Versatile Access Control for Massive IoT: Throughput, Latency, and Energy Efficiency

Han Seung Jang, *Member, IEEE*, Hu Jin, *Senior Member, IEEE*, Bang Chul Jung, *Senior Member, IEEE*, and Tony Q. S. Quek, *Fellow, IEEE*

Abstract—In this paper, we propose a novel access control (AC) mechanism for cellular internet of things (IoT) networks with massive devices, which effectively satisfies *various* performance metrics such as access throughput, access delay, and energy efficiency. Basic idea of the proposed AC mechanism is to adjust access class barring (ACB) factor according to performance targets. For a given performance target, we derive the optimal ACB factors by considering not only the conventional preamble collision detection technique but also the early preamble collision detection technique, respectively. In addition, the proposed AC mechanism considers overall radio resources to optimize the ACB factor, which includes the number of preambles, random access response (RAR) messages, and physical-layer uplink shared channels (PUSCHs), while most conventional ACB schemes consider only the number of preambles. In particular, the proposed AC mechanism is illustrated with two representative performance metrics: latency and energy efficiency. Through extensive computer simulations, it is shown that the proposed versatile AC mechanism outperforms the conventional ACB schemes in terms of various performance metrics under diverse resource constraints.

Index Terms—Internet of things, low-latency, energy-efficiency, access class barring (ACB), ACB factor, backlog estimation

1 Introduction

1.1 Background

Typical communication networks, which are optimized for human-to-human (H2H) communications, are evolving to support machine-to-machine (M2M) communications or internet of things (IoT) services [1]. In particular, 5G cellular wireless networks have been developed and standardized under careful consideration on IoT services in the third generation partnership project (3GPP) and international telecommunication union (ITU) [2], [3]. For supporting massive connections, the IoT networks may require various technologies including edge computing [4], artificial intelligence (AI) [5], etc. Optimizing M2M communication protocols is another interesting topic to satisfy various quality of service (QoS) requirements in the IoT networks [6], [7]. For example, in smart city, autonomous vehicles require very low latency communications with nearby vehicles, while temperature sensors require an ultra-high energy efficiency (EE) to prolong their battery life since the battery lifetime is required to be around 10 years. In particular, Sivanathan et al. [8] proposed a robust framework for IoT device classification using network traffic characteristics based on machine learning schemes.

Existing commercial cellular networks may encounter critical bottlenecks if they are adopted for massive IoT networks without significant modifications. First of all, the amount of radio resources in the cellular network is very limited to accommodate a rapidly growing number of IoT devices, and thus careful resource management is required [9], [10], [11]. Andrade et al. [9], [10] addressed that the control channels including physical downlink channel (PDCCH) may be limited to support overload control, and they proposed a prioritized control resource scheduling algorithm considering both of H2H and M2M communications. Xia et al. [11] provided a comprehensive technical survey on the resource management techniques for M2M communications in commercial LTE/LTE-A cellular networks.

1.2 Related Works

Technical challenges may come from massive nodes that are activated simultaneously due to bursty traffic. In general, the bursty traffic pattern causes severe congestion or overload problem in a radio access network (RAN). To overcome the congestion problem, several overload control schemes have been proposed [12], [13], [14], [15], [16], [17], [18], [19]. Among them, the access class barring (ACB) scheme has been considered as a promising technology since it effectively controls bursty traffic in cellular M2M networks [14]. In particular, the dynamic access class

- *H. S. Jang is with the School of Electrical, Electronic Communication, and Computer Engineering, Chonnam National University, Yeousu 59626, Republic of Korea. E-mail: hsjang@jnu.ac.kr.*
- *H. Jin is with the Division of Electrical Engineering, Hanyang University, Ansan 15588, Republic of Korea. E-mail: hjin@hanyang.ac.kr.*
- *B. C. Jung is with the Department of Electronics Engineering, Chungnam National University, Daejeon 34134, Republic of Korea. E-mail: bcjung@cnu.ac.kr.*
- *T. Q. S. Quek is with the Singapore University of Technology and Design, Singapore 487372, and also with the Department of Electronic Engineering, Kyung Hee University, Yongin 17104, South Korea. E-mail: tonyquek@sutd.edu.sg.*

Manuscript received 1 Sept. 2018; revised 23 Mar. 2019; accepted 26 Apr. 2019. Date of publication 2 May 2019; date of current version 1 July 2020. (Corresponding authors: Hu Jin and Bang Chul Jung.)
Digital Object Identifier no. 10.1109/TMC.2019.2914381

barring (D-ACB) schemes have been actively studied in order to support a varying number of access nodes in IoT networks [17]. The D-ACB schemes estimate the number of backlogged nodes to attempt access at each time slot, and then it generates an optimal ACB factor, also known as access probability, to allow nodes to send a preamble via physical random access channel (PRACH), by considering the available resources. Jin et al. [18] proposed a recursive pseudo-Bayesian ACB scheme in order to lower total service time of a bursty traffic, and a practical backlogged estimation scheme based on the number of undetected preambles at the first step of the random access procedure. On the other hand, Tavana et al. [15] proposed a backlog estimation method based on Kalman filtering for the adaptive ACB scheme, where it is shown that total service time becomes close to that with the optimal ACB scheme that exactly knows the number of backlogged nodes. Duan et al. proposed a dynamic preamble allocation scheme and an iterative algorithm to adaptively update the ACB factor for M2M devices. Vural et al. [19] proposed an adaptive preamble slicing algorithm, called multi-preamble random access, to dynamically determine the preamble set size according to different load conditions and service classes. However, most of existing studies focused on the dynamic access control mechanisms to meet a single performance target such as maximization of the number of successful access or equivalently minimization of access delay and total service time. In addition, they only considered the preamble resources at the first step of RA procedure even though other radio resources are involved in the overall RA procedure.

EE is one of the most important performance metrics especially for battery-powered IoT devices, and thus many techniques have been proposed in literature [20], [21], [22], [23], [24], [25], [26], [27], [28], [29]. Ho et al. [20] proposed a joint massive access control and resource allocation scheme to improve the EE, which adopts machine node grouping, coordinator selection, and optimal power allocation for coordinators. Zhang et al. [26] proposed an integrated energy efficient system for 5G IoT networks, which holistically combines the wireless and wired parts together in order to optimize the EE of the system. Azari and Miao [28] proposed a resource allocation framework based on the max-min lifetime-fairness, which exploits the information on the battery time of devices. Yang et al. [27] proposed an energy efficient resource allocation scheme with energy harvesting technology in time division multiple access and non-orthogonal multiple access strategies. However, the effect of the access control technique on the EE have been rarely investigated in massive IoT networks to the best our knowledge.

1.3 Contributions

In this paper, we propose a *versatile* access control mechanism that effectively satisfies various performance requirements including access throughput, access delay, and the EE for massive IoT networks. To obtain the ACB factor for satisfying latency or EE requirement, the proposed mechanism considers all the radio resources involved in the overall RA procedure: preambles on PRACH, RAR messages on physical downlink shared channel (PDSCH) or PDCCH, and uplink data resources on physical uplink shared channel (PUSCH). In addition to the conventional preamble collision detection (C-PCD) technique, we also exploit the early preamble collision detection (E-PCD) technique at the first RA procedure [30], [31]. Furthermore, we propose a comprehensive access control mechanism which can simultaneously support two groups of devices requesting low-latency and ultra-low power consumption services, respectively. Simulation results show that the proposed versatile AC mechanism efficiently satisfies various performance requirements under various resource constraints and it outperforms the existing AC schemes. The main contributions of this paper can be summarized as follows:

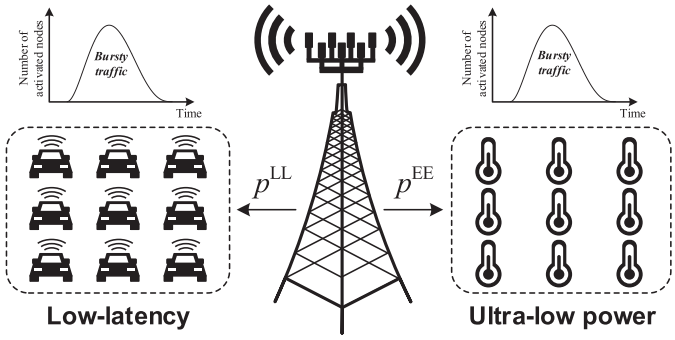


Fig. 1. System model of a massive IoT network with versatile access control mechanism.

