FASA: Accelerated S-ALOHA Using Access History for Event-Driven M2M Communications

Huasen Wu, *Student Member, IEEE*, Chenxi Zhu, *Member, IEEE*, Richard J. La, *Member, IEEE*, Xin Liu, *Member, IEEE*, and Youguang Zhang, *Member, IEEE*

***Abstract*—Supporting massive device transmission is challenging in machine-to-machine (M2M) communications. Particularly, in event-driven M2M communications, a large number of devices become activated within a short period of time, which in turn causes high radio congestions and severe access delay. To address this issue, we propose a Fast Adaptive S-ALOHA (FASA) scheme for random access control of M2M communication systems with bursty traffic. Instead of the observation in a single slot, the statistics of consecutive idle and collision slots are used in FASA to accelerate the tracking process of network status that is critical for optimizing S-ALOHA systems. With a design based on drift analysis, the estimate of the number of the active devices under FASA converges fast to the true value. Furthermore, by examining the T-slot drifts, we prove that the proposed FASA scheme is stable as long as the average arrival rate is smaller than e^{-1}, in the sense that the Markov chain derived from the scheme is geometrically ergodic. Simulation results demonstrate that under highly bursty traffic, the proposed FASA scheme outperforms traditional additive schemes such as PB-ALOHA and achieves near-optimal performance in reducing access delays. Moreover, compared to multiplicative schemes, FASA shows its robustness under heavy traffic load in addition to better delay performance.**

***Index Terms*—Adaptive S-ALOHA, drift analysis, machine-to-machine (M2M) communications, random access control, stability analysis.**

I. Introduction

MACHINE-TO-MACHINE (M2M) communication or machine-type communication (MTC) is expected to be one of the major drivers of cellular networks [1] and has become one of the focuses in 3GPP [2], [3]. Behind the proliferation of M2M communication, the congestion problem in M2M communication is becoming a big concern. The device density of M2M communication is much higher than that in traditional human-to-human (H2H) communication [2], [4]. For example, it is expected in [4] that 1000 devices/km^2 are deployed for environment monitoring and control. What is worse, in event-driven M2M applications, many devices may be triggered almost simultaneously and attempt to access the base station (BS) through the Random Access Channel (RACH) [5]. These event-driven M2M applications, such as remote control and alarm systems [3], [6], can be delay-sensitive. However, high burstiness of the traffic can result in congestion and increase access delays, which motivates our research.

Substantial efforts have been made in 3GPP to alleviate the radio congestion caused by M2M communications. Access Class Barring (ACB), which is a mechanism proposed for Radio Access Network (RAN) overload control [7], has also been accepted in M2M communications [3]. In ACB, devices are divided into several access classes, and devices configured for ACB are required to perform the access barring check before establishing a connection. This checking process is probability-based in the LTE/LTE-A system, i.e., a device attempts to undergo a RACH procedure by randomly sending a connection request with a given probability broadcast by the BS. The BS adjusts this probability to control the offered load. Recently, Extended Access Barring (EAB) mechanism is proposed to provide M2M-specific enhancement of the legacy ACB mechanism [3], and its practical barring procedures are still under development. In addition to EAB, there are other works trying to modify the RACH in LTE for M2M overload control [8], [9]. For example, in [8], the authors propose a cooperative ACB scheme to balance traffic load among BSs in a heterogeneous multitier cellular network. With cooperation among BSs, the congestion level can be reduced and the access delay can be significantly improved. In [9], the authors propose a code-expended scheme that combines preambles transmitted in multiple subframes for higher capacity. However, a critical step in implementing these schemes is to estimate the number of active devices, i.e., devices triggered by an event, and optimize the transmission probability. This problem is even more important in event-driven M2M applications, which are characterized by highly bursty traffic.

Essentially, the random access methods such as ACB are derivatives of slotted-ALOHA (S-ALOHA), which is widely applied for random access control. To address the instability issue of S-ALOHA [10], plenty of work has been done to tune the protocol parameters for stability, which is briefly summarized in Section II. In these schemes, the previous outcome is applied to estimate the network status and adjust protocol parameters. A drift analysis is then used to design the schemes and prove their stability [11]–[13]. However, these schemes usually rely on the assumption that the traffic can be modeled as a Poisson process and only apply the observation in the previous slot for estimating the number of active devices. Due to the burstiness, this assumption cannot be justified in the context of M2M applications, and the observation in a single slot is not satisfactory for adjusting the protocol parameters in time. Thus, we try to make full use of the information provided

Manuscript received May 06, 2012; revised November 04, 2012; accepted December 20, 2012; approved by IEEE/ACM TRANSACTIONS ON NETWORKING Editor E. Ekici. This work was supported in part by the NKBRP under Grant No. 2010CB731803, the NSFC under Grant No. 60921001, the NSF under Grant No. 0917251, a Fujitsu Research Grant, and the UC Davis Chancellor's Fellowship.

H. Wu and Y. Zhang are with the School of Electronic and Information Engineering, Beihang University, Beijing 100191, China (e-mail: huasenwu@gmail.com).

C. Zhu and R. J. La are with University of Maryland, College Park, MD 20742 USA.

X. Liu is with University of California, Davis, Davis, CA 95616 USA.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNET.2013.2241076

by the past outcomes for improving the performance of access control under bursty traffic.

In this paper, we study adaptive S-ALOHA scheme for event-driven M2M communications and provide rigorous analysis about its stability. As our main contribution, we propose a Fast Adaptive S-ALOHA (FASA) scheme for the random access control of M2M devices. A key characteristic of FASA is that the access outcomes in the past slots, in particular, consecutive idle or collision slots, are collected and applied to estimate the number of active devices. This enables a fast update of transmission probability under highly bursty traffic and thus reduces access delays. Furthermore, we prove the stability of FASA under bursty traffic. This is accomplished by examining the T-slot drifts of the network status, which are the expected changes of the network status, such as the number of active devices and its estimate, over T slots. Under interrupted Poisson traffic model [14], we show that the T-slot drifts of FASA have the required properties for stabilizing the scheme, and the system is stable when the arrival rate is less than e^{-1}. Numerical results demonstrate that using the FASA scheme, the transmission probability of S-ALOHA can be adjusted quickly, and access delays can be reduced close to the theoretical lower bound under highly bursty traffic.

The remainder of the paper is organized as follows. Section II summarizes the related work. In Section III, we present the system model, including the bursty traffic model for the event-driven M2M communications. In Section IV, after analyzing the limitations of traditional fixed-step-size policies, we propose the FASA scheme and design the parameters in the scheme based on drift analysis. Then, in Section V, we study the T-slot drifts of FASA and prove its stability. In Section VI, simulation results are presented to evaluate the performance of the proposed scheme, which is also compared to the theoretical optimum with full knowledge case and two traditional adaptive schemes. Finally, we conclude our paper in Section VII.

II. Related Work

A. Radio Congestion Control in M2M Communications

Because of the high density of devices and event-driven nature of their communication needs, a large number of M2M devices may simultaneously attempt to access the BS and result in radio congestion on the RACH. A classical scheme for reducing congestion is the backoff-based scheme [5], where devices retransmit after a backoff time if they have a collision. The backoff-based scheme can improve the performance under low congestion level, but it cannot solve the high-level congestions [15], [16]. In comparison, access-barring-type schemes are effective in reducing radio congestion [15]. Hence, the ACB mechanism is accepted as a baseline solution for overload control in M2M communications, and EAB is being developed to enhance the legacy ACB mechanism in M2M applications [3], [7]. To implement these mechanisms, the setting of operation parameters, e.g., transmission probability, should be adjusted according to the network status and is left to service providers. However, estimating the network status is challenging especially in event-driven M2M communications due to traffic burstiness. By using the statistics of access outcomes in the past, the proposed FASA scheme can track the network status quickly and reduce access delays in M2M communications with highly bursty traffic.

In addition to the efforts in 3GPP, there are a few publications addressing this issue. Since a group-based feature appears in many M2M applications, some researchers propose hierarchical architectures, such as a grouping scheme [17] and relay schemes [18], [19], for alleviating the radio congestion on the RACH. In these architectures, group heads are selected for collecting the messages from the group members. However, efficient schemes for communications between the group heads and their members are required to reduce the access delay. FASA can be used in these architectures and is expected to provide performance improvements of random access.

B. Adaptive S-ALOHA Schemes

S-ALOHA is a basic scheme for random access control, and many random access methods in wireless networks, including ACB, can be viewed as its derivatives. However, the uncontrolled S-ALOHA scheme is unstable in the sense that the number of backlogged devices grows unbounded over time with probability one [10]. The instability issue of S-ALOHA should be dealt with before being implemented in practical networks.

Two typical classes of schemes, additive and multiplicative schemes [12], [20]–[23], have been proposed for stabilizing S-ALOHA systems. In these schemes, the estimate of the network status is updated in an additive or multiplicative manner, respectively, and the transmission probability is adjusted accordingly. However, as discussed in more detail later, traditional additive schemes such as Pseudo Bayesian ALOHA (PB-ALOHA) [22] estimate the number of active devices based on the access outcome in the previous slot, but cannot adjust the transmission probability in a timely manner under highly busty traffic. On the other hand, because of the exponential increment in consecutive collision slots or decrement in consecutive idle slots, multiplicative schemes [23], e.g., Q-Algorithm [24] and its enhanced version Q^+-Algorithm [25], can track the network status in a short period. However, the throughput suffers in these schemes due to the fluctuations in the estimation [23]. Therefore, we aim to design adaptive schemes that can track the network status fast under bursty traffic while retaining high throughput.

By using access outcomes in consecutive slots in the past, the proposed FASA scheme provides a better tradeoff between the convergence time and the estimation fluctuation and thus reduces the access delays under bursty traffic. In our previous work [26], we propose a preliminary version of FASA based on some intuitive approximations and show its desirable properties through numerical simulations. However, no rigorous analysis on the stability of FASA is presented in [26], and we will investigate the stability of FASA based on drift analysis techniques in this paper.

C. Drift Analysis for Stabilization of S-ALOHA

Drift analysis has been used to study the stability of adaptive S-ALOHA schemes [12], [13], [20], [21], [27], [28]. The network status, which is represented by the number of active devices and its estimate, can be viewed as a stochastic sequence, and the convergence of adaptive schemes can be analyzed by examining the drift. In [12], it is shown that when the network status drift satisfies certain conditions, the system is stable in the sense that the returning time can be bounded

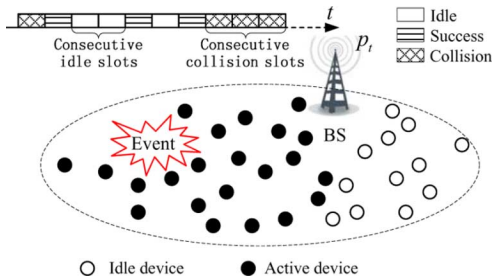


Fig. 1. S-ALOHA-based access control of M2M communications.

with high probability. Using the conclusion in [12], the most related work [13] studies the stability of PB-ALOHA scheme by defining a Lyapunov function to represent the network status and examining its drift. Later in [21], by applying the method of stochastic Lyapunov function, sufficient conditions of recurrence and nonrecurrence of Markov chains are derived, and the throughput of additive adaptive S-ALOHA schemes is analyzed by checking the conditions. In all the work mentioned above, the schemes update the parameters based on the observation in the previous slot. Thus, the 1-slot drifts, i.e., drifts between two adjacent slots, are sufficient for studying the stability of the systems. When access outcomes in multiple slots are used, however, the T-slot drifts are required to deal with the memory of our scheme. Unlike 1-slot drifts, calculating the exact T-slot drifts is impractical due to a large number of possible outcomes in consecutive T slots. Hence, we resort to approximations and obtain desirable properties of the T-slot drifts for showing the stability of FASA.

III. System Model

A. S-ALOHA Based Random Access

We consider a cellular network based M2M communication system for event detection. As shown in Fig. 1, the system consists of a BS and a large number of M2M devices. At any time, a device can be in one of two states: *idle* or *active*. Devices stay in the idle state until triggered by an event, and then go to the active state. The active devices attempt to undergo the RACH procedure for the initial network entry by sending connection requests.

We model the RACH procedure as an S-ALOHA system. Time is divided into slots, and each slot is indexed by an integer $t \in \mathbb{Z}_+ = \{0, 1, 2, \ldots\}$. We assume that all M2M devices are configured for ACB. Hence, in each slot t, active devices transmit connection requests with a given probability p_t, which is broadcast by the BS at the beginning of the slot. We assume that all connection requests are transmitted on a single channel. This corresponds to the case where all devices access the BS with a single preamble in LTE/LTE-A [5], [29]. The proposed scheme can be extended to multichannel system by borrowing the idea of [30], where a multichannel PB-ALOHA scheme is proposed.

In slot t, there may be zero, one, or more than one devices transmitting on the channel. We note that the transmission power of the request is set based on an open-loop estimation with full compensation for the path loss [29]. However, for the sake of tractability, we ignore the impact of power ramping and assume the transmission power of each device is fixed during the RACH procedure. Hence, we assume identical received power for all devices and use an ideal collision channel model, where a transmitted request will be successfully received by the BS if and only if no other requests are being transmitted in the same slot. Let Z_t denote the access outcome in slot t, and $Z_t = 0, 1$, or c depending on whether t is an *idle*, *success*, or *collision* slot. We assume that energy detection techniques [31], [32] are applied at the BS to distinguish the idle and collision slots, and thus the BS can tell which of the three possible outcomes has occurred. The BS broadcasts a message identifying the access outcome at the end of the slot. Devices that suffer a collision can retransmit their requests in the following slots. Let L represent the maximum number of retransmissions. We set $L = \infty$ when analyzing the stability of the proposed scheme in Section V, and set its value according to the suggestions of 3GPP when evaluating the access performance in Section VI.

