^m}{m!} e^{-\alpha} \quad (41)$$

which will be examined in Section IV. Then, we can write the joint probability that an idle period is $I = t$ when the system has m backlogged devices as

$$f_I(t, X = m) = f_I(t|X = m) \Pr[X = m] = m\beta e^{-m\beta t} \frac{\alpha^m}{m!} e^{-\alpha} = \beta \frac{\left(\alpha e^{-\beta t}\right)^m}{(m-1)!} e^{-\alpha}. \quad (42)$$

We can find $f_I(t)$ in (40) as

$$f_I(t) = \sum_{m=0}^{\infty} f_I(t, X = m) = \alpha\beta e^{-\beta t} \sum_{m=0}^{\infty} \frac{\left(\alpha e^{-\beta t}\right)^{m-1}}{(m-1)!} e^{-\alpha} = \alpha\beta e^{-\left(\alpha\left(1-e^{-\beta t}\right)+\beta t\right)}. \quad (43)$$

Using (42), we can obtain $E[X, I = t]$ as

$$\begin{aligned} E[X, I = t] &= \sum_{m=0}^{\infty} m f_I(t, X = m) \\ &= \alpha\beta e^{-\beta t} \sum_{m=0}^{\infty} m \frac{\left(\alpha e^{-\beta t}\right)^{m-1}}{(m-1)!} e^{-\alpha} \\ &= \alpha\beta e^{-\beta t} \sum_{m=0}^{\infty} (m-1+1) \frac{\left(\alpha e^{-\beta t}\right)^{m-1}}{(m-1)!} e^{-\alpha} \\ &= \alpha\beta e^{-(\beta t+\alpha)} \\ &\quad \times \left[\sum_{m=0}^{\infty} \frac{\left(\alpha e^{-\beta t}\right)^{m-1}}{(m-2)!} + \sum_{m=0}^{\infty} \frac{\left(\alpha e^{-\beta t}\right)^{m-1}}{(m-1)!} \right] \\ &= \alpha\beta e^{-\beta t-\alpha} \left(\alpha e^{-\beta t} + 1\right) e^{\alpha e^{-\beta t}} \end{aligned} \quad (44)$$

which is in fact the numerator of (40). From (43) and (44), we finally get

$$E[X|I = t] = \frac{E[X, I = t]}{f_I(t)} = 1 + \alpha e^{-\beta t}. \quad (45)$$

It can be seen in (45) that when a long idle period is observed, the mean of backlogged devices α exponentially decreases. Moreover, since an idle period eventually comes to an end, a 1 added in (45) accounts for that there is at least one backlogged device at the end of an idle period. Notice that α is the estimated average number of backlogged devices. As the idle period I becomes shorter with collisions, the impact of the exponential term is reduced, i.e., close to one. This causes α to increase as one is added, which subsequently results in an increase in the mean of the backoff period, $1/\beta$. However, if a shorter idle period is observed and is followed by success, the algorithm tries to keep the previous estimation on α in line 4.

Algorithm 1 Proposed Online Adaptive Backoff Algorithm

1: Initialize α, λ, at epoch $k = 0$.
2: **if** either success or collision begins **then**
3: $I = t'_k - t_{k-1}$
4: $\lambda = \theta\lambda + (1-\theta)\frac{1}{t_k - t_{k-1}}\mathbb{I}(S)$.
5: **if** success **then**
6: $\alpha = \alpha e^{-\beta I} + \lambda \cdot T$.
7: **else**
8: $\alpha = 1 + \alpha e^{-\beta I} + \lambda \cdot C$.
9: **end if**
10: **end if**
11: Broadcast $\beta = 1/(2\alpha)$ at t_k.