- We propose a versatile AC mechanism to efficiently control random access traffic from a massive number of IoT nodes while achieving various performance requirements such as low-latency and energy-efficiency.
- We derive the ACB factors for low-latency access and energy-efficient access with the conventional preamble collision detection technique and the early preamble collision detection technique, respectively.
- We also propose a comprehensive AC mechanism in order to simultaneously support both of low-latency access group and energy-efficient access group.

The rest of this paper is organized as follows. In Section 2, we describe a system model considered in this paper. In Section 3, we derive preliminary probabilities of collision-free and collided preambles which are utilized to optimize the ACB factor of the proposed mechanism. In Section 4, we propose two types of AC mechanisms: low-latency AC and energy-efficient AC mechanisms. Then, we explain the estimation algorithm for the number of backlogged nodes in Section 5. The proposed AC mechanisms are validated via extensive computer simulations in Section 6. Finally, conclusions are drawn in Section 7.

2 System Model

Fig. 1 shows the system model of a single cell massive IoT network which is considered in this paper. We assume that the eNodeB controls uplink access from two representative groups[1], each of which requires a certain performance

1. In practice, there may be a number of different groups. For example, a group of devices may require a mixed performance requirement among maximum throughput, low-latency, and low-power consumption.

metric such as low-latency (LL) and energy efficiency (EE). Activation of nodes is assumed to be bursty due to an external event for both groups. In addition, the eNodeB is assumed to adopt a probability-based dynamic access class barring (ACB) technique as used in the commercial cellular network, i.e., 3GPP LTE system [32], which inherently controls a number of simultaneous accesses by using a single number called ACB factor $p \in (0,1)$. Let p^{LL} and p^{EE} denote the ACB factors for LL access and EE access, respectively. In the proposed versatile AC mechanism, we consider following parameters:

- M: number of available preambles
- Q: number of available RAR messages
- U: number of available PUSCH resources
- N: total number of activated nodes in a group
- v_i: estimated number of backlogged nodes for the ith slot
- p_i: ACB factor for the ith slot

The above parameters will be explained in the next subsection in detail. Note that Q depends on PDSCH or PDCCH resources, and U depends on PUSCH resources. Since a single RAR message can deliver a single uplink resource grant for PUSCH [33], the number of allocable PUSCH resources is limited to

$$K = \min\{Q, U\}. \tag{1}$$

Thus, we only consider the number of available preambles M and the number of allocable PUSCH resources K throughout the paper.

2.1 Overall Random Access Procedure

The overall RA procedure consists of five steps, which include the access check (*Step 0*) and the typical four-step contention-based RA procedure (*Step 1-4*) [34].

(*Step 0*) Access check: The eNodeB calculates an ACB factor for a specific purpose based on the estimated number of backlogged nodes v_i and the number of available resources M and K, and then it notifies the ACB factor of the ith PRACH $p_i \in [0,1]$ at the beginning of the ith PRACH slot. Then, each activated node generates a random number $q \in [0,1]$ and compares q with p_i. If $q \leq p_i$, then the node attempts an RA on the ith PRACH slot. Otherwise, it defers an RA attempt to the $(i+1)$th PRACH slot.

(*Step 1) Preamble transmission and detection*: Each node which passed the access check at the Step 0 randomly selects a single preamble among M available preambles, and then sends the selected preamble on the ith PRACH slot. More than one node may select the same preamble due to the random preamble selection property.

(*Step 2) Random access response (RAR)*: After detecting the preambles from nodes, the eNodeB sends the RAR messages, each of which conveys identity of detected preambles, timing advance information, and an initial uplink PUSCH resource grant for uplink data transmission.

(*Step 3) Uplink data transmission*: Using the PUSCH resource indicated via the RAR message at the Step 2, the corresponding node transmits uplink data which conveys a data packet, a radio resource control (RRC) connection request, a tracking area update, or a scheduling request.

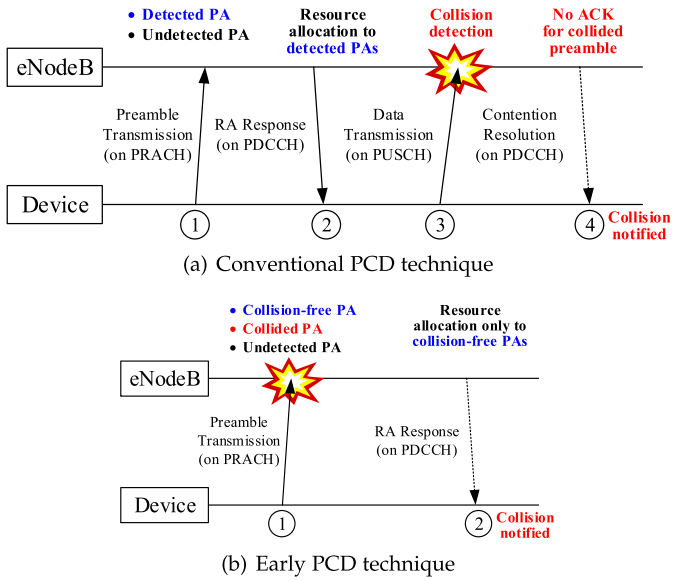


Fig. 2. Comparison of preamble collision detection (PCD) techniques: (a) conventional PCD technique and (b) early PCD technique.

(*Step 4) ACK message transmission*: If the eNodeB successfully decodes the data transmitted by a single node, it sends back the ACK message including the node identity (ID) which is obtained from the decoded data, otherwise the eNodeB sends nothing back.

2.2 Preamble Collision Detection Techniques

As noted before, we consider two preamble collision detection (PCD) techniques: conventional PCD (C-PCD) and early PCD (E-PCD) [35]. Fig. 2 compares the two techniques. The C-PCD technique is performed based on unsuccessful data decoding at Step 3 of the RA procedure. In other works, if multiple nodes send the same preamble at Step 1 and receive the identical RAR message from the eNodeB at Step 2, then all relevant nodes transmit their own data via the same uplink resource and the eNodeB may not decode the data successfully. From this unsuccessful decoding, the eNodeB detects the preamble collision. If the preamble collisions are detected, the eNodeB does not send the ACK message to the corresponding devices at Step 4. Nodes which do not receive the ACK message try to perform the RA procedure at the next PRACH slot.

On the other hand, the PCD can be performed at Step 1 of the RA procedure with the E-PCD techniques which exploit tagged preambles [30], [31], ID transmission [36], or machine learning algorithms [37]. If the preamble collisions are detected with E-PCD techniques at Step 1, the eNodeB does not send the RAR message to the corresponding nodes at Step 2. Then, nodes which do not receive the RAR message defer to attempt the RA to the next PRACH slot without transmitting uplink data. It is worth noting that the E-PCD techniques can separate the detected preambles into collision-free preambles and collied preambles and allocates the PUSCH resources only for the collision-free preambles through the RAR messages at Step 2.

2.3 Traffic Model

We assume that each of N nodes in a group is activated at the time $x \in (0, T_{\mathsf{act}})$, which follows Beta distribution with

parameters $\alpha = 3$ and $\beta = 4$ [38]:

$$f_X(x) = \frac{x^{\alpha-1}(T_{\text{act}} - x)^{\beta-1}}{(T_{\text{act}})^{\alpha+\beta-1} B(\alpha, \beta)},$$

where $B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1} dx$ denotes the beta function. It is assumed that there exist I_{act} slots within time T_{act} and the expected number of newly activated nodes in the ith RA slot is given by $\lambda_i = N \int_{t_{i-1}}^{t_i} f_X(x) dx$ for $i = 1, 2, \ldots, I_{\text{act}}$.