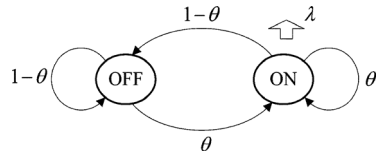


Fig. 2. Simplified interrupted Poisson process.

The BS coordinates the random access by adaptively adjusting the transmission probability p_t. At the end of slot t, the BS decides the transmission probability for next slot based on the past access outcomes $\{Z_0, Z_1, \ldots, Z_t\}$, i.e.,

$$p_{t+1} = \Gamma_t(Z_0, Z_1, \ldots, Z_t).$$

The objective of the BS is to maximize the throughput and minimize the access delays, where the throughput is the average number of successful accesses in a unit time, and the access delay experienced by a successful device is the delay between the time of activation and the completion of the random access. It is well known that when the number of active devices $N_t \geq 1$, using a transmission probability $p_t = 1/N_t$ in slot t maximizes the throughput of the S-ALOHA system [22]. However, the BS does not know N_t and has to obtain its estimate $\hat{N}_t$ based on the access outcomes in the past. The main focus of our paper is the estimation of N_t.

B. Traffic Model

In order to capture the burstiness of event-driven M2M traffic, instead of a traditional Poisson process, the arrival process is modeled as an *interrupted* Poisson process (IPP), which was suggested for simulating overflow traffic [14].

IPP can be viewed as a Poisson process modulated by a random switch and will be discretized according to the slotted structure of the scheme. We focus on a simplified version of the IPP model, where the event process is a Bernoulli process, as shown in Fig. 2. Specifically, let Y_t and A_t denote the number of events happening and the number of devices triggered in slot t, respectively. Assume that events happen independently in each slot with fixed probability θ. Namely, in each slot t, $\Pr(Y_t = 1) = \theta$ and $\Pr(Y_t = 0) = 1 - \theta$. In addition, assume that the number of triggered devices follows a Poisson distribution with mean λ when an event happens, and no devices become active otherwise, i.e., $A_t \sim \mathcal{P}(\lambda)$ when $Y_t = 1$ (ON state), and $A_t = 0$ when $Y_t = 0$ (OFF state). Therefore, random

variables $\{A_t : t \in \mathbb{Z}_+\}$ are independent and identically distributed (i.i.d.), and the long-term arrival rate can be calculated as

$$\bar{\lambda} = \Pr(Y_t = 0) \cdot 0 + \Pr(Y_t = 1) \cdot \lambda = \theta\lambda. \tag{1}$$

We will design and analyze an adaptive S-ALOHA scheme based on this model because it captures the ON–OFF nature of events and provides tractability. We note that the classical Poisson process is included as a special case of the simplified IPP model when $\theta = 1$. Indeed, as we will show later, the scheme proposed in this paper can stabilize the network under more general traffic models. It is worth noting that a limitation of the IPP model is that devices are triggered simultaneously when an event occurs, which is not valid in practice. To evaluate the performance of the proposed scheme for more realistic traffic, we also present simulation results in Section VI for beta-distributed pattern traffic, which is suggested by 3GPP for M2M communications [3], [4].

IV. Fast Adaptive S-ALOHA

In this section, we first analyze the drift of traditional fixed-step-size estimation schemes and show their limitations. Then, we propose a fast adaptive scheme, referred to as FASA.

A. Drift Analysis of Fixed-Step-Size Estimation Schemes

Many schemes have been proposed to estimate the number of active devices [11], [33]. Reference [11] mentions that a unified framework of additive schemes is proposed and studied in [20], where the estimate $\hat{N}_t$ is updated by the recursion

$$\hat{N}_{t+1} = \max\left\{1, \hat{N}_t + \sum_{i\in\{0,1,c\}} a_i I(Z_t = i)\right\} \tag{2}$$

where a_0, a_1, and a_c are constants and $I(\mathcal{E})$ is the indicator function of event $\mathcal{E}$.

With the estimate $\hat{N}_t$, the BS sets the transmission probability to $p_t = 1/\hat{N}_t$ for all active devices. Thus, the offered load, which is the expected number of devices attempting to access the channel, is $\rho = N_t p_t = N_t/\hat{N}_t$. To stabilize the S-ALOHA system, $\hat{N}_t$ needs to drift toward the actual number of active devices N_t, especially when N_t is large. When $N_t = n$ and $\hat{N}_t = \hat{n}$, the drift of the estimate can be calculated as follows [11], [20], [21]:

$$\begin{aligned} &E[\hat{N}_{t+1} - \hat{N}_t | N_t = n, \hat{N}_t = \hat{n}] \\ &\quad = (a_0 - a_c)\left(1 - \frac{1}{\hat{n}}\right)^n + (a_1 - a_c)\frac{n}{\hat{n}}\left(1 - \frac{1}{\hat{n}}\right)^{n-1} + a_c \\ &\quad \to (a_0 - a_c)e^{-\rho} + (a_1 - a_c)\rho e^{-\rho} + a_c \stackrel{\text{def}}{=} \phi(\rho) \end{aligned}$$

as $n \to \infty$, with $n/\hat{n} = \rho$ fixed.

With properly chosen parameters a_i $(i = 0, 1, c)$, the drift $\phi(\rho)$ satisfies $\phi(\rho) < 0$ if $\rho < 1$ and $\phi(\rho) > 0$ if $\rho > 1$, and the estimate $\hat{N}_t$ will drift toward the true value. However, these fixed-step-size schemes are not suitable for systems with bursty traffic. When the estimate $\hat{N}_t$ deviates far away from the true value N_t, we have $\lim_{\rho\to 0}\phi(\rho) = a_0$ and $\lim_{\rho\to\infty}\phi(\rho) = a_c$. These limits indicate that the drift tends to be constant even when the deviation is large, which could result in a large tracking time. Thus, it is necessary to design fast estimation schemes for event-driven M2M communications.

B. Framework of FASA

As discussed above, fixed-step-size estimation schemes such as PB-ALOHA may not be able to adapt in a timely manner for systems with bursty traffic because they always use a constant step size even when the estimate deviates far away from the true value. We note that in addition to the access outcome in the previous slot, the access outcomes in several consecutive slots can be used for improving the estimation performance as they reveal additional information about the true value. Intuitively, collisions in several consecutive slots are likely caused by a significant underestimation, i.e., $\hat{N}_t \ll N_t$, and the BS should aggressively increase its estimate. In contrast, several consecutive idle slots may indicate that the estimate $\hat{N}_t \gg N_t$, and it should be reduced aggressively.

Motivated by this intuition, we propose the FASA scheme. Let $K_{0,t}$ and $K_{c,t}$ be the numbers of consecutive idle and collision slots leading up to slot t, respectively. The estimate $\hat{N}_t$ is updated as follows:

$$\hat{N}_{t+1} = \begin{cases} \max\{1, \hat{N}_t - 1 - h_0(\nu)(K_{0,t} \wedge k_m)^\nu\}, & \text{if } Z_t = 0 \\ \hat{N}_t, & \text{if } Z_t = 1 \\ \hat{N}_t + \frac{1}{e-2} + h_c(\nu)(K_{c,t} \wedge k_m)^\nu, & \text{if } Z_t = c \end{cases} \tag{3}$$

where $k_m > 1$ is an integer, $K \wedge k_m = \min\{K, k_m\}$, $\nu > 0$ is the exponential factor we use to control the adjusting speed, and $h_0(\nu)$ and $h_c(\nu)$ are functions of ν that guarantee the right direction of the estimation drift and will be designed in Section IV-C. In order to make the scheme implementable and its stability analysis tractable, we bound the update step size with k_m in this paper, which is different from that we proposed in [26]. However, the two schemes behave similarly if we choose a sufficiently large k_m.

FASA belongs to the class of additive schemes. Under FASA, the estimate $\hat{N}_t$ stays unchanged in a success slot, is decremented in an idle slot, and is incremented in a collision slot. However, the key difference of FASA lies in the use of outcomes in consecutive slots to adjust the update step size. Specifically, for a sufficiently large k_m, the step size will be incremented during consecutive idle or collision slots, and the estimation process is sped up. Furthermore, with proper functions $h_0(\nu)$ and $h_c(\nu)$ that guarantee the convergence, the estimate $\hat{N}_t$ will converge quickly to the true value, and thus the response time under FASA is expected to be reduced.

C. Design of $h_0(\nu)$ and $h_c(\nu)$

The functions $h_0(\nu)$ and $h_c(\nu)$ are critical to guarantee the convergence of FASA. Next, we design $h_0(\nu)$ and $h_c(\nu)$ by analyzing the drift of $\hat{N}_t$. According to the structure of FASA, the evolution of $\hat{N}_t$ depends not only on the access outcome in the previous slot, but also on the outcomes in the past k_m slots. Therefore, unlike the fixed-step-size schemes, accurate drift analysis is difficult for FASA because of its memory of access outcomes over multiple slots. Thus, to make the problem tractable, we resort to an approximation based on Lemma 1, which indicates the feasibility of approximating the distribution of access outcomes in the past k_m slots using the network status in the previous slot.

Lemma 1: For given $\epsilon > 0$, Δn, and $\Delta\hat{n}$, there exists some $M > 0$ such that, for any $(n, \hat{n}) \in H_M$, where $H_M = \{(n, \hat{n}) : n \geq M \text{ or } \hat{n} \geq M, n + \Delta n \geq 0, \hat{n} + \Delta\hat{n} \geq 1\}$, we have

$$\left| e^{-\rho} - (1 - \frac{1}{\hat{n} + \Delta\hat{n}})^{n+\Delta n} \right| \leq \epsilon \tag{4}$$

and

$$\left| \rho e^{-\rho} - \frac{n + \Delta n}{\hat{n} + \Delta\hat{n}} (1 - \frac{1}{\hat{n} + \Delta\hat{n}})^{n+\Delta n-1} \right| \leq \epsilon \tag{5}$$

where $\rho = n/\hat{n}$.

Proof: The proof of this lemma follows the classic approach of showing the limit and is omitted here due to space limit. Details can be found in the technical report [34]. □

Note that the distribution of the access outcome Z_s in slot s is decided by N_s and $\hat{N}_s$. In addition, for any given $s \in \{t - k_m, t - k_m + 1, \ldots, t - 1\}$, we have $|\hat{N}_s - \hat{N}_t| \leq k_m/(e-2) + k_m^{1+\nu} \max\{h_0(\nu), h_c(\nu)\}$, and $|N_s - N_t|$ is bounded with high probability, i.e., $\Pr(|N_s - N_t| \leq B) \to 1$ as $B \to \infty$. Thus, according to Lemma 1, when N_t and $\hat{N}_t$ are known and at least one of them is sufficiently large, the distribution of access outcomes in the past k_m slots as well as those of $K_{0,t}$ and $K_{c,t}$ can be approximated. Therefore, in this section, we do not assume any knowledge of $K_{0,t}$ or $K_{c,t}$ in slot t and approximate the drift of $\hat{N}_t$ conditioned on N_t and $\hat{N}_t$ utilizing the bounds in Lemma 1 instead. In addition, we assume that $\hat{N}_t$ is large enough in the past k_m slots, and hence we can approximate $\max\{1, x\}$ in (3) as x in the analysis later. An analysis provided in Section V will show that the design with these approximations stabilizes the proposed scheme.

Suppose that in slot t, the number of active devices and its estimate are $N_t = n$ and $\hat{N}_t = \hat{n}$, respectively, and thus the offered load $\rho = n/\hat{n}$. When n or $\hat{n}$ is large, the drift of estimate $\hat{N}_t$ can be approximated as

$$E[\hat{N}_{t+1} - \hat{N}_t | N_t = n, \hat{N}_t = \hat{n}] \approx \sum_{i \in \{0,1,c\}} q_i(\rho) E[\Delta\hat{N}_t | (i, n, \hat{n})] \tag{6}$$

where $q_0(\rho) = e^{-\rho}$, $q_1(\rho) = \rho e^{-\rho}$, and $q_c(\rho) = 1 - q_0(\rho) - q_1(\rho)$ are the approximate probabilities of an *idle*, *success*, and *collision* slot, respectively; $E[\Delta\hat{N}_t | (i, n, \hat{n})]$ $(i = 0, 1, c)$ are the expected changes in $\hat{N}_t$ resulting from the corresponding updates when $(Z_t, N_t, \hat{N}_t) = (i, n, \hat{n})$.

Obviously, $E[\Delta\hat{N}_t | (1, n, \hat{n})] = 0$ since the estimated number remains unchanged when a packet is successfully transmitted in slot t. On the other hand, rather than maintaining the record of access outcomes in the past slots, $K_{0,t}$ and $K_{c,t}$ are treated as random variables. Hence, we estimate $E[\Delta\hat{N}_t | (0, n, \hat{n})]$ and $E[\Delta\hat{N}_t | (c, n, \hat{n})]$ based on the approximate distributions of $K_{0,t}$ and $K_{c,t}$.