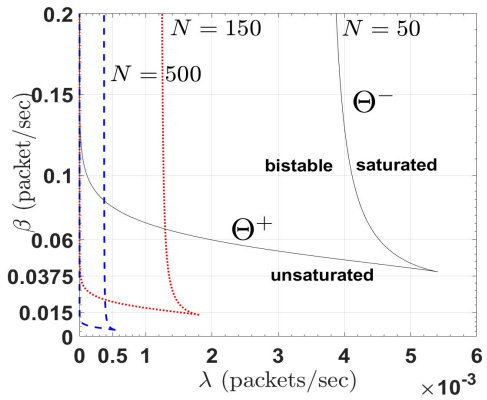


Fig. 6. Bifurcation sets of the unslotted ALOHA system.

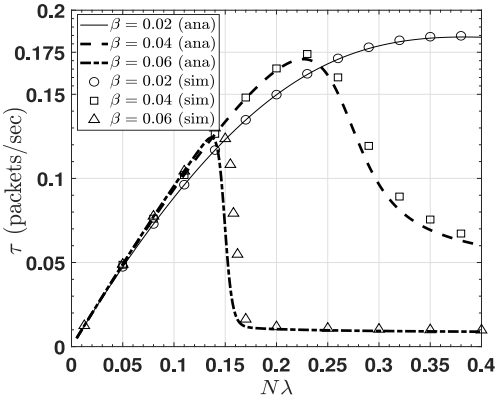
Returning to Algorithm 1, we build lines 6 and 8 upon (45) as follows. While utilizing (45) with observation I, upon the success we subtract one from it as line 6; that is, one device leaves the backlogged devices. Since the number of devices newly joining the backlog is $\lambda \cdot T$ during the period of success, it is added in line 4. On the other hand, upon collision, $\lambda \cdot C$ is added under the assumption that the new arrivals proportional to the length of the current collision period C, and we do not subtract one from (45). Finally, β is broadcast as in line 11.

IV. Numerical Studies

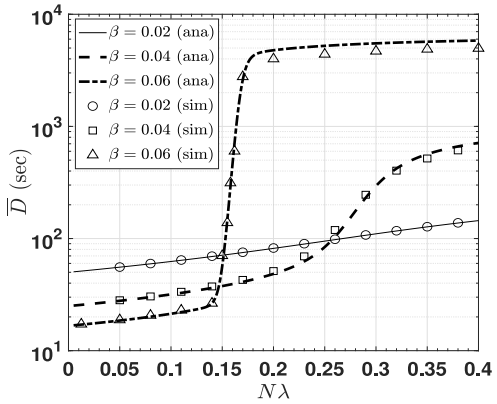
We build a discrete-event simulator for unslotted ALOHA systems with MATLAB, set the simulation run length to 10^7 (s), and collect simulation results as time averaged. In the figures, lines depict analytical results, whereas symbols show simulation results. Notice that the time unit is normalized by one packet length of T (s) long. When the packet length will be 1 ms long, the unit of delay should be also changed to ms.

In Fig. 6, we examine the bistability of unslotted ALOHA systems as N increases. It shows three operating regions, such as unsaturated stable, bistable, and saturated one. As N increases, the unsaturated stable region gets shrunk quickly, while the saturated region expands. Specifically, let us compare $N = 50$ with 150; the system becomes bistable and further saturated, if λ increases given a high backoff rate β, e.g., $\beta = 0.06$. On the other hand, given a packet arrival rate λ such as $\lambda = 0.004$ for $N = 50$, if β increases, the system goes bistable and even further saturated. Note that $\lambda = 3 \times 10^{-3}$ and $\beta = 0.06$ used in Fig. 3 fall into the bistable region. For $N = 500$, in order to run the system in the unsaturated region, the maximum of the mean rate of packet arrivals and the mean interval of backoff are limited up to $\lambda \approx 0.525 \times 10^{-3}$ and $1/\beta \approx 243.9$, respectively. It implies that for a large population of devices, very low source activity and long backoff interval can keep the system stable.