3 Preliminary Probabilities on Collision-Free Preamble and Collided Preamble

In this section, we first define the following discrete random variables (RVs):

- D: number of detected preambles,
- S: number of collision-free preambles,
- C: number of collided preambles,
- R: number of undetected preambles,

where $D = S + C$ and $R = M - D$. Then, we derive several preliminary probabilities related to the above RVs, which are used for designing the versatile AC mechanism.

At Step 1 of the RA procedure, each activated node which passes the access check sends an arbitrary preamble among M available preambles via PRACH, and then the eNodeB detects preambles. In general, detected preambles are classified into collision-free preambles and collided preambles. The collision-free preamble indicates the detected preamble which is sent by a single node, while the collided preamble indicates the detected preamble which is sent by more than two nodes. However, depending on the PCD techniques (e.g., E-PCD or C-PCD), the eNodeB can or cannot identify the detected preamble is collision-free or collided at Step 1 of the RA procedure.

We first consider the following joint probability that when m nodes simultaneously attempt the RAs, any $(r+s)$ preambles among M preambles are chosen, and then each of s preambles among chosen $(r+s)$ preambles is selected (collision-free) only by a sinlge node, and each of remaining r preambles is not selected by any nodes:

$$\begin{aligned}&\Psi_{r+s}^{M}(s|m)\\ &= \binom{M}{r+s}\binom{r+s}{s}\binom{m}{s} s! \frac{\{M-(r+s)\}^{m-s}}{M^m},\end{aligned} \tag{2}$$

where $\binom{M}{r+s}\binom{r+s}{s}$ represents the total number of cases that chooses $(r+s)$ preambles among M, and then chooses s collision-free preambles among the chosen $(r+s)$ preambles [15]. In addition, the term $\binom{m}{s} s!(d-s)^{m-s}$ represents the total number of cases that s nodes among m RA-attempting nodes transmits an exclusive preamble and the remaining $(m-s)$ nodes chooses one of preambles among $\{M-(r+s)\}$ preambles. Then, the denominator M^m represents the total number of cases that m RA-attempting nodes select a preamble among M preambles.

Let E_i denote the event that the ith preamble is selected by at most one node, and then the probability of the event E_i when m nodes attempt RAs simultaneously, i.e., m nodes pass the access check at Step 0, is given by the sum of the probability that ith preamble is not selected by any nodes and the probability that ith preamble is selected only by a single node, i.e.,

$$\Pr\{E_i|m\} = \left(1 - \frac{1}{M}\right)^m + \binom{m}{1}\frac{1}{M}\left(1-\frac{1}{M}\right)^{m-1}, \tag{3}$$

where $\frac{1}{M}$ denotes the probability that a node selects the ith preamble among M possibilities.

Based on the inclusion-exclusion principle [39], $\Pr\{E_1 \cup \cdots \cup E_M|m\}$ is written as

$$\begin{aligned}&\Pr\{\cup_{i=1}^{M} E_i|m\} = \sum_{i=1}^{M}(-1)^{i+1}\sum_{j=0}^{i}\Psi_i^M(j|m)\\ &= \sum_{i=1}^{M}\sum_{j=0}^{i}(-1)^{i+1}\binom{M}{i}\binom{i}{j}\binom{m}{j} j! \frac{(M-i)^{m-j}}{M^m},\end{aligned} \tag{4}$$

which denotes the probability that at least one preamble among M preambles is selected by at most one node with m RA-attempting nodes. Besides, the complementary probability to $\Pr\{\cup_{i=1}^{M} E_i|m\}$ is given by

$$\begin{aligned}&\Phi(M|m) = 1 - \Pr\{\cup_{i=1}^{M} E_i|m\}\\ &= 1 - \sum_{i=1}^{M}\sum_{j=0}^{i}(-1)^{i+1}\binom{M}{i}\binom{i}{j}\binom{m}{j} j! \frac{(M-i)^{m-j}}{M^m}\\ &= \sum_{i=0}^{M}\sum_{j=0}^{i}(-1)^{i}\binom{M}{i}\binom{i}{j}\binom{m}{j} j! \frac{(M-i)^{m-j}}{M^m},\end{aligned} \tag{5}$$

which represents the probability that each of all M preambles are selected by at least two nodes (collided) with m RA-attempting nodes. As a result, we obtain the joint probability that s preambles are collision-free among d detected preambles when m nodes simultaneously attempt the RA:

$$\begin{aligned}&\Pr\{D = d, S = s|m\} = \Psi_{r+s}^{M}(s|m)\Phi(M-r-s|m-s)\\ &= \Psi_{M-d+s}^{M}(s|m)\Phi(d-s|m-s)\\ &= \sum_{i=0}^{d-s}\sum_{j=0}^{i}\binom{M}{d}\binom{d}{s}\binom{m}{s} s! \frac{(d-s)^{m-s}}{M^m} \times\\ &\quad (-1)^i\binom{d-s}{i}\binom{i}{j}\binom{m-s}{j} j! \frac{(d-s-i)^{m-s-j}}{(d-s)^{m-s}},\end{aligned} \tag{6}$$

where we utilize $r = M - d$ and $\binom{M}{M-d+s}\binom{M-d+s}{s} = \binom{M}{d}\binom{d}{s}$.

We assume that a *Priori* distribution of the number of backlogged nodes n follows Poisson distribution with mean of ν, i.e.,

$$\mathbb{P}(n|\nu) = \frac{\nu^n}{n!} e^{-\nu}. \tag{7}$$

It is shown that this assumption significantly reduces the computation complexity of designing algorithms [40]. Then, for given ACB factor p and ν, the joint probability that s preambles are collision-free among d detected preambles when m out of n backlogged nodes simultaneously attempt the RA is given by

$$\Pr\{d,s,n,m|p,\nu\} = \underbrace{\Pr\{d,s|m\}}_{6}\underbrace{\Pr\{m|p,n\}}_{\mathbb{B}^n_m(p)}\underbrace{\Pr\{n|\nu\}}_{\mathbb{P}(n|\nu)}, \tag{8}$$

where $\mathbb{B}^n_m(p) = \binom{n}{m}p^m(1-p)^{(n-m)}$ denotes binomial distribution with ACB factor of p. By marginalizing (8) with regard to n and m, the joint probability that s preambles are collision-free among d detected preambles for given p and ν is given by

$$\begin{aligned}\Pr\{D=d,S=s|p,\nu\} &= \sum_{n=0}^{\infty}\sum_{m=0}^{n}\Pr\{d,s,n,m|p,\nu\}\\ &= \binom{M}{d}\binom{d}{s}e^{-p\nu}\Big(\frac{p\nu}{M}\Big)^s\Big(-1-\frac{p\nu}{M}+e^{\frac{p\nu}{M}}\Big)^{d-s}.\end{aligned} \tag{9}$$