First, we calculate the drift of estimate in an idle slot, i.e., $E[\Delta\hat{N}_t | (0, n, \hat{n})]$. Suppose that no packet is transmitted in slot t, then the estimated number will be reduced by $1 + h_0(\nu)(K_{0,t} \wedge k_m)^\nu$. Therefore

$$\begin{aligned} &E[\Delta\hat{N}_t | (0, n, \hat{n})] \\ &= -\sum_{k_0=1}^{k_m - 1} [1 + h_0(\nu) k_0^\nu] \Pr[K_{0,t} = k_0 | (0, n, \hat{n})] \\ &\quad - [1 + h_0(\nu) k_m^\nu] \Pr[K_{0,t} \geq k_m | (0, n, \hat{n})]. \end{aligned} \tag{7}$$

Notice that $K_{0,t} = k_0$ $(1 \leq k_0 < k_m)$ holds when slots $t - k_0 + 1, t - k_0 + 2, \ldots, t - 1$ are all idle while slot $t - k_0$ is not. Thus, for $1 \leq k_0 < k_m$, we have

$$\begin{aligned} &\Pr[K_{0,t} = k_0 | (0, n, \hat{n})] \\ &= \Pr[Z_{t-k_0} \neq 0 | (Z_{t-k_0+1}, \ldots, Z_t, N_t, N_t') = (0, \ldots, 0, n, \hat{n})] \\ &\cdot \prod_{s=t-k_0+1}^{t-1} \Pr[Z_s = 0 | (Z_{s+1}, \ldots, Z_t, N_t, N_t') = (0, \ldots, 0, n, \hat{n})]. \end{aligned}$$

According to Lemma 1, when $N_t = n$ or $\hat{N}_t = \hat{n}$ is sufficiently large, the distribution of access outcomes in the past k_m slots can be approximated as that in the previous slot, i.e.,

$$\begin{aligned} &\Pr[Z_{t-k_0} \neq 0 | (Z_{t-k_0+1}, \ldots, Z_t, N_t, N_t') = (0, \ldots, 0, n, \hat{n})] \\ &\quad \approx 1 - q_0(\rho) \end{aligned}$$

and

$$\begin{aligned} &\Pr[Z_s = 0 | (Z_{s+1}, \ldots, Z_t, N_t, N_t') = (0, \ldots, 0, n, \hat{n})] \\ &\quad \approx q_0(\rho), \quad s = t - k_0 + 1, t - k_0 + 2, \ldots, t - 1. \end{aligned}$$

Consequently, for $1 \leq k_0 < k_m$

$$\Pr[K_{0,t} = k_0 | (0, n, \hat{n})] \approx q_0^{k_0 - 1}(\rho)[1 - q_0(\rho)]. \tag{8}$$

Similarly, $K_{0,t} \geq k_m$ holds if slots $t - k_m + 1, t - k_m + 1, t - 1$ are all idle, and we can approximate the probability as

$$\Pr(K_{0,t} \geq k_m) \approx q_0^{k_m - 1}(\rho). \tag{9}$$

Substituting (8) and (9) into (7), we can calculate the drift of $\hat{N}_t$ in an idle slot as follows:

$$E[\Delta\hat{N}_t | (0, n, \hat{n})] \approx -[1 + h_0(\nu)\mu(\nu, q_0(\rho), k_m)] \tag{10}$$

where $\mu(\nu, q, k_m)$ is defined as

$$\mu(\nu, q, k_m) = \sum_{k=1}^{k_m - 1} k^\nu q^{k-1}(1 - q) + (k_m)^\nu q^{k_m - 1} \tag{11}$$

and $\mu(\nu, q_0(\rho), k_m)$ is the approximate expectation of $(K_{0,t} \wedge k_m)^\nu$ conditioned on $(Z_t, N_t, \hat{N}_t) = (0, n, \hat{n})$.

Second, we can calculate the drift of estimate in a collision slot in a similar fashion and obtain

$$E[\Delta\hat{N}_t | (c, n, \hat{n})] \approx (e - 2)^{-1} + h_c(\nu)\mu(\nu, q_c(\rho), k_m). \tag{12}$$

The drift of estimate for FASA can be approximated by substituting the expressions of $E[\Delta\hat{N}_t | (i, n, \hat{n})]$ $(i = 0, 1, c)$ into (6)

$$\begin{aligned} &E[\hat{N}_{t+1} - \hat{N}_t | N_t = n, \hat{N}_t = \hat{n}] \\ &\quad \approx -q_0(\rho)[1 + h_0(\nu)\mu(\nu, q_0(\rho), k_m)] \\ &\qquad + q_c(\rho)[(e - 2)^{-1} + h_c(\nu)\mu(\nu, q_c(\rho), k_m)] \\ &\quad \stackrel{\text{def}}{=} \varphi(\rho). \end{aligned} \tag{13}$$

In order to keep the offered load ρ in the neighborhood of the optimal value $\rho^* = 1$, it is reasonable to require that $\varphi(1) = 0$. In other words, letting $q_0^* = q_0(1) = e^{-1}$ and $q_c^* = q_c(1) = 1 - 2e^{-1}$, we expect that

$$\varphi(1) = -h_0(\nu) q_0^* \mu(\nu, q_0^*, k_m) + h_c(\nu) q_c^* \mu(\nu, q_c^*, k_m) = 0. \tag{14}$$

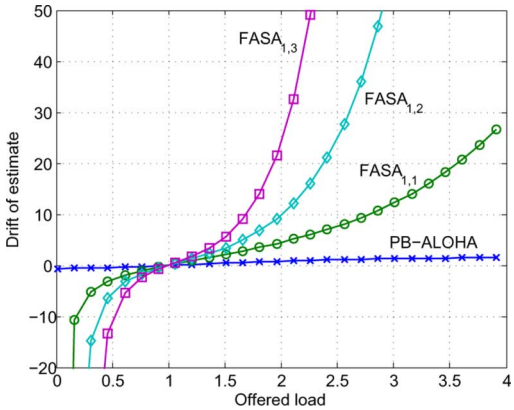


Fig. 3. Drift of estimation.

Hence, for given $k_m > 1$, to satisfy the condition in (14), we can select the following $h_0(\nu)$ and $h_c(\nu)$:

$$h_0(\nu) = \eta[q_0^* \mu(\nu, q_0^*, k_m)]^{-1} \tag{15}$$

$$h_c(\nu) = \eta[q_c^* \mu(\nu, q_c^*, k_m)]^{-1} \tag{16}$$

where $\eta > 0$ is the linear factor we use to control the adjusting speed.

The chosen $h_0(\nu)$ and $h_c(\nu)$ guarantee that $\varphi(1) = 0$ and thus provide a necessary condition for FASA to track the number of active devices. Furthermore, Theorem 1 shows a desirable property of FASA; it states that the estimated number $\hat{N}_t$ roughly drifts toward to the true value N_t, eventually yielding $\rho = N_t/\hat{N}_t \approx 1$.

Theorem 1: Given that $h_0(\nu)$ and $h_c(\nu)$ are defined in (15) and (16), respectively, the approximate drift of FASA $\varphi(\rho)$ is a strictly increasing function of ρ. In addition, $\varphi(\rho) < 0$ when $0 < \rho < 1$ and $\varphi(\rho) > 0$ when $\rho > 1$.

Proof: It is similar to the proof of [26, Proposition 1] and omitted here due to space limit. □

In order to understand better the behavior of the scheme, we now present the approximate drift of estimate for FASA with $\eta = 1$ and $\nu = 1, 2, 3$. Assume that k_m is large, and thus the distribution of $K_{i,t}$ $(i = 0, c)$ can be approximated using a geometric distribution with success probability $1 - q_i(\rho)$. In addition, for $\nu \in \mathbb{Z}_+$, $\mu(\nu, q_i(\rho), k_m)$ is approximately the νth moment of a geometrically distributed random variable with success probability $1 - q_i(\rho)$, and its closed-form expression can be obtained. Consequently, the results obtained in our previous work [26] can be applied directly. Fig. 3 shows the approximate drift of the estimate as a function of the offered load, where the subscript of FASA represents the values of (η, ν). Note that the drift of multiplicative schemes such as Q$^+$-Algorithm is not illustrated here since it depends not only on the offered load ρ, but also on the estimate $\hat{N}_t$. It can be observed from the figure that when the estimated number deviates far away from the actual number of active devices, i.e., $\rho \approx 0$ or $\rho \gg 1$, FASA adjusts its step size accordingly, while PB-ALOHA still uses the same step size. Though large step sizes can bring large fluctuations in a steady state [35], if we choose proper parameters, the FASA scheme will result in shorter adjusting times than PB-ALOHA, and thus improve the performance of M2M communication systems with bursty traffic.

V. Stability Analysis of FASA

In this section, we use drift analysis to study the stability of the proposed FASA scheme. The M2M traffic is modeled as an IPP presented in Section III. However, as we will demonstrate later, the proposed scheme can maintain stability under more general arrival processes.

Unlike traditional adaptive schemes, the access outcomes in the past consecutive slots are used in FASA to accelerate the speed of tracking, which makes it difficult to obtain the accurate drift of estimate. However, from the stability point of view, we concern mostly with the scenarios where the number of active devices or its estimate is large, and hence approximation can be applied in these cases. To deal with the issues caused by the memory property of FASA, we analyze its T-slot drifts rather than 1-slot drifts, which are introduced for analyzing traditional adaptive ALOHA schemes [12], [13], [21], [23], [28]. By constructing a virtual sequence, we show that the T-slot drifts of FASA have the properties required for stabilizing the system, and these are similar to the properties of PB-ALOHA. Therefore, with a slight modification, the Lyapunov-function-based method proposed for PB-ALOHA [13] can be used to prove the stability of FASA.

Consider the FASA scheme proposed in (3) under IPP arrival process with average arrival rate $\bar{\lambda}$. We define a sequence $X_t = (N_t, \hat{N}_t, K_t)$, where K_t represents the access outcomes in the past consecutive slots and is given by $K_0 = 0$ and, for $t > 0$

$$K_t = \begin{cases} -(K_{0,t-1} \wedge k_m), & \text{if } Z_{t-1} = 0 \\ 0, & \text{if } Z_{t-1} = 1 \\ K_{c,t-1} \wedge k_m, & \text{if } Z_{t-1} = c. \end{cases}$$

Recall that $K_{0,t-1}$ and $K_{c,t-1}$ are the numbers of consecutive idle and collision slots up to slot $t-1$. Given initial value $X_0 = (0, 1, 0)$, each component of X_t evolves as follows:

$$K_{t+1} = \begin{cases} -(|K_t| + 1) \wedge k_m, & \text{if } K_t < 0, Z_t = 0 \\ -1, & \text{if } K_t \geq 0, Z_t = 0 \\ 0, & \text{if } Z_t = 1 \\ 1, & \text{if } K_t \leq 0, Z_t = c \\ (K_t + 1) \wedge k_m, & \text{if } K_t > 0, Z_t = c \end{cases} \tag{17}$$

$$N_{t+1} = \max\{0, N_t + A_t - I(Z_t = 1)\} \tag{18}$$

$$\hat{N}_{t+1} = \begin{cases} \max\{1, \hat{N}_{t+1} - 1 - h_0(\nu)|K_{t+1}|^\nu\}, & \text{if } Z_t = 0 \\ \hat{N}_t, & \text{if } Z_t = 1 \\ \hat{N}_t + \frac{1}{e-2} + h_c(\nu)(K_{t+1})^\nu, & \text{if } Z_t = c. \end{cases} \tag{19}$$

From (17)–(19), we know that the distribution of X_{t+1} depends only on X_t, implying that X_t is a Markov chain on a countable state space $\mathbb{S}_X$. The main result of this section reveals the geometric ergodicity [12] of X_t, which is described in Theorem 2. As pointed out in [12], the geometric ergodicity is a weaker form of ergodicity and indicates the existence of steady distribution for each initial state.

Theorem 2: If $0 < \bar{\lambda} < e^{-1}$, $k_m > 1$, $h_0(\nu)$, and $h_c(\nu)$ are given by (15) and (16), then the Markov Chain X_t is geometrically ergodic.

Proof: The proof of Theorem 2 is based on drift analysis. Specifically, the proof involves three steps, which are outlined as follows and presented afterwards.

Step 1—Approximation of drifts: To deal with the impact of the memory in FASA, rather than 1-slot drifts examined in the

existing works, we study the T-slot drifts of FASA. The T-slot drifts are approximated by analyzing a virtual sequence X'_{t+s} that is constructed based on the state of X_t.

Step 2—Properties of drifts: Based on the approximation of T-slot drifts for FASA, we obtain the properties of the T-slot drifts that are used to prove the stability of the scheme.

Step 3—Stability analysis based on Lyapunov function: The Lyapunov function defined in [13] is adopted for the proposed scheme. Then, with the properties obtained in Step 2, we show the geometric ergodicity of X_t by analyzing the drift of the Lyapunov function.

1) Step 1—Approximation of Drifts: In order to analyze the stability of FASA, we evaluate the change of X_t from slot t to slot $t+T$. Let $\tilde{N}_t = \hat{N}_t - N_t$ denote the estimate error in slot t. Conditioned on the state of X_t, we define the T-slot drifts of N_t, $\hat{N}_t$, and $\tilde{N}_t$ as follows:

$$d_T(n,\hat{n},k) = E[N_{t+T} - N_t | X_t = (n,\hat{n},k)] \tag{20}$$

$$\hat{d}_T(n,\hat{n},k) = E[\hat{N}_{t+T} - \hat{N}_t | X_t = (n,\hat{n},k)] \tag{21}$$

$$\begin{aligned}\tilde{d}_T(n,\hat{n},k) &= E[\tilde{N}_{t+T} - \tilde{N}_t | X_t = (n,\hat{n},k)] \\ &= \hat{d}_T(n,\hat{n},k) - d_T(n,\hat{n},k).\end{aligned} \tag{22}$$

It is difficult to calculate the drifts defined above, and we try to obtain their properties by introducing an approximate version of X_t. Let $\{Z'_{t+s} : s \in \mathbb{Z}_+\}$ be a ternary value i.i.d. random sequence, whose distribution is given by

$$\Pr(Z'_{t+s} = i) = q_i(\rho), \qquad i = 0, 1, c$$

where $\rho = n/\hat{n}$. Then, we construct a virtual sequence $X'_{t+s} = (N'_{t+s}, \hat{N}'_{t+s}, K'_{t+s})$ based on X_t and Z'_{t+s} as follows: When $s = 0$, $X'_t = X_t = (n,\hat{n},k)$; when $s > 0$, K'_{t+s}, N'_{t+s}, and $\hat{N}'_{t+s}$ are updated in a similar way as (17)–(19), respectively, with Z_t replaced with Z'_{t+s}. However, unlike the updates in X_t, we allow N'_{t+s} to be negative and $\hat{N}'_{t+s}$ to be less than 1.