In Fig. 7(a) and (b), the system throughput τ (packets/s) and mean RA delay (s/packet) are presented for $N = 50$ according to the operating regions. Let us recall that for $\beta = 0.02$, as λ increases, the operating region never comes across the bistable region in Fig. 6; that is, the system turns into the saturated region from the unsaturated stable region without



(a)



(b)

Fig. 7. Performance of the unslotted ALOHA system. (a) System throughput with $N = 50$. (b) Mean access delay with $N = 50$.

passing through the bistable region. In such cases, it can be seen that the throughput and mean RA delay increase moderately without a sudden collapse in throughput or jump in delay. However, as β increases, it can be observed in Fig. 6 that the system enters the bistable region with small λ and gets across a larger bistable region before reaching the saturated region. Accordingly, steeper collapse or abrupt jump is found in throughput and delay in Fig. 7(a) and (b). For $N\lambda < 0.125$ the throughput seems insensitive to β in Fig. 7(a), while the access delay is much improved as β gets smaller. As expected, it can be concluded that β must be adaptive to traffic load $N\lambda$.

In Fig. 8(a) and (b), we examine the RA delay violation probability $\Pr[D \geq d]$ for ERB and URB. In order to make the mean backoff interval of ERB equal to that of URB, we set the window size of URB as $U = 2\lceil(1/\beta)\rceil$, where $\lceil x \rceil$ is the ceiling function that takes the least integer greater than or equal to x. Fig. 8(a) presents $\Pr[D \geq d]$ for $N = 50$ and $\lambda = 0.035$, which means that the total offered load is $N\lambda = 0.175$ (highly loaded). It is notable that the analysis agrees well with simulation results. This implies that the analysis based on ERB is valid to get p_s for URB. It might be also valid when the mean backoff interval of an arbitrary backoff algorithm, which generates i.i.d. intervals, is matched. Especially for $d \leq U$, the delay violation probability of ERB is lower than that of URB. However, it becomes higher for $d > U$. In Fig. 8(b) we compare the delay violation probability $\Pr[D \geq 30]$ of ERB

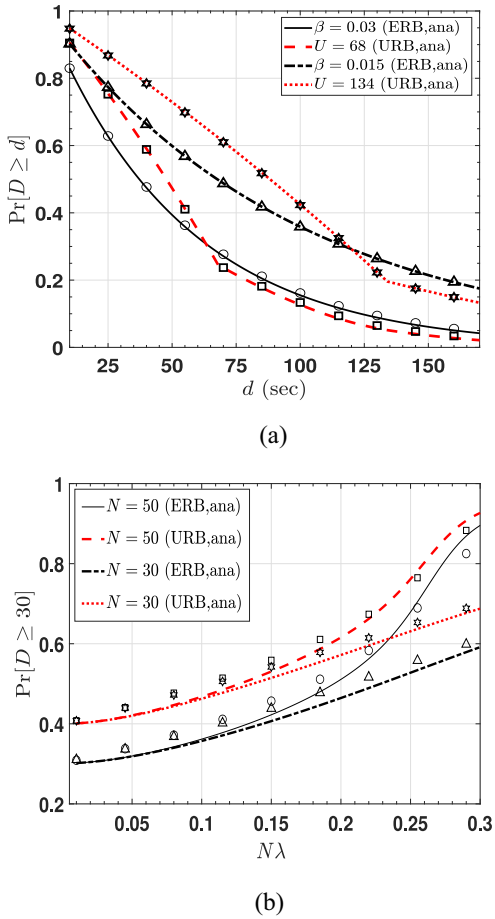


Fig. 8. Delay violation probability. (a) Delay violation probability with various backoff intervals. (b) Delay violation probability with various population sizes.