Then, by marginalizing $\Pr\{D=d,S=s|p,\nu\}$ with regard to S and D, respectively, $\Pr\{D=d|p,\nu\}$ and $\Pr\{S=s|p,\nu\}$ are also given by

$$\begin{aligned}\Pr\{D=d|p,\nu\} &= \sum_{s=0}^{d}\Pr\{D=d,S=s|p,\nu\}\\ &= \binom{M}{d}e^{-p\nu}\sum_{s=0}^{d}\binom{d}{s}\Big(\frac{p\nu}{M}\Big)^s\Big(-1-\frac{p\nu}{M}+e^{\frac{p\nu}{M}}\Big)^{d-s}\\ &= \binom{M}{d}e^{-p\nu}\Big(-1+e^{\frac{p\nu}{M}}\Big)^d,\end{aligned} \tag{10}$$

and

$$\begin{aligned}\Pr\{S=s|p,\nu\} &= \sum_{d=0}^{M}\Pr\{D=d,S=s|p,\nu\}\\ &= e^{-p\nu}\Big(\frac{p\nu}{M}\Big)^s\sum_{d=0}^{M}\binom{M}{d}\binom{d}{s}\Big(-1-\frac{p\nu}{M}+e^{\frac{p\nu}{M}}\Big)^{d-s}\\ &= \binom{M}{s}e^{-p\nu}\Big(\frac{p\nu}{M}\Big)^s\sum_{d=0}^{M}\binom{M-s}{d-s}\Big(-1-\frac{p\nu}{M}+e^{\frac{p\nu}{M}}\Big)^{d-s}\\ &= \binom{M}{s}e^{-p\nu}\Big(\frac{p\nu}{M}\Big)^s\Big(-\frac{p\nu}{M}+e^{\frac{p\nu}{M}}\Big)^{M-s},\end{aligned} \tag{11}$$

respectively. Since $C = D - S$, we have $\Pr\{C=c,S=s|p,\nu\}$ by substituting $s+c$ for d in Eq. (9). Thus, the probability that the number of collided preambles is equal to c is given by

$$\begin{aligned}\Pr\{C=c|p,\nu\} &= \sum_{s=0}^{M-c}\Pr\{C=c,S=s|p,\nu\}\\ &= \binom{M}{c}e^{-p\nu}\Big(-1-\frac{p\nu}{M}+e^{\frac{p\nu}{M}}\Big)^c\Big(1+\frac{p\nu}{M}\Big)^{M-c}.\end{aligned} \tag{12}$$

As a result, for given p and ν, the expected values of RVs D, S and C are given by

$$\begin{aligned}\mathbb{E}[D|p,\nu] &= \sum_{d=0}^{M}d\binom{M}{d}e^{-p\nu}\Big(-1+e^{\frac{p\nu}{M}}\Big)^d\\ &= Me^{-p\nu}\Big(-1+e^{\frac{p\nu}{M}}\Big)\sum_{d=0}^{M}\binom{M-1}{d-1}\Big(-1+e^{\frac{p\nu}{M}}\Big)^{d-1}\\ &= Me^{-p\nu}\Big(-1+e^{\frac{p\nu}{M}}\Big)e^{\frac{p\nu}{M}(M-1)}\\ &= M\Big(1-e^{-\frac{p\nu}{M}}\Big),\end{aligned} \tag{13}$$

$$\begin{aligned}\mathbb{E}[S|p,\nu] &= \sum_{s=0}^{M}s\binom{M}{s}\Big(\frac{p\nu}{M}\Big)^s e^{-p\nu}\Big(-\frac{p\nu}{M}+e^{\frac{p\nu}{M}}\Big)^{M-s}\\ &= p\nu e^{-p\nu}\sum_{s=0}^{M}\binom{M-1}{s-1}\Big(\frac{p\nu}{M}\Big)^{s-1}\Big(-\frac{p\nu}{M}+e^{\frac{p\nu}{M}}\Big)^{M-s}\\ &= p\nu e^{-p\nu}e^{\frac{p\nu}{M}(M-1)}\\ &= p\nu e^{-\frac{p\nu}{M}},\end{aligned} \tag{14}$$

and

$$\begin{aligned}\mathbb{E}[C|p,\nu] &= \mathbb{E}[D|p,\nu] - \mathbb{E}[S|p,\nu]\\ &= M\Big(1-e^{-\frac{p\nu}{M}}\Big) - p\nu e^{-\frac{p\nu}{M}}\\ &= M\Big(-1-\frac{p\nu}{M}+e^{\frac{p\nu}{M}}\Big)e^{-\frac{p\nu}{M}},\end{aligned} \tag{15}$$

respectively.

4 Versatile Access Control Mechanisms

In this section, we propose two representative AC mechanisms: low-latency AC mechanism and energy-efficient AC mechanism. For each mechanism, we derive the ACB factors by considering both the C-PCD technique and the E-PCD technique. Here, we assume that the estimated number of backlogged nodes is equal to ν. The estimation algorithm for the number of backlogged nodes will be explained in the next section.

4.1 Low-Latency Access Control

The low-latency requirement can be satisfied by maximizing the access throughput in each slot. Thus, we derive the ACB factor for maximizing the access throughput which is defined as the number of nodes that succeed in the RA.

4.1.1 With C-PCD Technique

Basically, the RA of a certain node is successful when it sends an exclusive preamble and receives an exclusive PUSCH resource grant. Let A denote the access throughput, i.e., the number of nodes that succeed in the RA. Recall that K denotes the number of allocable PUSCH resources at Step 2 of the RA procedure and D denotes the number of detected preambles. By using the well-known total probability theorem, the probability that $A = a$ can be obtained as

$$\begin{aligned}\Pr\{A=a\} &= \Pr\{A=a,D\le K\} + \Pr\{A=a,D>K\}\\ &= \Pr\{S=a,D\le K\} + \Pr\{S\ge a,D>K\}\mathbb{B}^s_a\big(\tfrac{K}{D}\big),\end{aligned} \tag{16}$$

where $\mathbb{B}^s_a\big(\frac{K}{D}\big) = \binom{s}{a}\big(\frac{K}{D}\big)^a\big(1-\frac{K}{D}\big)^{s-a}$ and $\frac{K}{D} < 1$. The term $\frac{K}{D} < 1$ can be regarded as the resource scheduling probability. Then, the conditional expected value of A for given p and ν is given by

$$\begin{aligned}\mathbb{E}[A|p,\nu] = &\sum_{a=0}^{K}\sum_{d=0}^{K}a\Pr\{D=d,S=a|p,\nu\}+\\ &\sum_{a=0}^{K}\sum_{d=K+1}^{M}\sum_{s=a}^{d}a\Pr\{D=d,S=s|p,\nu\}\mathbb{B}^s_a\big(\tfrac{K}{D}\big).\end{aligned} \tag{17}$$

Then, we obtain the optimal ACB factor p to maximize the average access throughput as follows:

$$p_{\mathsf{conv}}^{\mathsf{LL}} = \underset{0\le p\le 1}{\arg\max}\, \mathbb{E}[A|p,\nu]. \tag{18}$$

In particular, when $K = M$, Eq. (17) can be rewritten by

$$\begin{aligned}\mathbb{E}[A|p,\nu] &= \sum_{a=0}^{M}\sum_{d=0}^{M} a\Pr\{D=d, S=a|p,\nu\} \\ &= \mathbb{E}[S|p,\nu].\end{aligned} \tag{19}$$

Thus, in this case, the optimal $p_{\mathsf{conv}}^{\mathsf{LL}}$ is found by taking derivative of $\mathbb{E}[S|p,\nu] = p\nu e^{-\frac{p\nu}{M}}$ with respect to p, i.e.,

$$p_{\mathsf{conv}}^{\mathsf{LL}} = \frac{M}{\nu}. \tag{20}$$

When $K < M$, on the other hand, it is difficult to find the closed-form expression for the optimal ACB factor and thus we approximate $\mathbb{E}[A|p,\nu]$ as

$$\mathbb{E}[A|p,\nu] \approx \mathbb{E}[S|p,\nu]\cdot\min\left\{1, \frac{K}{\mathbb{E}[D|p,\nu]}\right\}, \tag{21}$$