Obviously, X'_{t+s} is a Markov chain, and its transition probabilities are determined by the state of X_t. We define its T-slot drifts $d'_T(n,\hat{n},k)$, $\hat{d}'_T(n,\hat{n},k)$, and $\tilde{d}'_T(n,\hat{n},k)$ similarly to (20)–(22). Given T, we show in Lemma 2 that the drifts of X'_{t+s} can be used to approximate the drifts of X_t from slot t to slot $t+T$, when either N_t or $\hat{N}_t$ is sufficiently large.

Lemma 2: Given $T > 0$ and $\epsilon > 0$, there exists some $M > 0$, such that if $N_t = n \geq M$ or $\hat{N}_t = \hat{n} \geq M$, then

$$|d_T(n,\hat{n},k) - d'_T(n,\hat{n},k)| \leq \epsilon \tag{23}$$

$$|\hat{d}_T(n,\hat{n},k) - \hat{d}'_T(n,\hat{n},k)| \leq \epsilon. \tag{24}$$

Proof: See Appendix A. □

According to Lemma 2, for given T, the differences between the T-slot drifts of X'_{t+s} and X_{t+s} can be made as close to zero as desired by letting n or $\hat{n}$ be sufficiently large. Next, we analyze the drifts of X'_{t+s}.

1) $d'_t(n,\hat{n},K)$: Since in the virtual sequence X'_{t+s}, N'_{t+s} is allowed to be negative, we have

$$d'_T(n,\hat{n},k) = E\left[\sum_{s=0}^{T-1}(A_{t+s} - Z'_{t+s}) | X'_t = (n,\hat{n},k)\right] = T(\bar{\lambda} - \rho e^{-\rho}).$$

2) $\hat{d}'_t(n,\hat{n},K)$: In slot $t+s$, the update of $\hat{N}'_{t+s}$ depends on both K'_{t+s} and Z'_{t+s}. Notice that the sequence K'_{t+s} is a Markov chain on a finite state space $\mathbb{S}_K = \{-k_m, -k_m+1, \ldots, 0, \ldots, k_m\}$. Since the distribution of Z'_{t+s} is fixed, by showing the ergodicity of K'_{t+s}, we are able to approximate the T-slot drift by analyzing the stationary behavior of K'_{t+s}. Specifically, the transition of K'_{t+s} depends on the value of Z'_{t+s} and the 1-step transition probabilities are given by

$$p_{kj}(\rho) = \begin{cases} q_0(\rho), & \text{if } k \leq 0, j = \max\{k-1, -k_m\} \\ q_0(\rho), & \text{if } k > 0, j = -1 \\ q_1(\rho), & \text{if } j = 0 \\ q_c(\rho), & \text{if } k < 0, j = 1 \\ q_c(\rho), & \text{if } k \geq 0, j = \min\{k+1, k_m\} \\ 0, & \text{else.} \end{cases}$$

When $\rho > 0$, K'_{t+s} is irreducible and aperiodic. Thus, K'_{t+s} is ergodic and there is a unique stationary distribution. It can be verified that the stationary distribution of K'_{t+s} is

$$\boldsymbol{\pi}(\rho) = [\pi_{-k_m}(\rho), \pi_{-k_m+1}(\rho), \ldots, \pi_0(\rho), \ldots, \pi_{k_m}(\rho)]$$

where each component $\pi_k(\rho)$ ($k \in \mathbb{S}_K$) represents the stationary probability of k and is given by

$$\pi_k(\rho) = \begin{cases} q_0^{k_m}(\rho), & \text{if } k = -k_m \\ q_0^{|k|}(\rho)[1 - q_0(\rho)], & \text{if } -k_m + 1 \leq k \leq -1 \\ q_1(\rho), & \text{if } k = 0 \\ q_c^{k}(\rho)[1 - q_c(\rho)], & \text{if } 1 \leq k \leq k_m - 1 \\ q_c^{k_m}(\rho), & \text{if } k = k_m. \end{cases}$$

Using the expression of $\boldsymbol{\pi}(\rho)$, we can verify that $\varphi(\rho)$ defined in (13) represents the stationary drift of $\hat{N}'_{t+s}$, which is the 1-slot drift of $\hat{N}'_{t+s}$ when K'_{t+s} is in the steady state. With the ergodicity of K'_{t+s}, we have

$$\lim_{T\to\infty} \frac{1}{T}\hat{d}'_T(n,\hat{n},k) = \varphi(\rho).$$

Moreover, in order to use Lemma 2, we expect to find a common T such that (23) and (24) hold for given $\epsilon > 0$ and for all $(n,\hat{n},k) \in \mathbb{S}_X$. This requires the uniform convergence of $\frac{1}{T}\hat{d}'_T(n,\hat{n},k)$. In fact, by analyzing the evolution of K'_{t+s}, we can show that $\frac{1}{T}\hat{d}'_T(n,\hat{n},k)$ converges uniformly in $(n,\hat{n},k)$. First, by multiplying the transition probability matrix k_m times, we can see that for any $k, j \in \mathbb{S}_K$, the k_m-step transition probability $p_{k,j}^{(k_m)}(\rho) = \pi_j(\rho)$. Consequently, for any state $j \in \mathbb{S}_K$, we have $\Pr(K'_{t+s} = j) = \pi_j(\rho)$ for $s \geq k_m$. Thus, when $s \geq k_m$, the drift of $\hat{N}'_{t+s}$ in each slot is exactly $\varphi(\rho)$. Then, with the fact that for any $(n,\hat{n},k) \in \mathbb{S}_X$

$$\begin{aligned}&\left|\hat{d}'_{k_m-1}(n,\hat{n},k) - (k_m-1)\varphi(\rho)\right| \\ &\qquad \leq 2(k_m-1)\left[\frac{1}{e-2} + k_m^{\nu}\max\{h_0(\nu), h_c(\nu)\}\right]\end{aligned}$$

we know that as T tends to infinity, $\frac{1}{T}\hat{d}'_T(n,\hat{n},k)$ converges to $\varphi(\rho)$ uniformly in $(n,\hat{n},k)$. Thus, the difference between $\frac{1}{T}\hat{d}'_T(n,\hat{n},k)$ and $\varphi(\rho)$ can be made as close to zero as desired by choosing a common T for all $(n,\hat{n},k) \in \mathbb{S}_X$.

3) $\tilde{d}'_t(n,\hat{n},K)$: Since $\tilde{d}'_T(n,\hat{n},k) = \hat{d}'_T(n,\hat{n},k) - d'_T(n,\hat{n},k)$, we introduce the following function to approximate $\frac{1}{T}\tilde{d}'_T(n,\hat{n},k)$:

$$\begin{aligned}\psi(\rho,\bar{\lambda}) &= \varphi(\rho) - (\bar{\lambda} - \rho e^{-\rho}) \\ &= [\frac{1}{e-2} + h_c(\nu)\mu(\nu, q_c(\rho), k_m)](1 - e^{-\rho} - \rho e^{-\rho}) \\ &\quad - [1 + h_0(\nu)\mu(\nu, q_0(\rho), k_m)]e^{-\rho} + \rho e^{-\rho} - \bar{\lambda}.\end{aligned}$$

With the uniform convergence of $\frac{1}{T}\hat{d}'_T(n,\hat{n},k)$, we know that for any given $\bar{\lambda}>0$, as $T\to\infty$

$$\frac{1}{T}\tilde{d}'_T(n,\hat{n},k)\to\psi(\rho,\bar{\lambda})$$

uniformly in $(n,\hat{n},k)$.

2) Step 2—Properties of Drifts: The evolution of estimate error $\tilde{N}_t$ is critical for showing the stability of the scheme. We first show in Lemma 3 that the approximate drift $\psi(\rho,\bar{\lambda})$ has the same properties as those for the 1-slot drift of PB-ALOHA and then present the required properties of T-slot drifts of FASA in Lemma 4.

Lemma 3: Given k_m as a positive integer, $\psi(\rho,\bar{\lambda})$ has the following properties.

a) For any $\bar{\lambda}$, the function $\psi(\rho,\bar{\lambda})$ is strictly increasing in ρ.
b) For any $\bar{\lambda}\in(0,e^{-1}]$, there exists a unique $\rho=\omega(\bar{\lambda})\in(0,1]$, such that $\psi(\rho,\bar{\lambda})=0$.
c) If $\bar{\lambda}\in(0,e^{-1})$, then $\omega(\bar{\lambda})e^{-\omega(\bar{\lambda})}>\bar{\lambda}$.

Proof: See Appendix B. □

Let $\beta=\omega(\bar{\lambda})$ denote the root of $\psi(\rho,\bar{\lambda})=0$ for given $\bar{\lambda}$. Similarly to the method in [13], we partition the state space into the following four parts:

$$\begin{aligned}
S_{\gamma,M}=&\{(n,\hat{n},k):\\
&\quad n\ge M \text{ or } \hat{n}\ge M, \beta-\gamma\le\frac{n}{\hat{n}}\le 1+\gamma, k\in\mathbb{S}_K\},\\
R^-_{\gamma,M}=&\{(n,\hat{n},k):\hat{n}\ge M,\frac{n}{\hat{n}}<\beta-\gamma,k\in\mathbb{S}_K\},\\
R^+_{\gamma,M}=&\{(n,\hat{n},k):n\ge M,\frac{n}{\hat{n}}>1+\gamma,k\in\mathbb{S}_K\},\\
Q_M=&\{(n,\hat{n},k):n<M,\hat{n}<M,k\in\mathbb{S}_K\},
\end{aligned}$$

and let $R_{\gamma,M}=R^-_{\gamma,M}\cup R^+_{\gamma,M}$.

With Lemmas 2 and 3, we present the properties of the T-slot drifts in these regions in the following lemma.

Lemma 4: There exist some $\gamma>0$, $\delta>0$, $T>0$, and $M>0$, such that $5\gamma<\beta$ and

$$\tilde{d}_T(n,\hat{n},k)\le -T\delta \qquad \forall\,(n,\hat{n},k)\in S_{5\gamma,M} \tag{25}$$
$$\tilde{d}_T(n,\hat{n},k)\le -T\delta \qquad \forall\,(n,\hat{n},k)\in R^-_{\gamma,M} \tag{26}$$
$$\tilde{d}_T(n,\hat{n},k)\ge T\delta \qquad \forall\,(n,\hat{n},k)\in R^+_{\gamma,M}. \tag{27}$$

Proof: See Appendix C. □

Intuitively, according to Lemma 4, when the estimate $\hat{N}_t$ is close enough to N_t, there will be more departures than arrivals in the following T slots. On the other hand, the deviation of the estimate $\hat{N}_t$ from N_t is expected to decrease when it is larger than a certain threshold. These properties guarantee the stability of FASA, as presented in Step 3.

3) Step 3—Stability Analysis Based on Lyapunov Function: Lemma 4 shows that with a sufficiently large T, the T-slot drifts have similar properties to the drifts of PB-ALOHA. Hence, when observing the system every T slots, the Lyapunov-function-based method for PB-ALOHA can be used for analyzing the stability of FASA. Next, we provide an outline of using the Lyapunov-function-based method to prove the stability of FASA. For more details about this method, please refer to [13].

We choose T, M, γ, and δ such that (25)–(27) hold. We use the Lyapunov function defined in [13]

$$V(n,\hat{n},k)=\max\left\{n,\frac{1+3\gamma}{3\gamma}(n-\hat{n}),\frac{\beta-3\gamma}{1-\beta+3\gamma}(\hat{n}-n)\right\}$$
$$=\begin{cases}n, & \text{if } (n,\hat{n},k)\in S_{3\gamma,M}\\ \frac{1+3\gamma}{3\gamma}(n-\hat{n}), & \text{if } (n,\hat{n},k)\in R^+_{3\gamma,M}\\ \frac{\beta-3\gamma}{1-\beta+3\gamma}(\hat{n}-n), & \text{if } (n,\hat{n},k)\in R^-_{3\gamma,M}.\end{cases} \tag{28}$$

We will show that if J is sufficiently large, there exists some $\Delta>0$ such that

$$E[V(N_{t+JT},\hat{N}_{t+JT},K_{t+JT})-V(N_t,\hat{N}_t,K_t)+\Delta;\\ (N_t,\hat{N}_t,K_t)\notin Q_{M+(JT)^2}|\mathcal{F}_t]\le 0 \tag{29}$$

where $\mathcal{F}_t$ is the σ-field generated by $\{A_{s-1},N_s,\hat{N}_s,K_s:s\le t\}$. Note that for a random variable X and an event $\mathcal{E}$, the notation $E[X;\mathcal{E}|\mathcal{F}]$ stands for $E[XI(\mathcal{E})|\mathcal{F}]$.