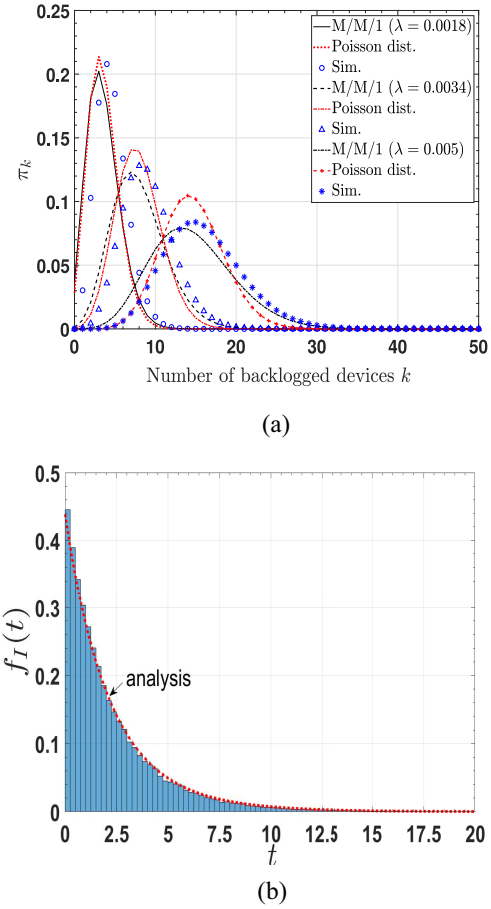
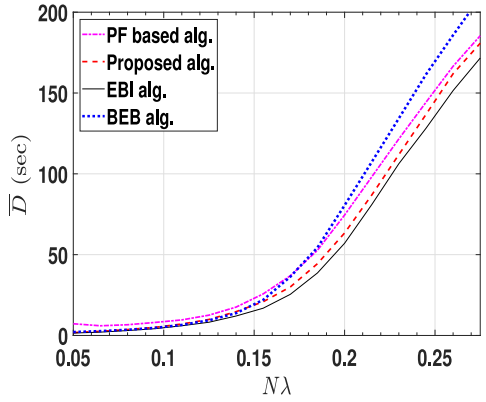


Fig. 9. Validation of underlying assumptions. (a) State distribution for unsaturated stable systems. (b) Unconditional pdf of idle period (43).

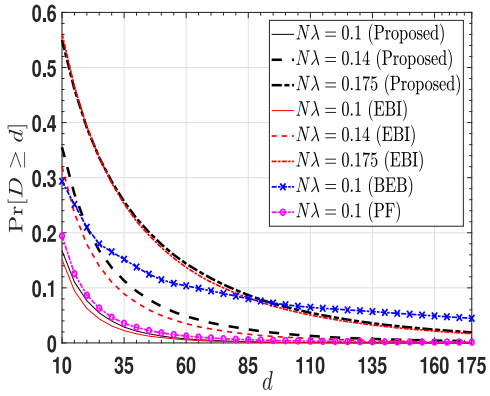
with $\beta = 0.04$ with that of URB with $U = 50$. Since d is less than U, $\Pr[D \geq 30]$ of URB is higher than that of ERB. For $N\lambda < 0.15$, whether the system has $N = 30$ or 50, the delay violation probability is not much different. As it is much more highly loaded, the system with a large population gets saturated.

In Fig. 9(a) and (b), we validate the assumption on Poisson process for the number of backlogged devices for the system with $N = 50$ and $\beta = 0.03$. Since the maximum throughput of unslotted ALOHA systems is $0.5e^{-1} = 0.18$, we can think of three traffic loads $N\lambda = 0.09$, 0.17, and 0.25 as moderate, high, and extremely high, respectively. The state probabilities from (9), i.e., M/M/1 approximation, are compared to Poisson distribution with mean $\overline{m} = 3.5627$ and 7.9403, 14.6131, where $\overline{m}$ is the mean obtained from the analysis. They are further compared to the distribution obtained from simulation. It can be seen that they resemble each other. Moreover, in Fig. 9(b), we compare (43) with a histogram of idle periods obtained from simulations. Since Poisson distribution of the system with $\lambda = 0.005$ in Fig. 9(a) shows slight difference against simulation, we compare its unconditional pdf $f_I(t)$ with simulation in Fig. 9(b). We use the analytical result for the mean α in (43). They show quite good agreement. Thus, we come to a conclusion that Poisson assumption on backlog size seems reasonably good enough to use.