where we have two components: $\mathbb{E}[S|p,\nu]$ represents the average number of nodes that succeed in preamble transmissions, and $\min\left\{1, \frac{K}{\mathbb{E}[D|p,\nu]}\right\}$ represents the successful resource scheduling probability. In case that $\mathbb{E}[D|p_{\mathsf{conv}}^{\mathsf{LL}} = \frac{M}{\nu}, \nu] \le K$, the optimal ACB factor can be approximated as Eq. (20) since $\min\{1, \frac{K}{\mathbb{E}[D|p_{\mathsf{conv}}^{\mathsf{LL}}, \nu]}\} = 1$ and $\mathbb{E}[A|p,\nu] \approx \mathbb{E}[S|p,\nu]$. On the other hand, when $\mathbb{E}[D|p_{\mathsf{conv}}^{\mathsf{LL}} = \frac{M}{\nu}, \nu] > K$, $\mathbb{E}[A|p,\nu]$ increases until the ACB factor p approaches to $\tilde{p}_{\mathsf{conv}}^{\mathsf{LL}}$ such that $\mathbb{E}[D|\tilde{p}_{\mathsf{conv}}^{\mathsf{LL}}, \nu] = K$ and $\mathbb{E}[A|p,\nu]$ becomes decreased if $p > \tilde{p}_{\mathsf{conv}}^{\mathsf{LL}}$ since $\mathbb{E}[S|p,\nu] < \mathbb{E}[D|p,\nu]$ and $\mathbb{E}[D|p,\nu] = M\left(1 - e^{-\frac{p\nu}{M}}\right)$ is an increasing function of p. It implies that the eNodeB should control the ACB factor p so that it allocates PUSCH resources to all detected preambles in order to maximize access throughput. In other words, the resource scheduling probability needs to approach to 1 for maximizing the access throughput.

As a result, we obtain the following ACB factor by solving $\mathbb{E}[D|p,\nu] = M\left(1 - e^{-\frac{p\nu}{M}}\right) = K$:

$$\tilde{p}_{\mathsf{conv}}^{\mathsf{LL}} = -\ln\left(1 - \frac{K}{M}\right)\frac{M}{\nu}, \quad K < M. \tag{22}$$

4.1.2 With E-PCD Technique

With E-PCD techniques, the eNodeB allocates resources only for the collision-free preambles, and thus the average access throughput is given by

$$\mathbb{E}[A|p,\nu] = \sum_{a=0}^{M}\min(a,K)\Pr\{S=a|p,\nu\}, \tag{23}$$

where S denotes the number of collision-free preambles. Then, we obtain the optimal ACB factor p to maximize the average access throughput as follows:

$$p_{\mathsf{early}}^{\mathsf{LL}} = \underset{0\le p\le 1}{\arg\max}\ \mathbb{E}[A|p,\nu]. \tag{24}$$

Similar to the case of C-PCD technique, when $K = M$,

$$\mathbb{E}[A|p,\nu] = \sum_{a=0}^{M} a\Pr\{S=a|p,\nu\} = \mathbb{E}[S|p,\nu], \tag{25}$$

and thus the optimal $p_{\mathsf{early}}^{\mathsf{LL}}$ is also found by

$$p_{\mathsf{early}}^{\mathsf{LL}} = \frac{M}{\nu}. \tag{26}$$

Note that, when $K = M$, the optimal ACB factors with the C-PCD and the E-PCD are identical, i.e., $p_{\mathsf{conv}}^{\mathsf{LL}} = p_{\mathsf{early}}^{\mathsf{LL}} = \frac{M}{\nu}$.

When $K < M$, on the other hand, it is difficult to find the closed-form expression for the optimal ACB factor, and thus, we approximate $\mathbb{E}[A|p,\nu]$ as

$$\mathbb{E}[A|p,\nu] \approx \mathbb{E}[S|p,\nu]\cdot\min\left\{1, \frac{K}{\mathbb{E}[S|p,\nu]}\right\}, \tag{27}$$

where we have two components: $\mathbb{E}[S|p,\nu]$ represents the average number of nodes that succeed in preamble transmissions, and $\min\left\{1, \frac{K}{\mathbb{E}[S|p,\nu]}\right\}$ represents the successful resource scheduling probability. Note that Eq. (27) is different from Eq. (21), which comes from the fact that the E-PCD technique allocates resources only for the collision-free preambles. In case that $\mathbb{E}[S|p_{\mathsf{early}}^{\mathsf{LL}} = \frac{M}{\nu}, \nu] \le K$, the optimal ACB factor can be approximated as Eq. (26) since $\min\{1, \frac{K}{\mathbb{E}[S|p_{\mathsf{early}}^{\mathsf{LL}}, \nu]}\} = 1$ and $\mathbb{E}[A|p,\nu] \approx \mathbb{E}[S|p,\nu]$. However, when $\mathbb{E}[S|p_{\mathsf{early}}^{\mathsf{LL}} = \frac{M}{\nu}, \nu] > K$, $\mathbb{E}[A|p,\nu]$ increases until the ACB factor p approaches to $\tilde{p}_{\mathsf{early}}^{\mathsf{LL}}$ such that $\mathbb{E}[S|\tilde{p}_{\mathsf{early}}^{\mathsf{LL}}, \nu] = K$ and $\mathbb{E}[A|p,\nu] \approx K$ if $\tilde{p}_{\mathsf{early}}^{\mathsf{LL}} < p \le p_{\mathsf{early}}^{\mathsf{LL}}$. It implies that the eNodeB cannot allocate resources even for some collision-free preambles when $p > \tilde{p}_{\mathsf{early}}^{\mathsf{LL}}$, which is called *resource allocation failure* in this paper. The average number of collision-free preambles that experience the resource allocation failure is given by

$$F(p,\nu) = \begin{cases} 0, & \text{if } p \le \tilde{p}_{\mathsf{early}}^{\mathsf{LL}}, \\ \mathbb{E}[S|p,\nu] - K, & \text{if } \tilde{p}_{\mathsf{early}}^{\mathsf{LL}} < p < p_{\mathsf{early}}^{\mathsf{LL}}. \end{cases}$$

We obtain the ACB factor p such that $\mathbb{E}[S|p,\nu] = K$ to maximize the access throughput and minimize the number of resource allocation failures (the number of unnecessary accesses). As a result, we obtain the following ACB factor by solving $\mathbb{E}[S|p,\nu] = p\nu e^{-\frac{p\nu}{M}} = K$:

$$\tilde{p}_{\mathsf{early}}^{\mathsf{LL}} = -\mathcal{W}_0\left(-\frac{K}{M}\right)\frac{M}{\nu}, \quad K < M, \tag{28}$$

where $\mathcal{W}_0(x)$ represents the principle (upper) branch Lambert W function with $\mathcal{W}_0(x) > -1$ [41].

However, if $\tilde{p}^{\mathsf{LL}}_{\mathsf{early}}$ is used as the ACB factor, the average access throughput becomes slightly lower than K because the transmissions occur in a probabilistic manner. To validate this, let p denote the ACB factor and $\mathbb{B}^N_n(p) = \binom{N}{n}(p)^n(1-p)^{(N-n)}$ denotes the probability that n nodes attempt RAs among N backlogged nodes with the ACB factor of p. Then, the average number of successful RA-attempts is given by

$$\mathcal{S}(p,K) = \sum_{n=0}^{N} \min\left[n\left(1-\tfrac{1}{M}\right)^{n-1}, K\right]\mathbb{B}^N_n(p), \tag{29}$$

where $\mathcal{S}(\tilde{p}^{\mathsf{LL}}_{\mathsf{early}}, K)$ is smaller than $\mathcal{S}(p^{\mathsf{LL}}_{\mathsf{early}}, K)$.

On the other hand, let us define *unsuccessful* RA-attempts as the attempts either with collided preambles or with collision-free preambles experiencing the resource allocation failure. The average number of unsuccessful RA-attempts is given by

$$\mathcal{U}(p,K) = \sum_{n=0}^{N} n\mathbb{B}^N_n(p) - \mathcal{S}(p,K). \tag{30}$$

Note that the unsuccessful RA-attempt causes energy waste of nodes. Hence, the ACB factor needs to be carefully controlled between $\tilde{p}^{\mathsf{LL}}_{\mathsf{early}}$ and $p^{\mathsf{LL}}_{\mathsf{early}}$ by considering both $\mathcal{S}(p,K)$ and $\mathcal{U}(p,K)$ in practice. We will further discuss a detailed setting for p in the section of numerical results.