For given $t\ge 0$ and an integer J, let

$$\tau_J=\min\{j\ge 0:\sum_{s=0}^{jT}A_{t+s}\ge JT\}.$$

Similarly to [13], we then analyze the drift of the Lyapunov function by considering the unlikely event $\{\tau_J\le J\}$ and likely event $\{\tau_J>J\}$ separately.

Using Chernoff bound [36], we can show that the following results also hold for IPP traffic:

$$\lim_{J\to\infty}\Pr(\tau_J\le J)=0 \tag{30}$$
$$\lim_{J\to\infty}E\left[l_1JT+l_2\sum_{s=0}^{JT}A_{t+s};\tau_J\le J\right]=0 \tag{31}$$

where l_1, l_2 are arbitrary given constants. Thus, for any $(n,\hat{n},k)\in\mathbb{S}_X$, as $J\to\infty$, we have

$$E[|V(N_{t+JT},\hat{N}_{t+JT},K_{t+JT})-V(N_t,\hat{N}_t,K_t)|;\\ \tau_J\le J|X_t=(n,\hat{n},k)]\to 0 \tag{32}$$

implying that this expectation can be made as close to 0 as desired by choosing a sufficiently large J.

Now consider the event $\tau_J>J$. Based on the value of $X_t=(n,\hat{n},k)$, we study the drift of the Lyapunov function in the following five cases:

a) $(n,\hat{n},k)\in S_{2\gamma,M+(JT)^2}$;
b) $(n,\hat{n},k)\in R^+_{4\gamma,M+(JT)^2}$;
c) $(n,\hat{n},k)\in R^-_{4\gamma,M+(JT)^2}$;
d) $(n,\hat{n},k)\in S_{4\gamma,M+(JT)^2}\cap R^+_{2\gamma,M+(JT)^2}$;
e) $(n,\hat{n},k)\in S_{4\gamma,M+(JT)^2}\cap R^-_{2\gamma,M+(JT)^2}$.

In any of these cases, following the approach in [13], we can show that when J is sufficiently large, there exists some $\Delta>0$, such that (29) holds.

Take case (a) as an example. According to [13, Lemma 3.4], if $X_t=(n,\hat{n},k)\in S_{2\gamma,M+(JT)^2}$, then we choose a sufficiently large J, such that $(N_{t+s},\hat{N}_{t+s},K_{t+s})\in S_{3\gamma,M}$ for all $s=0,1,\ldots,JT$, and

$$V(N_{t+jT},\hat{N}_{t+jT},K_{t+jT})=N_{t+jT},\qquad \text{for all } j\in[0,J].$$

Thus, choosing a sufficiently large J such that $\Pr(\tau_J>J)>1/2$, we have

$$\begin{aligned}&E[V(N_{t+JT},\hat{N}_{t+JT},K_{t+JT})-V(N_t,\hat{N}_t,K_t);\\ &\quad\tau_J>J|X_t=(n,\hat{n},k)]\\ &=E[N_{t+JT}-N_t;\tau_J>J|X_t=(n,\hat{n},k)]\end{aligned}$$

$$= \sum_{j=0}^{J-1} E[d_T(N_{t+jT}, \hat{N}_{t+jT}, K_{t+jT}); \tau_J > J | X_t = (n, \hat{n}, k)]$$

$$\leq -\delta TJ \Pr(\tau_J > J) \leq -\frac{\delta TJ}{2}. \quad (33)$$

Combining (32) and (33), we know that there exists some J such that (29) holds for some $\Delta > 0$.

Now, we can complete the proof by using the results on the hitting time bounds [12], [13]. Let

$$\Lambda = [M + (JT)^2] \max\left\{1, \frac{1+3\gamma}{3\gamma}, \frac{\beta - 3\gamma}{1 - \beta + 3\gamma}\right\}.$$

Note that for any $K_t \in \mathbb{S}_K$, if $V(N_t, \hat{N}_t, K_t) < \Lambda$, then $(N_t, \hat{N}_t, K_t) \notin Q_{M+(JT)^2}$ and hence (29) holds. Moreover, $\{V(N_t, \hat{N}_t, K_t), \mathcal{F}_t)\}$ is exponential-type. Then, [13, Proposition 2.1] applies. Therefore, for any initial state, the returning time $\tau_\Lambda^* = \min\{t > 0 : V(N_t, \hat{N}_t, K_t) < \Lambda\}$ is exponential-type, which implies that X_t is geometrically ergodic and concludes the proof of Theorem 2. ■

Similarly to the discussion in [13], from the proof of Theorem 2, we know that the proposed FASA scheme is stable under more general traffic, as long as the average arrival rate $\bar{\lambda} < e^{-1}$ and the traffic model satisfies the conditions in (30) and (31).

VI. Simulation Results

In this section, we evaluate the performance of the proposed scheme through simulations. We first examine the tracking performance and the effect of control parameters. We then study the delay performance of the proposed scheme, including both cases of single-event and multiple-events reporting.

We compare the performance of FASA, the ideal policy with perfect network state information, PB-ALOHA [22], and Q$^+$-Algorithm [25]. With perfect knowledge of N_t, the ideal policy sets transmission probability at $p_t = 1/N_t$ for $N_t > 0$. Thus, the ideal policy achieves the minimum access delay of S-ALOHA and serves as a benchmark in the comparison. For PB-ALOHA, we use the estimated arrival rate $\hat{\lambda}_t = e^{-1}$, as suggested in [13]. In Q$^+$-Algorithm, $\hat{N}_t$ is updated by $\hat{N}_{t+1} = \max\{1, [I(Z_t = 0)/\zeta_0 + I(Z_t = 1) + \zeta_c I(Z_t = c)]\hat{N}_t\}$, where $\zeta_0 = 2^{0.25} \approx 1.1892$ and $\zeta_c = 2^{0.35} \approx 1.2746$ are suggested in [25] for optimal performance.

A. Performance of Tracking

In order to gain insights into the operation of the estimation schemes, we treat adaptive S-ALOHA schemes as dynamic systems and study their step responses, where the number of active devices $N_t = 0$ when $t < 0$ and $N_t = n$ for all $t \geq 0$. We examine the tracking performance of schemes for $n = 200$, 500, and 1000.

1) Evolution of Estimation: Before quantitive analysis, we first show in Fig. 4 the evolution of the estimate $\hat{N}_t$ under different conditions, which helps us understand the behavior of these schemes. In this figure, average values of $\hat{N}_t$ in each slot are obtained from 4000 independent experiments, and the maximum number of consecutive slots is set at $k_m = 16$ unless it is expressly stated.

From Fig. 4, when comparing the evolution of $\hat{N}_t$ for different schemes with $n = 500$, we can see that unlike the almost linear increment under PB-ALOHA, the estimate $\hat{N}_t$ given by FASA increases slowly at the beginning, but speeds up due to consecutive collisions, i.e., increasing of K_t. When the estimate gets close to the true value, success and idle slots occur more often, and hence the increment of the estimate slows down. The estimate of Q$^+$-Algorithm follows the same trend as FASA and speeds up even faster on average than FASA because of the exponential increment. However, Q$^+$-Algorithm will result in large fluctuation especially in heavy traffic.

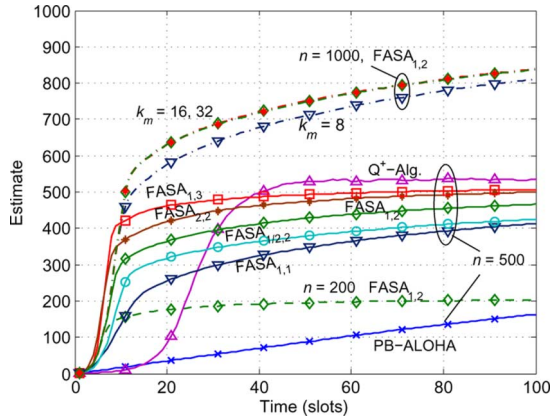


Fig. 4. Estimation process under different settings.

Next, we examine the effect of the control parameters in FASA. First, in Fig. 4, we present the evolution of $\hat{N}_t$ for $n = 500$ under FASA with different values of η and ν. From the curves, we can see that with larger η or ν, the estimate adjusts faster. However, this faster convergence of the estimates comes at the price of large fluctuations in the estimates. Second, we also present the evolution of $\hat{N}_t$ for $n = 1000$ under FASA with $k_m = 8$, 16, and 32, respectively. We can see from Fig. 4 that, while increasing k_m improves the convergence speed initially, the estimation becomes insensitive to the value of k_m beyond a certain threshold. This is because the number of consecutive idle or collision slots rarely becomes very large due to the underlying geometric distribution. In particular, K_t exceeds $k_m = 16$ only a few times in the simulation, and the evolution of estimate is almost the same with $k_m = 16$ and 32. Therefore, we set $k_m = 16$ in the following simulations.

2) Evolution of Instantaneous Throughput: Because the access delay depends on the throughput, two throughput-oriented metrics are introduced to measure the convergence speed and steady state performance: *0%–x% throughput rising time* and *stationary throughput*. The 0%–x% throughput rising time is defined as the time required for the expected throughput to rise from 0% to x% of the optimal value e^{-1}. For $x = 10$, 50, and 90, they are equal to the time required for the estimated number of active devices $\hat{N}_t$ to rise from 0% to 20.45%, 37.34%, and 65.25% of the true value n, respectively. Stationary throughput is the average throughput after the time that the expected throughput reaches 90% of the optimal value.

As shown in Table I, given x, the 0%–x% rising time of PB-ALOHA increases almost linearly in n and is much larger than that of Q$^+$-Algorithm and FASA. For instance, when $n = 500$, the 0%–50% rising time of FASA with $\eta = 1$ and $\nu = 2$ is about 1/10 of that of PB-ALOHA. Moreover, due to the aggressive update in FASA, the 0%–x% rising time increases more slowly than linearly with the number of devices n. Comparing the rising time of FASA with different η and ν, we

TABLE I
0%–x% THROUGHPUT RISING TIME (UNIT: SLOTS)

n	x	PB-ALOHA	Q^+-Alg.	$FASA_{1,1}$	$FASA_{1,3}$	$FASA_{1,2}$	$FASA_{1/2,2}$	$FASA_{2,2}$
	10	24.3	17.1	6.0	5.0	5.0	6.0	4.0
200	50	48.6	19.3	8.9	5.1	6.5	7.5	5.4
	90	125.0	24.6	20.4	9.4	13.4	18.3	10.0
	10	59.8	21.0	9.1	6.0	7.1	8.1	5.0
500	50	117.9	23.5	14.1	8.0	9.1	11.8	7.5
	90	313.1	27.5	43.9	14.5	21.3	31.9	15.6
	10	118.5	23.0	13.2	7.1	8.3	10.1	7.1
1000	50	237.0	26.6	21.6	9.4	12.0	15.0	9.0
	90	627.4	30.7	83.2	24.8	38.1	60.0	20.6

TABLE II
STATIONARY THROUGHPUT

n	PB-ALOHA	Q^+-Alg.	$FASA_{1,1}$	$FASA_{1,3}$	$FASA_{1,2}$	$FASA_{1/2,2}$	$FASA_{2,2}$
200	0.3682	0.3528	0.3662	0.3633	0.3638	0.3666	0.3603
500	0.3686	0.3529	0.3679	0.3656	0.3670	0.3676	0.3653
1000	0.3684	0.3521	0.3678	0.3663	0.3675	0.3681	0.3668

observe that the increment of η or ν results in a decrease in rising time. For Q^+-Algorithm, with multiplicative adjustment, it takes longer time than FASA to reach 10% of the optimal value, but shorter time to increase the expected throughput from 10% to 90% of the optimal value. However, aggressive adjustment of the estimate usually results in large fluctuations. This fact becomes clear when comparing the stationary throughput shown in Table II. Particularly, when the number of devices is small, e.g., $n = 200$, the fluctuation of the estimate is comparable to n, and the stationary throughput suffers. Note that the access delay is affected by both convergence speed and stationary throughput. We will see in Section VII that the performance of FASA is not sensitive to the values of η and ν.

B. Delay Performance

In this section, simulation results about the access delay of adaptive S-ALOHA schemes are presented. In some event-driven M2M applications, a response can be taken with messages from a smaller subset of the devices, and not all devices need to report an event. Thus, both the distribution of access delay for single-event reporting and the long-term average delay for repetitive-event reporting are evaluated.

1) Single-Event Reporting: We evaluate the performance of FASA in the scenarios where a large number of devices are activated to report a single event. We use the beta-distributed pattern traffic proposed in 3GPP [3]. Assume that $N_0 = n$ devices are triggered when an event is detected and their active times follow beta distribution [3]. We note that in event-driven M2M communications, devices can be triggered in a very short time, e.g., from 500 ms to 10 s [6]. Thus, we assume that the devices are activated in 2 s rather than 10 s as suggested in [3]. Similar to [3], the Physical Random Access Channel (PRACH) period is configured to be 5 ms, implying that there are 200 slots per second. The maximum number of retransmissions is $L = 10$. We study the scenarios with the number of active devices $n = 200$, 500, and 1000. When there are 30 preambles allocated to M2M applications, these are corresponding to the total number of 6000, 15 000, and 30 000, respectively, which are typical numbers of devices considered in 3GPP [3].