In Figs. 10(a) and 11(a), the performance of the proposed backoff algorithm is examined. First, Fig. 10(a) depicts the mean of RA delay of the proposed backoff algorithm in comparison with the other three algorithms for $N = 100$. The one named EBI is to use the exact backlog size information X_{t_k} such that devices can set their backoff interval by using $1/(2X_{t_k})$. The one named "PF" is the PF-based algorithm in [14]. The algorithm named "BEB" is a BEB algorithm, which works as follows. At the kth retransmission, the device picks up a retransmission interval from a continuous uniform random variable in the interval $[0, \min(W_0 2^{k-1}, 2^{10})]$ and $W_0 = 2$. It can be seen that the mean of RA delay of the proposed algorithm is quite close to that of the EBI algorithm. While the proposed algorithm shows better performance than the PF algorithm, it is noteworthy that the PF-based algorithm in [14] does not take into account estimation on new packet arrivals nor subtract one successfully transmitted packet from the estimated backlog size upon success. While the BEB algorithm seems as good as PF-based and the proposed algorithm for low traffic loads $N\lambda < 0.18$, we shall see in Fig. 10(b) that the delay violation is the worst. With PF-based, the proposed, and EBI algorithms, some sudden jump in $\overline{D}$ is not observed as the traffic load increases. It can be expected that the system transits from unsaturated stable to saturated one. We thus come to a conclusion that bistability is eliminated by three algorithms.



(a)

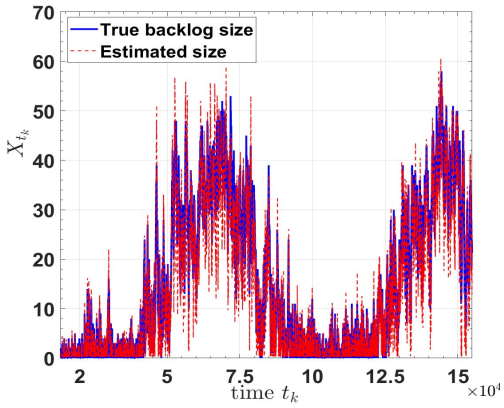


(b)

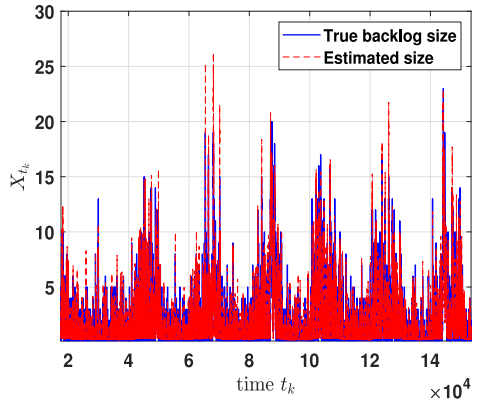
Fig. 10. Performance of the proposed algorithm. (a) Delay performance of the proposed algorithm. (b) Delay violation probability of the proposed algorithm.

Fig. 10(b) presents the delay violation probabilities for the proposed and EBI algorithms for three traffic loads, such as $N\lambda = 0.1$, 0.14, and 0.175, when the system has $N = 100$. Especially for $N\lambda = 0.1$ (low traffic load), we also compare PF-based and BEB algorithms. It can be seen that the performances of the proposed and EBI algorithms are very close. Even for $N\lambda = 0.175$ they almost overlap. Notice that the delay violation probabilities for the traffic load $\lambda N = 0.175$ for $N = 50$ are depicted in Fig. 8(a) when ERB and URB are used. For $N\lambda = 0.1$, the delay violation probability of the PF-based algorithm is as good as the proposed algorithm. It is notable that the BEB algorithm shows a large delay violation probability for $N\lambda = 0.1$, which is worse than the proposed algorithm for $N\lambda = 0.14$.