4.2 Energy Efficient Access Control

Energy efficiency (EE) is defined as the ratio of energy consumed for the successful uplink transmissions to energy consumed for all uplink transmissions during the RA procedure at a node.

4.2.1 With C-PCD Technique

The EE with C-PCD technique is given by

$$\begin{aligned}\mathbb{E}[\eta^{\mathsf{EE}}_{\mathsf{conv}}(p)] &= \frac{(1+\chi)\mathbb{E}[S|p,\nu]\min\left(1,\mathbb{E}\left[\frac{K}{D}|p,\nu\right]\right)}{p\nu+\chi p\nu\min\left(1,\mathbb{E}\left[\frac{K}{D}|p,\nu\right]\right)}\\ &= \frac{(1+\chi)\mathbb{E}[S|p,\nu]\min\left(1,\mathbb{E}\left[\frac{K}{D}|p,\nu\right]\right)}{p\nu\left\{1+\chi\min\left(1,\mathbb{E}\left[\frac{K}{D}|p,\nu\right]\right)\right\}},\end{aligned} \tag{31}$$

where χ denotes the ratio of energy consumed for the preamble transmission at Step 1 over energy consumed for the data transmission at Step 3 at a node. If we control the ACB factor such that all nodes with detected preambles receive resource grants, i.e., $\mathbb{E}[D|p,\nu] \le K$, $\min\left(1,\mathbb{E}\left[\frac{K}{D}|p,\nu\right]\right) = 1$, then $\mathbb{E}[\eta^{\mathsf{EE}}_{\mathsf{conv}}(p)]$ is rewritten as

$$\mathbb{E}[\eta^{\mathsf{EE}}_{\mathsf{conv}}(p)] = \frac{(1+\chi)\mathbb{E}[S|p,\nu]}{(1+\chi)p\nu} = \frac{p\nu e^{-\frac{p\nu}{M}}}{p\nu} = e^{-\frac{p\nu}{M}}. \tag{32}$$

The EE with C-PCD technique is a decreasing function of p, and thus there exists a trade-off between the average access throughput and the energy efficiency. It is worth noting that (32) does not depend on the energy ratio of preamble over data transmissions χ. In particular, if $p^{\mathsf{LL}}_{\mathsf{conv}}$ and $\tilde{p}^{\mathsf{LL}}_{\mathsf{conv}}$ are used for the ACB factor in (32), then we have

$$\mathbb{E}[\eta^{\mathsf{EE}}_{\mathsf{conv}}(p^{\mathsf{LL}}_{\mathsf{conv}})] = e^{-1}, \tag{33}$$

and

$$\mathbb{E}[\eta^{\mathsf{EE}}_{\mathsf{conv}}(\tilde{p}^{\mathsf{LL}}_{\mathsf{conv}})] = 1 - \tfrac{K}{M}, \tag{34}$$

respectively.

For achieving a certain EE of θ for a group of nodes, the ACB factor is controlled as:

$$p^{\mathsf{EE}:\theta}_{\mathsf{conv}} = -\ln(\theta)\frac{M}{\nu}, \tag{35}$$

which satisfies

$$\mathbb{E}[\eta^{\mathsf{EE}}_{\mathsf{conv}}(p)] = e^{-\frac{p\nu}{M}} = \theta.$$

For energy-efficient AC, the target value of θ needs to be higher than both $\mathbb{E}[\eta^{\mathsf{EE}}_{\mathsf{conv}}(p^{\mathsf{LL}}_{\mathsf{conv}})]$ and $\mathbb{E}[\eta^{\mathsf{EE}}_{\mathsf{conv}}(\tilde{p}^{\mathsf{LL}}_{\mathsf{conv}})]$, which is set to

$$\theta > \begin{cases} e^{-1}, & \text{if } \mathbb{E}[D|p^{\mathsf{LL}}_{\mathsf{conv}} = \frac{M}{\nu},\nu] \le K,\\ 1-\frac{K}{M}, & \text{if } \mathbb{E}[D|p^{\mathsf{LL}}_{\mathsf{conv}} = \frac{M}{\nu},\nu] > K.\end{cases}$$

4.2.2 With E-PCD Technique

The EE with E-PCD technique is given by

$$\mathbb{E}[\eta^{\mathsf{EE}}_{\mathsf{early}}(p)] = \frac{(1+\chi)\mathbb{E}[S|p,\nu]\min\left(1,\mathbb{E}\left[\frac{K}{S}|p,\nu\right]\right)}{p\nu+\chi\mathbb{E}[S|p,\nu]\min\left(1,\mathbb{E}\left[\frac{K}{S}|p,\nu\right]\right)}. \tag{36}$$

If we control the ACB factor such that all nodes with collision-free preambles receive resource grants, i.e., $\mathbb{E}[S|p,\nu] \le K$, $\min\left(1,\mathbb{E}\left[\frac{K}{S}|p,\nu\right]\right) = 1$, then $\mathbb{E}[\eta^{\mathsf{EE}}_{\mathsf{early}}(p)]$ is rewritten as

$$\begin{aligned}\mathbb{E}[\eta^{\mathsf{EE}}_{\mathsf{early}}(p)] &= \frac{(1+\chi)\mathbb{E}[S|p,\nu]}{p\nu+\chi\mathbb{E}[S|p,\nu]} = \frac{(1+\chi)p\nu e^{-\frac{p\nu}{M}}}{p\nu+\chi p\nu e^{-\frac{p\nu}{M}}}\\ &= \frac{(1+\chi)e^{-\frac{p\nu}{M}}}{1+\chi e^{-\frac{p\nu}{M}}}.\end{aligned} \tag{37}$$

The EE with E-PCD technique is a decreasing function of p, and thus there exists a trade-off between the average access throughput and the energy efficiency as well. In contrast to $\mathbb{E}[\eta^{\mathsf{EE}}_{\mathsf{conv}}(p)]$ in (32), the EE with E-PCD technique $\mathbb{E}[\eta^{\mathsf{EE}}_{\mathsf{early}}(p)]$ depends on the energy ratio of preamble over data transmissions χ. In particular, if $p^{\mathsf{LL}}_{\mathsf{early}}$ and $\tilde{p}^{\mathsf{LL}}_{\mathsf{early}}$ are used for the ACB factor in (37), then we have

$$\mathbb{E}\left[\eta^{\mathsf{EE}}_{\mathsf{early}}(p^{\mathsf{LL}}_{\mathsf{early}})\right] = \frac{(1+\chi)e^{-1}}{1+\chi e^{-1}}, \tag{38}$$

and

$$\mathbb{E}\left[\eta^{\mathsf{EE}}_{\mathsf{early}}(\tilde{p}^{\mathsf{LL}}_{\mathsf{early}})\right] = \frac{(1+\chi)e^{\mathcal{W}_0\left(-\frac{K}{M}\right)}}{1+\chi e^{\mathcal{W}_0\left(-\frac{K}{M}\right)}}, \tag{39}$$

respectively.