Table III presents the average access delay and $y\%$ access delay under different schemes. The $y\%$ access delay is the access delay achieved by $y\%$ of the active devices, and y is set to 10 and 90. From Table III, we can see that the proposed FASA scheme outperforms other schemes when the number of active devices is large. In particularly, for $n = 1000$, the performance of the proposed FASA scheme is close to the benchmark with perfect information. For PB-ALOHA, when the average number of arrivals is large, it takes a long time to track the number of active devices, and few devices can access successfully during this period. Hence, the 10% delay of PB-ALOHA is much larger than that of other schemes. For example, when $n = 1000$, the 10% delay of PB-ALOHA is 322.0 slots while that of $\mathrm{FASA}_{1,2}$ is 136.0 slots. The Q^+-Algorithm can track the number of active devices in a short time because of the exponential increment in consecutive collision slots. However, it takes longer for all the devices to access the channel under Q^+-Algorithm than under FASA due to the large estimation fluctuations in Q^+-Algorithm. Comparing the access delay achieved by FASA with different η and ν, we observe that the 10% access delay is slightly smaller for larger η or ν since they provide a quicker response. However, the larger fluctuations cause the 90% and average access delay for larger η and ν to be close to, or even larger than, that for smaller η and ν. We also note that in the case with $n = 200$ triggered in 400 slots, the performance of FASA is comparable to or slightly worse than PB-ALOHA. This is because, in this case, the fluctuation of the estimate is comparable to the number of devices in the system and the throughput is reduced.

TABLE III
$y\%$ ACCESS DELAY (UNIT: SLOTS)

n	y	Perf.Info.	PB-ALOHA	Q^+-Alg.	$FASA_{1,1}$	$FASA_{1,3}$	$FASA_{1,2}$	$FASA_{1/2,2}$	$FASA_{2,2}$
	Av.	130.5	152.1	150.9	149.2	153.9	150.8	146.7	157.8
200	10	2.0	4.0	5.0	5.0	5.0	5.0	5.0	5.0
	90	317.0	356.0	351.0	346.0	357.0	351.0	343.0	363.0
	Av.	523.2	630.5	564.2	551.3	553.5	550.1	553.5	552.6
500	10	25.0	48.0	33.0	33.0	33.0	33.0	33.0	33.0
	90	1049.0	1219.0	1114.0	1089.0	1097.0	1090.0	1094.0	1096.0
	Av.	1196.0	1470.0	1273.7	1245.4	1235.1	1236.3	1251.7	1230.1
1000	10	110.0	322.0	133.0	144.0	134.0	136.0	150.0	130.0
	90	2270.0	2663.0	2402.0	2335.0	2332.0	2326.0	2344.0	2325.0

2) Repetitive-Events Reporting With IPP Traffic: The events happen sequentially in the real system, and we now study the long-term average delay of adaptive schemes under IPP traffic with different arrival rates and bursty levels. In addition to average delay, we also define the normalized divergence to quantify the divergence from the theoretical optimum performance, i.e., $e(D) = \frac{D-D^*}{D^*}$, where D is the average delay of a particular scheme and D^* is the delay achieved by the ideal policy. As pointed out in the single-event reporting case, the performance of FASA with different η and ν is rather close. Thus, only the performance of FASA with $\eta = 1$ and $\nu = 2$ is shown here for more concise presentation.

Fig. 5 compares the average delay and the normalized divergence of adaptive schemes under different arrival rates with fixed ON-probability $\theta = 0.0001$. From Fig. 5(a), we observe that both PB-ALOHA and FASA are stable when the average arrival rate $\bar{\lambda} < e^{-1}$ and experience finite access delays. For Q^+-Algorithm, however, when the arrival rate is larger than about 0.352, the access delay grows unbounded, indicating that the algorithm with the given parameters is unstable for some $\bar{\lambda} < e^{-1}$. As pointed out in [12], the parameters in Q^+-Algorithm should be chosen according to the value of $\bar{\lambda}$ to stabilize the scheme, which is not required in either PB-ALOHA or FASA. As shown in Fig. 5(b), when the arrival rate is close to

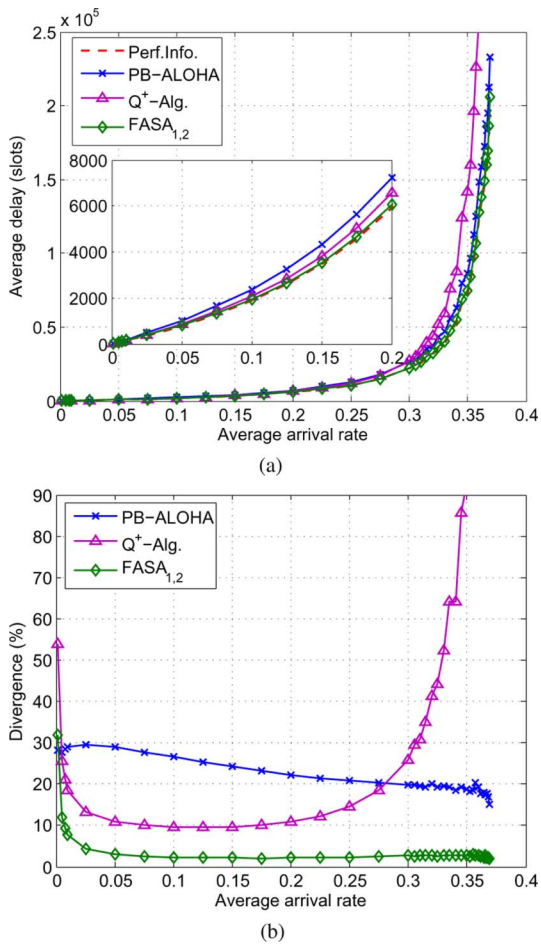


Fig. 5. Average access delay with different arrival rates. (a) Average delay ($\theta = 0:0001$) (b) Divergence ($\theta = 0:0001$).

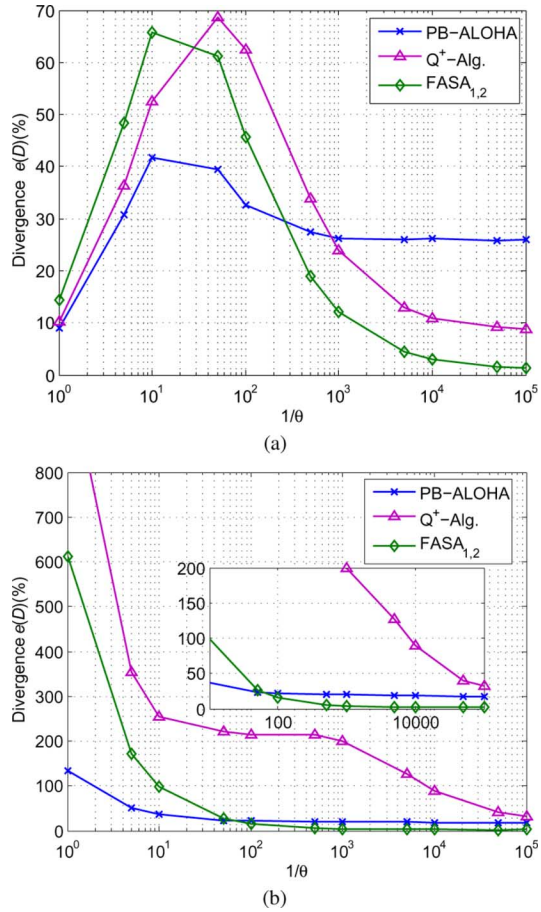


Fig. 6. Divergence of average access delay under different burstiness. (a) $\bar{\lambda} = 0.05$. (b) $\bar{\lambda} = 0.35$.

zero, the divergence of Q$^+$-Algorithm and FASA is larger than that of PB-ALOHA due to the fluctuation in estimation, while all the delays are very small. As the average arrival rate increases, the divergence of FASA decreases and the average access delay gets close to the optimal value since the estimate error becomes relatively small compared to the increasing number of active devices in the system. For $\bar{\lambda}$ larger than 0.1, the divergence of FASA is about 2.5%, while that of PB-ALOHA is about 20%.

We point out that since the average access delay could be dominated by the time waiting in the system after the estimate catches up the true value, the improvement of performance by FASA does not seem to be significant from the average delay point of view. However, as discussed in the single-event reporting cases, the 10% access delay can be improved significantly by FASA, which is very important in event-driven M2M communications.

For given $\bar{\lambda}$, the burstiness is reflected by $1/\theta$, which is the ratio between the arrival rate in ON state and the long-term average arrival rate. In order to examine the impact of burstiness, we present in Fig. 6 the divergence of access delay versus $1/\theta$ for $\bar{\lambda} = 0.05$ and 0.35. In the light traffic scenarios with $\bar{\lambda} = 0.05$, when the bursty level is low, the access delays obtained under all these schemes are close to the optimal value. The reason is that there is usually only one device triggered in one slot when the estimate is usually set to 1 after several idle slots. As the traffic becomes more bursty, the divergence first increases due to the rising time and fluctuation of estimate. Then, the divergence of FASA and Q$^+$-Algorithm decreases for highly bursty traffic because with aggressive update, they are able to track the status of the network quickly while the fluctuation becomes smaller compared to the total number of active devices. In the heavy traffic scenarios with $\bar{\lambda} = 0.35$, the divergences keep decreasing as the bursty level increases, while the proposed FASA scheme performs better than both PB-ALOHA and Q$^+$-Algorithm under highly bursty traffic. For pure Poisson traffic, i.e., $1/\theta = 1$, the burstiness is low, and the performance of $\mathrm{FASA}_{1,2}$ is close to or even worse than PB-ALOHA. However, FASA outperforms QQ^+-Algorithm under heavy traffic, regardless of the bursty level.

From the above analysis, we can see that the proposed FASA scheme outperforms PB-ALOHA under bursty traffic, e.g., when 500 devices are triggered in 400 slots. In particular, the 10% delay can be reduced significantly by FASA. On the other hand, compared to Q$^+$-Algorithm, FASA performs much better under heavy traffic with both high and low burstiness. Since the traffic in the event-driven M2M applications is expected to be bursty, we believe that our proposed scheme will perform well for these applications.

VII. Conclusion

In this paper, we proposed the FASA scheme for event-driven M2M communications. By adjusting the estimate of network status with statistics of consecutive idle and collision slots, a BS can track the number of active devices more quickly. This is a

main advantage compared to fixed-step-size additive schemes, e.g., PB-ALOHA. Moreover, we studied the stability of FASA under IPP traffic. By analyzing the T-slot drifts of FASA, we showed that without modifying the values of parameters, FASA is stable for any average arrival rate less than e^{-1}, in the sense that the system is geometrically ergodic. This property results in a much better long-term average performance under heavy traffic loads, compared to that of multiplicative schemes. In summary, the proposed scheme is an effective and stable S-ALOHA scheme and is suitable for the random access control of event-driven M2M communications and other systems characterized by bursty traffic. For future work, we will study the impact of noise and capture effect on the estimation, and a generalization of the scheme to multiclass systems, where different classes of users transmit with different probabilities.

Appendix A
Proof of Lemma 2

Because of the similarity, we present a complete analysis for (23) and briefly discuss (24) at the end.

Recall that we assume identical realization of arrival process A_{t+s} for X_{t+s} and X'_{t+s}. Hence, the difference between the T-slot drifts of N_{t+s} and N'_{t+s} is bounded by T, i.e., $|d_T(n,\hat{n},k) - d'_T(n,\hat{n},k)| \le T$. Therefore, for any given $\epsilon > 0$, there exists a $B_A > 0$ such that

$$|d_T(n,\hat{n},k) - d'_T(n,\hat{n},k)| \Pr\left(\sum_{s=0}^{T-1} A_{t+s} > B_A\right) \le \epsilon/2. \tag{34}$$

Now we consider the event $\sum_{s=0}^{T-1} A_{t+s} \le B_A$. Let $\boldsymbol{A}_{t,T} = (A_t, A_{t+1}, \ldots, A_{t+T-1})$, $\boldsymbol{Z}_{t,T} = (Z_t, Z_{t+1}, \ldots, Z_{t+T-1})$, and $\boldsymbol{Z}'_{t,T} = (Z'_t, Z'_{t+1}, \ldots, Z'_{t+T-1})$. $(\boldsymbol{A}_{t,T}, \boldsymbol{Z}_{t,T})$ and $(\boldsymbol{A}_{t,T}, \boldsymbol{Z}'_{t,T})$ have the same possible value set, which is finite and denoted by Ω. Each pair $(\boldsymbol{a},\boldsymbol{z}) = (a_0 \ldots a_{T-1}, z_0 \ldots z_{T-1}) \in \Omega$ results in corresponding drifts in both N_{t+s} and N'_{t+s}, denoted by $\Delta N_{(\boldsymbol{a},\boldsymbol{z},n,\hat{n},k)}$ and $\Delta N'_{(\boldsymbol{a},\boldsymbol{z},n,\hat{n},k)}$, respectively.

A difference between X_{t+s} and X'_{t+s} is that $N'_{t+s} \ge 0$ always holds in X_{t+s}, but $N'_{t+s} < 0$ is allowed in the virtual sequence. However, if $n \ge T$, $N'_{t+s} < 0$ will not occur in X'_{t+s}, and hence $\Delta N_{(\boldsymbol{a},\boldsymbol{z},n,\hat{n},k)} = \Delta N'_{(\boldsymbol{a},\boldsymbol{z},n,\hat{n},k)}$ for any $(\boldsymbol{a},\boldsymbol{z}) \in \Omega$ in this case. We first study the case where $n \ge T$, and analyze later the other cases where n is not large enough.