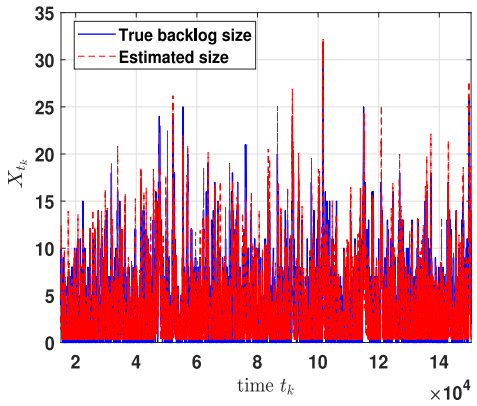
Finally, Fig. 11(a)–(c) shows a sample path of X_{t_k} to see how the proposed algorithm keeps track of the number of backlogged devices over time for the system with $N = 100$. The blue solid line depicts X_{t_k}, i.e., the true backlog size at t_k, while the red dashed line shows the estimation on it in each figure. Let us recall that the proposed algorithm estimates the mean of the true backlog size. In Fig. 11(a), new packet arrivals vary over time as follows. In the beginning, it is set to $\lambda = 0.5 \times 10^{-3}$ to each device. Every 10^4 s, λ is raised by 0.5×10^{-3}, which is equal to overall traffic load 0.05. When λ reaches 0.003 (overall traffic load is



(a)



(b)



(c)

Fig. 11. Tracking performance of the proposed algorithm. (a) Homogeneous time-varying Poisson arrivals. (b) Inhomogeneous time-varying Poisson arrivals. (c) Static poisson arrivals with $N\lambda = 0.175$ and $N = 100$.

0.3), it is reduced by 0.5×10^{-3} again every 10^4 s. This is called homogeneous time-varying Poisson arrivals in this article since N devices have the same mean. On the other hand, in Fig. 11(b), 50 devices have a fixed $\lambda = 0.001$, while the other 50 devices have a time-varying λ that changes back and forth between 0.001 and 0.0025 every 10^4 s. Accordingly, the overall packet arrivals vary between 0.1 and 0.35. This is called inhomogeneous time-varying Poisson arrivals. To see the difference between the actual backlog size and the

estimated one, the mean absolute error (MAE), normalized root mean square error (NRMSE), and mean absolute percentage error (MAPE) are observed. They are found as 3.71, 5.3173, and 2.9597×10^{-6} in Fig. 11(a), and 1.28, 1.82, and 5.2497×10^{-6} in Fig. 11(b), respectively. Finally, in Fig. 11(c), λ does not change at all and is set to 1.75×10^{-3}. For static arrival rate, our proposed algorithm shows MAE, NRMSE, and MAPE as 0.8480, 1.1734, and 6.7593×10^{-6} for $N\lambda = 0.125$ with $N = 100$, and 1.8658, 2.4302, and 3.1912×10^{-6} for $N\lambda = 0.175$ [in Fig. 11(c)], respectively. It can be seen that the proposed algorithm can track reasonably well the backlog size of the system with time-varying and static arrival rates.

V. Conclusion

This work has developed a performance analysis model for unslotted ALOHA systems for LPWAN by using an M/M/1 queueing model. With the model, we have also derived RA delay distributions for ERB and URB algorithms, which are numerically demonstrated. As a result, when the mean backoff interval of two algorithms is equal, RA delay violation of URB seems larger that of ERB, for as long as the RA delay constraint is less than the first window size of URB. Furthermore, the bistability of unslotted ALOHA systems has been investigated. It has been shown that bistability can take place depending on the population size, retransmission rate, and new packet arrival rate. It resulted in a sudden collapse of system performance. To prevent bistability, we have shown that a throughput maximizing backoff algorithm can eliminate bistability and have designed an online adaptive backoff algorithm to achieve it. As future work, it is interesting to analyze the performance of unslotted ALOHA systems with variable length packets.

References

[1] M. R. Palattella *et al.*, "Internet of Things in the 5G era: Enablers, architecture, and business models," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 3, pp. 510–527, Mar. 2016.

[2] U. Raza, P. Kulkarni, and M. Sooriyabandara, "Low power wide area networks: An overview," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 855–873, 2nd Quart., 2017.

[3] J. Metzner, "On improving utilization in ALOHA networks," *IEEE Trans. Commun.*, vol. C-24, no. 4, pp. 447–448, Apr. 1976.

[4] N. Abramson, "The throughput of packet broadcasting channels," *IEEE Trans. Commun.*, vol. COM-25, no. 1, pp. 117–128, Jan. 1977.