For achieving a certain EE of θ for a group of nodes, the ACB factor is controlled as:

$$p^{\mathsf{EE}:\theta}_{\mathsf{early}} = -\ln\left(\frac{\theta}{1+(1-\theta)\chi}\right)\frac{M}{\nu}, \tag{40}$$

which satisfies

$$\mathbb{E}[\eta^{\mathsf{EE}}_{\mathsf{early}}(p)] = \frac{(1+\chi)e^{-\frac{p\nu}{M}}}{1+\chi e^{-\frac{p\nu}{M}}} = \theta.$$

For energy-efficient AC, the target value of θ needs to be higher than both $\mathbb{E}[\eta^{\mathsf{EE}}_{\mathsf{early}}(p^{\mathsf{LL}}_{\mathsf{early}})]$ and $\mathbb{E}[\eta^{\mathsf{EE}}_{\mathsf{early}}(\tilde{p}^{\mathsf{LL}}_{\mathsf{early}})]$, which is set to

$$\theta > \begin{cases} \frac{(1+\chi)e^{-1}}{1+\chi e^{-1}}, & \text{if } \mathbb{E}[S|p^{\mathsf{LL}}_{\mathsf{early}} = \frac{M}{\nu}, \nu] \le K, \\ \frac{(1+\chi)e^{\mathcal{W}_0\left(-\frac{K}{M}\right)}}{1+\chi e^{\mathcal{W}_0\left(-\frac{K}{M}\right)}}, & \text{if } \mathbb{E}[S|p^{\mathsf{LL}}_{\mathsf{early}} = \frac{M}{\nu}, \nu] > K. \end{cases}$$

4.3 Comprehensive Access Control

In Sections 4.1 and 4.2, we proposed low-latency AC and energy-efficient AC mechanisms, respectively. However, in practical situation, the eNodeB is required to support accesses both from nodes with low-latency QoS and from nodes with energy-efficient QoS at the same time. First, we assume that eNodeB can identify the access ratio between low-latency access and energy-efficient access. Let γ and $(1-\gamma)$ denote the access ratios of the low-latency access and the energy-efficient access, respectively. Then, when there exists N_t backlogged nodes at time slot t in total, γN_t and $(1-\gamma)N_t$ backlogged nodes attempt random access with low-latency QoS and energy-efficient QoS, respectively, on the average. In the proposed comprehensive access control mechanism, the eNodeB broadcasts two respective ACB factors p^{LL} and $p^{\mathsf{EE}:\theta}$, and each node utilizes one of ACB factors depending on its QoS (low-latency or energy-efficiency) requirement. Consequently, we have the following equation:

$$\gamma N_t p^{\mathsf{LL}} + (1-\gamma)N_t p^{\mathsf{EE}:\theta} = N_t p^{\mathsf{comp.}}, \tag{41}$$

where $p^{\mathsf{comp.}} = \gamma p^{\mathsf{LL}} + (1-\gamma)p^{\mathsf{EE}:\theta}$ denotes a comprehensive ACB factor, which is utilized to estimate N_t as ν. Based on the estimated backlogged size ν, ACB factors p^{LL} and $p^{\mathsf{EE}:\theta}$ are calculated based on the proposed low-latency AC in (20), (22), (26), and (28) and energy-efficient AC mechanisms in (35) and (40), respectively.

5 Estimation on the Number of Backlogged Nodes

We extend the pseudo-Bayesian backlog estimation scheme [18] in order to estimate the number of backlogged nodes for the proposed versatile AC mechanism. The pseudo-Bayesian backlog estimation scheme operates based on the number of available preambles M and the number of undetected (idle) preambles r, which can be exactly identified during the preamble detection procedure at Step 1 of the RA procedure for both C-PCD and E-PCD techniques. Algorithm 1 summarizes the estimation and update algorithm of ν. The eNodeB first estimates ν_0 as the number of preambles M. Then, the estimated value is updated as $\nu = \nu + \Delta\nu$. The estimation offset $\Delta\nu > 0$ is given by [18]

$$\Delta\nu(p) = \mathbb{E}[n|r,p,\nu] - \nu = \nu p\left(\frac{e^{-\frac{p\nu}{M}} - \frac{r}{M}}{1-e^{-\frac{p\nu}{M}}}\right). \tag{42}$$

For various ACB factors, (42) is rewritten as:

$$\Delta\nu(p^{\mathsf{LL}}_{\mathsf{conv}}) = \Delta\nu(p^{\mathsf{LL}}_{\mathsf{early}}) = \frac{Me^{-1} - r}{1-e^{-1}}, \tag{43}$$

$$\Delta\nu(\tilde{p}^{\mathsf{LL}}_{\mathsf{conv}}) = -\ln\left(1-\frac{K}{M}\right)\left(\frac{M}{K}\right)(M-K-r), \tag{44}$$

$$\Delta\nu(\tilde{p}^{\mathsf{LL}}_{\mathsf{early}}) = -\mathcal{W}_0\left(-\frac{K}{M}\right)M\left(\frac{e^{\mathcal{W}_0\left(-\frac{K}{M}\right)} - \frac{r}{M}}{1-e^{\mathcal{W}_0\left(-\frac{K}{M}\right)}}\right), \tag{45}$$

$$\Delta\nu(p^{\mathsf{EE}:\theta}_{\mathsf{conv}}) = -\ln(\theta)M\left(\frac{\theta - \frac{r}{M}}{1-\theta}\right), \tag{46}$$

$$\Delta\nu(p^{\mathsf{EE}:\theta}_{\mathsf{early}}) = -\ln(\xi)M\left(\frac{\frac{\theta}{1+(1-\theta)\chi} - \frac{r}{M}}{1-\frac{\theta}{1+(1-\theta)\chi}}\right). \tag{47}$$

For the comprehensive AC mechanism, the estimation offset is calculated by $\Delta\nu(p^{\mathsf{comp.}})$, where $p^{\mathsf{comp.}} = \gamma p^{\mathsf{LL}} + (1-\gamma)p^{\mathsf{EE}:\theta}$. As in [18], we introduce a boosting factor k_i in order to reflect characteristics of the bursty traffic. It implies that when we observe that the traffic continuously increases based on $\Delta\nu > 0$, we boost the estimation ν. At the last line, the new estimation of ν_i is calculated by $\nu_i = \nu_{i-1} - c_{i-1}$, where c_{i-1} denotes the number of successful accesses on the $(i-1)$th PRACH slot.

Algorithm 1. Estimation and Update Algorithm of ν

1: Initialize $\nu_0 = M$, $k_0 = 0$, and p_0.
2: Count the number of undetected preambles r_{i-1}
3: Calculate $\Delta\nu$ according to (42) with r_{i-1}, p_{i-1}, and ν_{i-1}
4: Update $\nu_{i-1} = \nu_{i-1} + \Delta\nu$
5: **if** $\Delta\nu > 0$ **then**
6: $k_i = k_{i-1} + 1$ and $\nu_{i-1} = \nu_{i-1} + k_i \cdot \Delta\nu$ ▷ Boosting
7: **else**
8: $k_i = 0$ and $\nu_{i-1} = \nu_{i-1}$
9: **end if**
10: $\nu_i = \max(M, \nu_{i-1} - c_{i-1})$ ▷ New estimation of ν for slot i

6 Numerical Results

Table 1 summarizes parameters used for computer simulations. We assume that random access opportunity (PRACH) appears every 1 ms, i.e., $T_{\mathsf{interval}} = 1$ ms, and the number of preambles is equal to $M = 64$.[2] The number of allocable resources K is assumed to vary from 10 to 64 and each group of nodes consists of 10000 nodes which try to connect to the network through the RA procedure. The proposed low-latency AC mechanism aims at the lowest access delay or equivalently the maximum access throughput. For the proposed energy-efficient AC mechanism, the target EE value is assumed to be θ. In particular, for the

2. The proposed versatile AC mechanism can dynamically adjust ACB factors according to the number of preambles, but we focus on the case of the fixed number of preambles in this paper.

TABLE 1
Simulation Parameters and Values

Parameters	Values
The number of preambles, M	64
The number of allocable PUSCH resources, K	$10 \sim 64$
Energy consumption ratio, χ	3
PRACH interval, T_{interval}	1 ms
Activation duration, T_{act}	500 ms
Total number of RA-attempting nodes, N	10000
Access ratio, γ	0.5
Target energy efficiency, θ	0.6/0.8/0.9

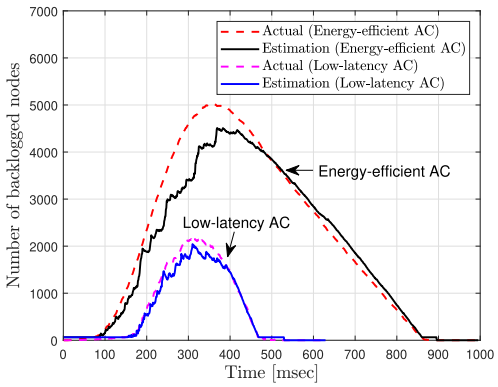


Fig. 3. Estimation of backlogged nodes for $N = 10000$ machine nodes in case of energy-efficient AC and low-latency AC with $K = 64$.