Conditioned on $\sum_{s=0}^{T-1} A_{t+s} \le B_A$, we define the following probabilities:

$$\begin{aligned}
&f_{\boldsymbol{A}}(\boldsymbol{a}) = \Pr(\boldsymbol{A}_{t,T} = \boldsymbol{a})\\
&\quad f_{(\boldsymbol{A},\boldsymbol{Z})|X}(\boldsymbol{a},\boldsymbol{z}|n,\hat{n},k)\\
&\qquad = \Pr[(\boldsymbol{A}_{t,T},\boldsymbol{Z}_{t,T}) = (\boldsymbol{a},\boldsymbol{z})|X_t = (n,\hat{n},k)]\\
&\quad f_{(\boldsymbol{A},\boldsymbol{Z}')|X'}(\boldsymbol{a},\boldsymbol{z}|n,\hat{n},k)\\
&\qquad = \Pr[(\boldsymbol{A}_{t,T},\boldsymbol{Z}'_{t,T}) = (\boldsymbol{a},\boldsymbol{z})|X'_t = (n,\hat{n},k)]\\
&\quad f_{\boldsymbol{Z}|\boldsymbol{A},X}(\boldsymbol{z}|\boldsymbol{a},n,\hat{n},k)\\
&\qquad = \Pr[\boldsymbol{Z}_{t,T} = \boldsymbol{z}|\boldsymbol{A}_{t,T} = \boldsymbol{a}, X_t = (n,\hat{n},k)]\\
&\quad f_{\boldsymbol{Z}'|\boldsymbol{A},X'}(\boldsymbol{z}|\boldsymbol{a},n,\hat{n},k)\\
&\qquad = \Pr[\boldsymbol{Z}'_{t,T} = \boldsymbol{z}|\boldsymbol{A}_{t,T} = \boldsymbol{a}, X'_t = (n,\hat{n},k)].
\end{aligned}$$

Thus, when $n \ge T$, we have

$$\begin{aligned}
&|d_T(n,\hat{n},k) - d'_T(n,\hat{n},k)| \Pr\left(\sum_{s=0}^{T-1} A_{t+s} \le B_A\right)\\
&\quad \le \Big| \sum_{(\boldsymbol{a},\boldsymbol{z})\in\Omega} [f_{(\boldsymbol{A},\boldsymbol{Z})|X}(\boldsymbol{a},\boldsymbol{z}|n,\hat{n},k)\\
&\qquad - f_{(\boldsymbol{A},\boldsymbol{Z}'|X'}(\boldsymbol{a},\boldsymbol{z}|n,\hat{n},k)]\Delta N_{(\boldsymbol{a},\boldsymbol{z},n,\hat{n},k)}\Big|\\
&\quad = \Big| \sum_{(\boldsymbol{a},\boldsymbol{z})\in\Omega} [f_{\boldsymbol{Z}|\boldsymbol{A},X}(\boldsymbol{z}|\boldsymbol{a},n,\hat{n},k)\\
&\qquad - f_{\boldsymbol{Z}'|\boldsymbol{A},X}(\boldsymbol{z}|\boldsymbol{a},n,\hat{n},k)]f_{\boldsymbol{A}}(\boldsymbol{a})\Delta N_{(\boldsymbol{a},\boldsymbol{z},n,\hat{n},k)}\Big|\\
&\quad \le C \sum_{(\boldsymbol{a},\boldsymbol{z})\in\Omega} \left|f_{\boldsymbol{Z}|\boldsymbol{A},X}(\boldsymbol{z}|\boldsymbol{a},n,\hat{n},k) - f_{\boldsymbol{Z}'|\boldsymbol{A},X}(\boldsymbol{z}|\boldsymbol{a},n,\hat{n},k)\right|
\end{aligned}$$

where C is the maximum value of $|f_{\boldsymbol{A}}(\boldsymbol{a})\Delta N_{(\boldsymbol{a},\boldsymbol{z},n,\hat{n},k)}|$ for all $(\boldsymbol{a},\boldsymbol{z}) \in \Omega$ and $k \in \{-k_m, -k_m+1, \ldots, 0, \ldots, k_m-1, k_m\}$. Moreover, by checking the state in each slot, we have

$$\begin{aligned}
&f_{\boldsymbol{Z}|\boldsymbol{A},X}(\boldsymbol{z}|\boldsymbol{a},n,\hat{n},k)\\
&\quad = \prod_{s=0}^{T-1} \Pr[Z_{t+s} = z_s|(N_{t+s},\hat{N}_{t+s}) = (n_s,\hat{n}_s)_{\boldsymbol{a},\boldsymbol{z},n,\hat{n},k}]\\
&f_{\boldsymbol{Z}'|\boldsymbol{A},X}(\boldsymbol{z}|\boldsymbol{a},n,\hat{n},k)\\
&\quad = \prod_{s=0}^{T-1} \Pr[Z'_{t+s} = z_s|(N'_{t+s},\hat{N}'_{t+s}) = (n_s,\hat{n}_s)_{\boldsymbol{a},\boldsymbol{z},n,\hat{n},k}]
\end{aligned}$$

where $(n_s,\hat{n}_s)_{\boldsymbol{a},\boldsymbol{z},n,\hat{n},k}$ is the number of active devices and its estimate in slot $t+s$, given the initial state $X_t = X'_t = (n,\hat{n},k)$, $\boldsymbol{A}_{t,T} = \boldsymbol{a}$, and $\boldsymbol{Z}_{t,T} = \boldsymbol{Z}'_{t,T} = \boldsymbol{z}$.

According to the construction of X'_{t+s}, the distribution of Z'_{t+s} is fixed, i.e., $\Pr(Z'_{t+s} = i) = q_i(\rho)$ $(i = 0, 1, c)$. On the other hand, $|N_{t+s} - n|$ and $|\hat{N}_{t+s} - \hat{n}|$ are bounded in X_{t+s}, for all $s = 1, 2, \ldots, T-1$. Therefore, using Lemma 1, we know that there exists a M'_1, such that if $n \ge M'_1$ or $\hat{n} \ge M'_1$, then for all $s = 1, 2, \ldots, T-1$, and $(\boldsymbol{a},\boldsymbol{z}) \in \Omega$

$$\begin{aligned}
&\big|\Pr[Z_{t+s} = z_s|(N_{t+s},\hat{N}_{t+s}) = (n_s,\hat{n}_s)_{\boldsymbol{a},\boldsymbol{z},n,\hat{n},k}]\\
&\quad - \Pr[Z'_{t+s} = z_s|(N'_{t+s},\hat{N}'_{t+s}) = (n_s,\hat{n}_s)_{\boldsymbol{a},\boldsymbol{z},n,\hat{n},k}]\big| \le \frac{\epsilon}{2TC|\Omega|}
\end{aligned}$$

where $|\Omega|$ is the number of elements in Ω. Hence

$$\left|f_{\boldsymbol{Z}|\boldsymbol{A},X}(\boldsymbol{z}|\boldsymbol{a},n,\hat{n},k) - f_{\boldsymbol{Z}'|\boldsymbol{A},X}(\boldsymbol{z}|\boldsymbol{a},n,\hat{n},k)\right| \le \frac{\epsilon}{2C|\Omega|}$$

and thus

$$|d_T(n,\hat{n},k) - d'_T(n,\hat{n},k)| \Pr\left(\sum_{s=0}^{T-1} A_{t+s} \le B_A\right) \le \epsilon/2. \tag{35}$$

Therefore, combining (35) with (34) implies that (23) holds when $n \ge T$, and either $n \ge M'_1$ or $\hat{n} \ge M'_1$.

Now consider the cases where $n < T$. In these cases, $(\boldsymbol{A}_{t,T}, \boldsymbol{Z}_{t,T}) = (\boldsymbol{a},\boldsymbol{z})$ is an impossible event for the $(\boldsymbol{a},\boldsymbol{z})s$ that result in $N_{t+s} < 0$. For these $(\boldsymbol{a},\boldsymbol{z})s$, we set $\Delta N_{(\boldsymbol{a},\boldsymbol{z},n,\hat{n},k)} = \Delta N'_{(\boldsymbol{a},\boldsymbol{z},n,\hat{n},k)}$, which does not affect the calculation of drifts since the corresponding probabilities are zero in X_{t+s}. Notice that for these $(\boldsymbol{a},\boldsymbol{z})s$, there is at least one component of $\boldsymbol{z}$ equal to 1. Because $\lim_{\rho\to 0} \rho e^{-\rho} = 0$,

there exists $\rho_1 \in (0,1)$, such that $\rho e^{-\rho} \le \epsilon/(2TC|\Omega|)$ for all $\rho = n/\hat{n} \le \rho_1$. Then, the difference between $f_{\boldsymbol{Z}|\boldsymbol{A},X}(\boldsymbol{z}|\boldsymbol{a},n,\hat{n},k)$ and $f_{\boldsymbol{Z}'|\boldsymbol{A},X}(\boldsymbol{z}|\boldsymbol{a},n,\hat{n},k)$ is bounded by $\epsilon/(2TC|\Omega|)$, and the same conclusion holds when $n < T$ and $n/\hat{n} \le \rho_1$.

Consequently, from the analysis above, we know that (23) holds when $n \ge M_1$ or $\hat{n} \ge M_1$, where $M_1 = \max\{T/\rho_1, M_1'/\rho_1\}$.

Finally, we discuss about (24). For given $(\boldsymbol{a},\boldsymbol{z}) \in \Omega$, denote the corresponding T-slot drifts of $\hat{N}_{t+s}$ and $\hat{N}'_{t+s}$ by $\Delta\hat{N}_{(\boldsymbol{a},\boldsymbol{z},n,\hat{n},k)}$ and $\Delta\hat{N}'_{(\boldsymbol{a},\boldsymbol{z},n,\hat{n},k)}$, respectively. In this case, $\Delta\hat{N}_{(\boldsymbol{a},\boldsymbol{z},n,\hat{n},k)} \ne \Delta\hat{N}'_{(\boldsymbol{a},\boldsymbol{z},n,\hat{n},k)}$ may occur for some $(\boldsymbol{a},\boldsymbol{z})$ when n or $\hat{n}$ is not large enough. Since for all $(\boldsymbol{a},\boldsymbol{z}) \in \Omega$, both $\Delta\hat{N}_{(\boldsymbol{a},\boldsymbol{z},n,\hat{n},k)}$ and $\Delta\hat{N}'_{(\boldsymbol{a},\boldsymbol{z},n,\hat{n},k)}$ are bounded uniformly in $(n,\hat{n},k)$, the impact of $\Delta\hat{N}_{(\boldsymbol{a},\boldsymbol{z},n,\hat{n},k)} \ne \Delta\hat{N}'_{(\boldsymbol{a},\boldsymbol{z},n,\hat{n},k)}$ can be made ignorable by making the probability of this event as close to zero as desired with the facts that $\lim_{\rho\to\infty} \rho e^{-\rho} = 0$ and $\lim_{\rho\to\infty} e^{-\rho} = 0$. Hence, similarly to (23), we can analyze the following three cases to obtain the threshold for (24):

a) $n \ge T$, $\hat{n} \ge T[1 + k_m^{\nu} h_0(\nu)] + 1$;
b) $n < T$, $\hat{n} \ge T[1 + k_m^{\nu} h_0(\nu)] + 1$;
c) $n \ge T$, $\hat{n} < T[1 + k_m^{\nu} h_0(\nu)] + 1$.

Therefore, the proof of this lemma can be concluded by choosing M as the greater threshold for (23) and (24).

Appendix B
Proof of Lemma 3

a) Let

$$\begin{aligned}
\psi^{(0)}(\rho,\bar{\lambda}) &= -e^{-\rho} + \frac{1}{e-2}(1 - e^{-\rho} - \rho e^{-\rho}) \\
\psi^{(1)}(\rho,\bar{\lambda}) &= \rho e^{-\rho} - \bar{\lambda} \\
\psi^{(2)}(\rho,\bar{\lambda}) &= -e^{-\rho} h_0(\nu)\mu(\nu, q_0(\rho), k_m) \\
&\quad + (1 - e^{-\rho} - \rho e^{-\rho}) h_c(\nu)\mu(\nu, q_c(\rho), k_m).
\end{aligned}$$

Then, $\psi = \psi^{(0)} + \psi^{(1)} + \psi^{(2)}$.

First, because

$$\frac{\partial(\psi^{(0)} + \psi^{(1)})}{\partial\rho} = e^{-\rho} + \frac{\rho e^{-\rho}}{e-2} + e^{-\rho} - \rho e^{-\rho} > 0$$

$\psi^{(0)} + \psi^{(1)}$ is strictly increasing in ρ.

In addition, using a similar approach in [26, Appendix], we can show that $\mu(\nu, q_0(\rho), k_m)$ is strictly decreasing in ρ and $\mu(\nu, q_c(\rho), k_m)$ is strictly increasing in ρ. Thus, $\psi^{(2)}(\rho,\bar{\lambda})$ and $\psi = \psi^{(0)} + \psi^{(1)} + \psi^{(2)}$ are strictly increasing in ρ.

b) For any $\bar{\lambda} \in (0, e^{-1}]$, we have $\psi(1,\bar{\lambda}) = e^{-1} - \bar{\lambda} \ge 0$, and $\psi(\rho,\bar{\lambda}) \to -h_0(\nu)(k_m)^{\nu+1} < 0$ as $\rho \to 0$. In addition, the function ψ is continuous and strictly monotonic in ρ. Thus, there is a unique solution $\rho = \omega(\bar{\lambda}) \in (0,1]$ for $\psi(\rho,\bar{\lambda}) = 0$.

c) For given $\bar{\lambda} \in (0, e^{-1})$, $\psi^{(0)}$ and $\psi^{(2)}$ are both strictly increasing in ρ. In addition, $\psi^{(0)} = \psi^{(2)} = 0$ when $\rho = 1$. Because the solution $\rho = \omega(\bar{\lambda}) < 1$, we have $\psi^{(0)}(\omega(\bar{\lambda}),\bar{\lambda}) + \psi^{(2)}(\omega(\bar{\lambda}),\bar{\lambda}) < 0$, and thus $\psi^{(1)}(\omega(\bar{\lambda}),\bar{\lambda}) = \omega(\bar{\lambda})e^{-\varphi(\bar{\lambda})} - \bar{\lambda} > 0$, i.e., $\omega(\bar{\lambda})e^{-\omega(\bar{\lambda})} > \bar{\lambda}$.