[5] S. Bellini and F. Borgonovo, "On the throughput of an ALOHA channel with variable length packets," *IEEE Trans. Commun.*, vol. COM-28, no. 11, pp. 1932–1935, Nov. 1980.

[6] A. L. Mcbridge, "Application of renewal theory to unslotted ALOHA in VSAT networks," *IEEE Trans. Commun.*, vol. 36, no. 5, pp. 617–621, May 1988.

[7] M. J. Ferguson, "An approximate analysis of delay for fixed and variable length packets in an unslotted Aloha channel," *IEEE Trans. Commun.*, vol. COM-25, no. 7, pp. 644–654, Jul. 1977.

[8] M. J. Ferguson, "A bound and approximation of delay distribution for fixed-length packets in an unslotted ALOHA channel and a comparison with time division multiplexing (TDM)," *IEEE Trans. Commun.*, vol. COM-25, no. 1, pp. 136–139, Jan. 1977.

[9] D. Sant, "Throughput of unslotted ALOHA channels with arbitrary packet interarrival time distributions," *IEEE Trans. Commun.*, vol. COM-28, no. 8, pp. 1422–1425, Aug. 1980.

[10] H. Kobayashi, Y. Onozato, and D. Huynh, "An approximate method for design and analysis of an ALOHA System," *IEEE Trans. Commun.*, vol. COM-25, no. 1, pp. 148–157, Jan. 1977.

[11] J. A. Roberts and T. J. Healy, "Packet radio performance over slow Rayleigh fading channels," *IEEE Trans. Commun.*, vol. COM-28, no. 2, pp. 279–286, Feb. 1980.

[12] V. C. M. Leung and R. W. Donaldson, "Effects of channel errors on the delay-throughput performance and capacity of ALOHA multiple access systems," *IEEE Trans. Commun.*, vol. COM-34, no. 5, pp. 497–501, May 1986.

[13] R. J. van der Vleuten, W. C. van Etten, and H. P. A. van den Boom, "Optimal controlled ALOHA for two-way data communication in a cable television network," *IEEE Trans. Commun.*, vol. 42, no. 7, pp. 2453–2459, Jul. 1994.

[14] J.-B. Seo and H. Jin, "Optimally controlled pure Aloha systems for wireless sensor networks," *IEEE Commun. Lett.*, vol. 21, no. 17, pp. 2460–2463, Nov. 2017.

[15] Z. J. Haas and J. Deng, "On optimizing the backoff interval for random access schemes," *IEEE Trans. Commun.*, vol. 51, no. 12, pp. 2081–2090, Dec. 2003.

[16] H. Wang and A. O. Fapojuwo, "Design and performance evaluation of successive interference cancellation-based pure Aloha for Internet-of-Things networks," *IEEE Internet Things J.*, vol. 6, no. 4, pp. 6578–6592, Aug. 2019.

[17] I. B. Arun and T. G. Venkatesh, "Order statistics based analysis of pure ALOHA in channels with multipacket reception," *IEEE Commun. Lett.*, vol. 17, no. 10, pp. 2012–2015, Oct. 2013.

[18] A. Baiocchi and F. Ricciato, "Analysis of pure and slotted ALOHA with multi-packet reception and variable packet size," *IEEE Commun. Lett.*, vol. 22, no. 7, pp. 1482–1485, Jul. 2018.

[19] B. Blaszczyszyn and P. Mühlethaler, "Stochastic analysis of non-slotted Aloha in wireless ad-hoc networks," in *Proc. IEEE INFOCOM*, 2010, pp. 2570–2578.

[20] A. Munari, P. Mähönen, and M. Petrova, "A stochastic geometry approach to asynchronous Aloha full-duplex networks," *IEEE/ACM Trans. Netw.*, vol. 25, no. 6, pp. 3695–3708, Dec. 2017.

[21] J. E. Marengo, D. L. Farnsworth, and L. Stefanic, "A geometric derivation of the Irwin-Hall distribution," *Int. J. Math. Math. Sci.*, vol. 2017, pp. 296–300, Sep. 2017.