Appendix C
Proof of Lemma 4

We have shown that the T-slot drifts of X_t can be approximated by the drifts of X'_{t+s}, which can be further approximated by closed-form expressions. Thus, we first examine the properties of these expressions, and then show that the actual drifts have the same properties as their approximations by properly choosing the parameters.

Notice that $\bar{\lambda} - \rho e^{-\rho} < 0$ when $\rho = 1$ or $\rho = \beta$. In addition, it is a continuous and monotonically increasing in $\rho \in [\beta, 1]$. Therefore, there exist some δ_1 and γ such that $\bar{\lambda} - \rho e^{-\rho} < \delta_1$ for all $\rho \in [\beta - 5\gamma, 1 + 5\gamma]$. For $\psi(\rho,\bar{\lambda})$, with its strict monotonicity in ρ, we conclude that $\psi(\rho,\bar{\lambda}) < \psi(\beta-\gamma,\bar{\lambda}) < 0$ for all $\rho \in (0, \beta-\gamma)$ and $\psi(\rho,\bar{\lambda}) > \psi(1+\gamma,\bar{\lambda}) > 0$ for all $\rho \in (1+\gamma, \infty)$. Thus, there exists some δ such that

$$\bar{\lambda} - \rho e^{-\rho} \le -3\delta/2, \qquad \forall \rho \in [\beta - 5\gamma, 1 + 5\gamma] \tag{36}$$

$$\psi(\rho,\bar{\lambda}) \le -3\delta, \qquad \forall \rho \in (0, \beta - \gamma) \tag{37}$$

$$\psi(\rho,\bar{\lambda}) \ge 3\delta, \qquad \forall \rho \in (1 + \gamma, \infty). \tag{38}$$

Now, fix γ and δ. Using the uniform convergence of $\frac{1}{T}\tilde{d}'_T(n,\hat{n},k)$, we can choose a sufficiently large T, such that for any $(n,\hat{n},k) \in \mathbb{S}_X$

$$\left|\frac{1}{T}\tilde{d}'_T(n,\hat{n},k) - \psi(\rho,\bar{\lambda})\right| \le \delta \tag{39}$$

where $\rho = n/\hat{n}$.

According to Lemma 2, with the chosen T, there exists some $M > 0$, such that if $n \ge M$ or $\hat{n} \ge M$, then

$$\left|\frac{1}{T}d_T(n,\hat{n},k) - \frac{1}{T}d'_T(n,\hat{n},k)\right| \le \delta/2 \tag{40}$$

$$\left|\frac{1}{T}\hat{d}_T(n,\hat{n},k) - \frac{1}{T}\hat{d}'_T(n,\hat{n},k)\right| \le \delta/2 \tag{41}$$

and thus

$$\left|\frac{1}{T}\tilde{d}_T(n,\hat{n},k) - \frac{1}{T}\tilde{d}'_T(n,\hat{n},k)\right| \le \delta. \tag{42}$$

With $\frac{1}{T}d'_T(n,\hat{n},k) = \bar{\lambda} - \rho e^{-\rho}$, (36) and (40) together imply that for any $(n,\hat{n},k) \in S_{5\gamma,M}$, we have $\frac{1}{T}d_T(n,\hat{n},k) \le -\delta$ and, thus, (25) holds.

Similarly, (26) follows by combining (37), (39), and (42); (27) follows by combining (38), (39), and (42).

References

[1] Cisco, San Jose, CA, USA, "Cisco visual networking index: Global mobile data traffic forecast update, 2010–2015," Feb. 2011.

[2] 3GPP, "Service requirements for machine-type communications," 3GPP TS 22.368 V11.0.2, Jun. 2011.

[3] 3GPP, "Study on RAN improvements for machine-type communications," 3GPP TR 37.868 V11.0.0, Sep. 2011.

[4] Huawei and CATR, "R2-100204: Traffic model for M2M services," presented at the 3GPP TSG RAN WG2 Meeting #68bis, Jan. 2010.

[5] 3GPP, "Medium access control (MAC) protocol specification," 3GPP TS 36.321 V11.0.0, Sep. 2012.

[6] IEEE, "Machine to machine (M2M) communications technical report," IEEE 802.16p-10/0005, Nov. 2010.

[7] 3GPP, "Access class barring and overload protection," 3GPP TR 23.898 V7.0.0, Mar. 2005.

[8] S.-Y. Lien, T.-H. Liau, C.-Y. Kao, and K.-C. Chen, "Cooperative access class barring for machine-to-machine communications," *IEEE Trans. Wireless Commun.*, vol. 11, no. 1, pp. 27–32, Jan. 2012.

[9] N. K. Pratas, H. Thomsen, Č. Stefanović, and P. Povovski, "Code-expanded random access for machine-type communications," in *Proc. IEEE IWM2M*, Dec. 2012, accepted for publication.

[10] W. A. Rosenkrantz and D. Towsley, "On the instability of the slotted ALOHA multiaccess algorithm," *IEEE Trans. Autom. Control*, vol. AC-28, no. 10, pp. 994–996, Oct. 1983.

[11] F. P. Kelly, "Stochastic models of communication systems," *J. Royal Stat. Soc. B*, vol. 47, no. 3, pp. 379–395, 1985.

[12] B. Hajek, "Hitting time and occupation time bounds implied by drift analysis with applications," *Adv. Appl. Probab.*, vol. 14, pp. 502–525, Sep. 1982.

[13] J. N. Tsitsiklis, "Analysis of a multiaccess control scheme," *IEEE Trans. Autom. Control*, vol. AC-32, no. 11, pp. 1017–1020, Nov. 1987.

[14] A. Kuczura, "The interrupted Poisson process as an overflow process," *Bell Syst. Tech. J.*, vol. 52, no. 3, pp. 437–448, Mar. 1973.

[15] ZTE, "R2-104662: MTC simulation results with specific solutions," presented at the 3GPP TSG RAN WG2 Meeting #71, Aug. 2010.

[16] S.-Y. Lien, K.-C. Chen, and Y. Lin, "Toward ubiquitous massive accesses in 3GPP machine-to-machine communications," *IEEE Commun. Mag.*, vol. 49, no. 4, pp. 66–74, Apr. 2011.

[17] K.-R. Jung, A. Park, and S. Lee, "Machine-Type-Communication (MTC) device grouping algorithm for congestion avoidance of MTC oriented LTE network," *Commun. Comput. Inf. Sci.*, vol. 78, pp. 167–178, 2010.

[18] R. Y. Kim, "Efficient wireless communications schemes for machine to machine communications," *Commun. Comput. Inf. Sci.*, vol. 181, no. 3, pp. 313–323, 2011.

[19] S. D. Andreev, O. Galinina, and Y. Koucheryavy, "Energy-efficient client relay scheme for Machine-to-Machine communication," in *Proc. IEEE GLOBECOM*, 2011, pp. 1–5.

[20] V. A. Mikhailov, "Methods of random multiple access," Ph.D. dissertation, Moscow Inst. Phys. Technol., Moscow, Russia, 1979.

[21] V. A. Mikhailov, "Geometrical analysis of the stability of Markov chains in R^n and its application to throughput evaluation of the adaptive random multiple access algorithm," *Prob. Inf. Transm.*, vol. 24, no. 1, pp. 47–56, Mar. 1988.

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

[23] B. Hajek and T. van Loon, "Decentralized dynamic control of a multiaccess broadcast channel," *IEEE Trans. Autom. Control*, vol. AC-27, no. 3, pp. 559–569, Jun. 1982.

[24] *ISO/IEC 18000-6:2010 Information Technology—Radio Frequency Identification for Item Management—Part 6: Parameters for Air Interface Communications at 860 MHz to 960 MHz*, ISO/IEC, 2010.

[25] D. Lee, K. Kim, and W. Lee, "Q^+-algorithm: An enhanced RFID tag collision arbitration algorithm," *Lect.e Notes Comput. Sci., Ubiq. Intell. Comput.*, vol. 4611, no. 31, pp. 23–32, 2007.

[26] H. Wu, C. Zhu, R. La, X. Liu, and Y. Zhang, "Fast adaptive S-ALOHA scheme for event-driven machine-to-machine communications," in *Proc. IEEE VTC-Fall*, Québec City, QC, Canada, Sep. 2012, pp. 1–5.

[27] 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.

[28] M. K. Gurcan and A. Al-Amir, "Joint drift analysis for multigroup slotted ALOHA: Stability with maximum utilization," *IEEE Trans. Veh. Technol.*, vol. 50, no. 6, pp. 1415–1425, Nov. 2001.

[29] 3GPP, "Physical layer procedures," 3GPP TS 36.213 V11.0.0, Sep. 2012.

[30] D. Shen and V. O. K. Li, "Stabilized multi-channel ALOHA for wireless OFDM networks," in *Proc. IEEE GLOBECOM*, Nov. 2002, pp. 701–705.

[31] F. F. Digham, "On the energy detection of unknown signals over fading channels," *IEEE Trans. Commun.*, vol. 55, no. 1, pp. 21–24, Jan. 2007.

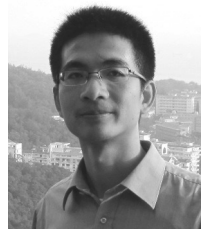
[32] Y. Chen, "Improved energy detector for random signals in Gaussian noise," *IEEE Trans. Wireless Commun.*, vol. 9, no. 2, pp. 558–563, Feb. 2010.

[33] G. A. Cunningham, III, "Delay versus throughput comparisons for stabilized slotted ALOHA," *IEEE Trans. Commun.*, vol. 38, no. 11, pp. 1932–1934, Nov. 1990.

[34] H. Wu, C. Zhu, R. La, X. Liu, and Y. Zhang, "FASA: Accelerated S-ALOHA using access history for event-driven M2M communications," 2012 [Online]. Available: http://www.cs.ucdavis.edu/~liu/preprint/FASA_TR_0111.pdf

[35] L. Merakos and D. Kazakos, "On retransmission control policies in multiple-access communication networks," *IEEE Trans. Autom. Control*, vol. AC-30, no. 2, pp. 109–117, Feb. 1985.

[36] M. Mitzenmacher and E. Upfal, *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2005.



Huasen Wu (S'12) received the B.S. degree in electronics and information engineering from Beihang University, Beijing, China, in 2007, and is currently pursuing the Ph.D. degree in electronic and information engineering at Beihang University.

From 2010 to 2012, he worked as a Visiting Student with the Department of Computer Science, University of California, Davis, CA, USA. His research interests include wireless scheduling, machine-to-machine communication, and cognitive radio.

Chenxi Zhu (M'01) received the B.S. degree in electronics engineering from Tsinghua University, Beijing, China, in 1993, and the Ph.D. degree in electronical engineering from the University of Maryland, College Park, MD, USA, in 2001.

He was with Flarion Technologies, Inc., Bedminster, NJ, USA, from 2001 to 2004. Since 2004, he has been with Fujitsu Laboratories of America, Sunnyvale, CA, USA, currently as a Senior Technical Advisor. He is on the research faculty of the Institute for Systems Research, University of Maryland. He has published over 30 papers and over 70 patents and patent applications in the areas of wireless communications.

Richard J. La (S'98–M'01) received the B.S.E.E. degree from the University of Maryland, College Park, MD, USA, in 1994, and the M.S. and Ph.D. degrees in electrical engineering from the University of California, Berkeley, CA, USA, in 1997 and 2000, respectively.

From 2000 to 2001, he was with the Mathematics of Communication Networks Group, Motorola, Inc., Arlington Heights, IL, USA. Since 2001, he has been on the faculty of the Department of Electrical and Computer Engineering, University of Maryland, where he is currently an Associate Professor.

Xin Liu (M'09) received the Ph.D. degree in electrical engineering from Purdue University, West Lafayette, IN, USA, in 2002.

She is currently an Associate Professor with the Computer Science Department, University of California, Davis, CA, USA. Before joining UC Davis, she was a Postdoctoral Research Associate with the Coordinated Science Laboratory, University of Illinois at Urbana–Champaign, Urbana, IL, USA. Her research is on wireless communication networks with a focus on resource allocation and dynamic spectrum management.

Dr. Liu received the Best Paper of Year Award from *Computer Networks* in 2003 for her work on opportunistic scheduling. She received the US National Science Foundation CAREER Award in 2005 for her research on smart-radio-technology-enabled opportunistic spectrum utilization. She received the Outstanding Engineering Junior Faculty Award from the College of Engineering, UC Davis, in 2005.

Youguang Zhang (M'12) received the B.S. degree in mathematics from Zhejiang Normal University, Jinhua, China, in 1984, the M.S. degree in mathematics from Peking University, Beijing, China, in 1987, and the Ph.D. degree in electronics from Biehang University, Beijing, China, in 1990.

Since September 1990, he has been with the School of Electronic and Information Engineering, Beihang University, where he currently holds an academic post of Professor. His research interests include broadband wireless communications, cognitive radio, and signal processing.