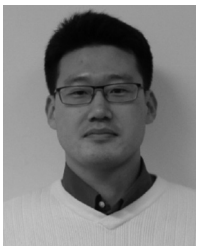
[22] A. B. Carleial and M. E. Hellman, "Bistable behavior of ALOHA-type systems," *IEEE Trans. Commun.*, vol. COM-23, no. 4, pp. 401–410, Apr. 1975.

[23] Y. Onozato and S. Noguchi, "On the thrashing cusp in slotted ALOHA systems," *IEEE Trans. Commun.*. vol. COM-33, no. 11, pp. 1171–1182, Nov. 1985.

[24] 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.

[25] R. Nelson, "Stochastic catastrophe theory in computer performance modeling," *J. Assoc. Comput. Mach.*, vol. 34, no. 3, pp. 661–685, 1987.

[26] A. Fuchs, *Nonlinear Dynamics in Complex Systems*. Heidelberg, Germany: Springer-Verlag, 2013.



Jun-Bae Seo (Member, IEEE) received the B.S. and M.Sc. degrees in electrical engineering from Korea University, Seoul, South Korea, in 2000 and 2003, respectively, and the Ph.D. degree from the University of British Columbia (UBC), Vancouver, BC, Canada, in 2012.

From 2003 to 2006, he was a member of the research staff with the Electronics and Telecommunications Research Institute, Daejeon, South Korea, carrying out research on IEEE 802.16 systems. He was also a Postdoctoral Fellow with UBC until 2014. From 2015 to August 2019, he was an Assistant Professor with the Indian Institute of Technology Delhi, New Delhi, India. From August 2019 to August 2020, he was a Research Professor with Hanyang University (ERICA Campus), Ansan, South Korea. Since September 2020, he has been with the Department of Information and Communication Engineering, Gyeongsang National University, Tongyeong, South Korea. His research interests include stochastic modeling and optimizing queueing systems with applications to wireless mobile and computer communications networks.

Bang Chul Jung (Senior Member, IEEE) received the B.S. degree in electronics engineering from Ajou University, Suwon, South Korea, in 2002, and the M.S. and Ph.D. degrees in electrical and computer engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2004 and 2008, respectively.

He was a Senior Researcher/Research Professor with KAIST Institute for Information Technology Convergence from January 2009 to February 2010. From March 2010 to August 2015, he was a Faculty Member with Gyeongsang National University (Tongyeong Campus), Tongyeong, South Korea. He is currently a Professor with the Department of EE, Chungnam National University, Daejeon. His research interests include 6G wireless communications, wireless IoT communications, statistical signal processing, information theory, wireless localization, interference management, radar signal processing, spectrum sharing, multiple antennas, multiple access techniques, radio resource management, machine learning, and GNSS receiver signal processing.

Prof. Jung was a recipient of the 5th IEEE Communication Society Asia–Pacific Outstanding Young Researcher Award in 2011, the KICS Haedong Young Scholar Award in 2015, and the 29th KOFST Science and Technology Best Paper Award in 2019. He has been serving as an Associate Editor for *IEEE Vehicular Technology Magazine* since May 2020 and *IEICE Transactions on Fundamentals of Electronics, Communications, and Computer Sciences* since 2018.

Hu Jin (Senior Member, IEEE) received the B.E. degree in electronic engineering and information science from the University of Science and Technology of China, Hefei, China, in 2004, and the M.S. and Ph.D. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2006 and 2011, respectively.

From 2011 to 2013, he was a Postdoctoral Fellow with The University of British Columbia, Vancouver, BC, Canada. From 2013 to 2014, he was a Research Professor with Gyeongsang National University (Tongyeong Campus), Tongyeong, South Korea. Since 2014, he has been with the Division of Electrical Engineering, Hanyang University, Ansan, South Korea, where he is currently an Associate Professor. His research interests include medium-access control and radio resource management for random access networks and scheduling systems considering advanced signal processing and queueing performance.