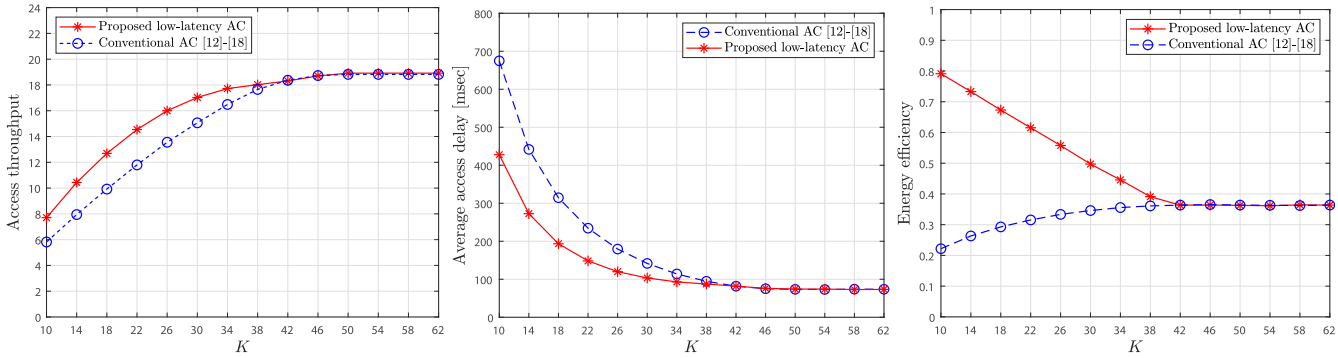
comprehensive AC mechanism, we set the access ratio γ to 0.5. We evaluate the performances of the proposed AC mechanisms in terms of access throughput, average access delay, and energy efficiency. The access delay is defined as the time difference between the access completion time and the activation time of a node. For comparison, the performance of the conventional low-latency AC mechanisms [12], [13], [14], [15], [16], [17], [18] with the ACB factor of $p = \frac{M}{\nu}$ is also illustrated. Among [12], [13], [14], [15], [16], [17], [18], the backlog estimation scheme of [18] is utilized to obtain ν representatively.

6.1 With C-PCD Technique

First of all, we evaluate the performance of backlog estimation. Fig. 3 shows the pseudo-Bayesian estimation result for the number of backlogged nodes in case of energy-efficient AC and low-latency AC, respectively, with $K = 64$. We observed that there exist a gap between the estimation and actual values, especially during the bursty period. However, the estimation algorithm keeps track of the actual values quite well during the traffic descent period. In particular, the backlog estimation under the low-latency AC is quite accurate.

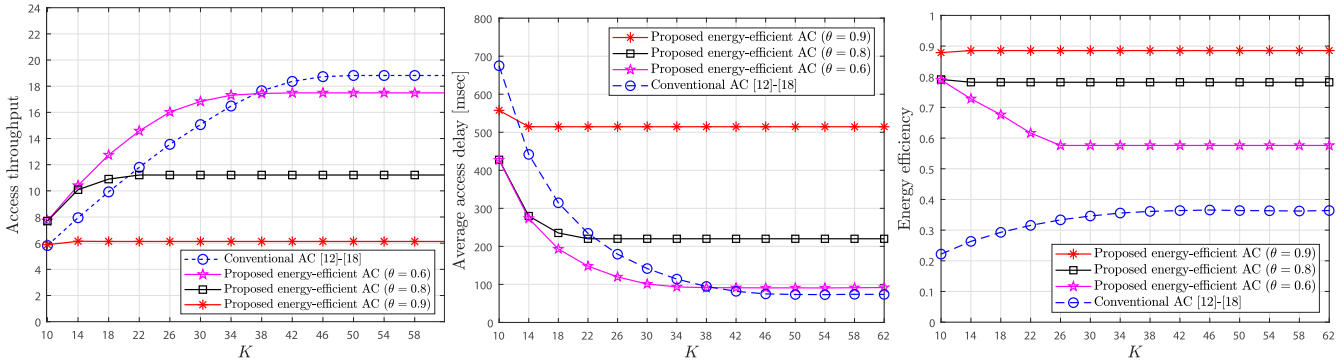
Fig. 4 shows (a) access throughput, (b) average access delay, and (c) energy efficiency of the proposed *low-latency* AC mechanism with the C-PCD technique. The proposed low-latency AC mechanism achieves higher access throughput and energy efficiency, and lower average access delay than those of the conventional low-latency AC mechanism when $K \leq 42$ since it optimizes the ACB factor by considering not only M but also K, while the conventional AC mechanism takes into account only M. According to the result, both the proposed and the conventional low-latency AC mechanism results in the same performances when the number of allocable PUSCH resources is sufficiently large. The performance gap between two mechanisms is significant especially when K is small. For example, the average access delay of the proposed low-latency AC mechanism is lower than 200 ms, while the average access delay of the conventional low-latency AC mechanism is higher than 300 ms when $K = 18$. Furthermore, the EE of the proposed low-latency mechanism (0.80) is approximately 4 times higher than the EE of the conventional low-latency AC mechanism (0.21) when $K = 10$. According to simulation statistics even though they are not shown in the figure, when $K = 10$, the proposed low-latency AC mechanism spends 10983 PUSCH resources and 983 PUSCH resources are wasted due to preamble collisions, while the conventional low-latency AC mechanism spends 16738 PUSCH resources and 6738 PUSCH resources are wasted due to preamble collisions for accommodating 10000 nodes.

Fig. 5 shows (a) access throughput, (b) average access delay, and (c) energy efficiency of the proposed *energy-efficient* AC mechanism with the C-PCD technique for various target EEs ($\theta = 0.9$, 0.8, and 0.6). As shown in Fig. 5c, the proposed energy-efficient AC mechanism satisfies the target EE regardless of K at the cost of increase of average access delay as shown in Fig. 5b. As expected, there exists a tradeoff relationship between the EE and throughput/delay performances. Note that the conventional AC mechanism only yields the maximum EE of 0.38 when $K \geq 42$. For the case that



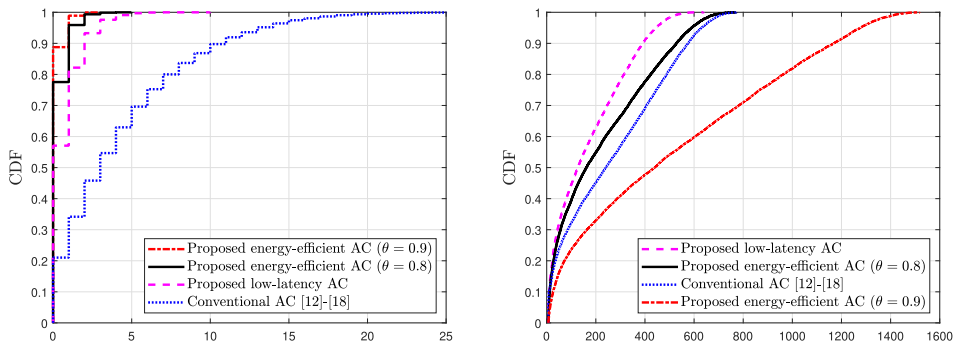
(a) Access throughput (b) Average access delay (c) Energy efficiency

Fig. 4. Performances of the proposed low-latency AC mechanism with the C-PCD technique for varying K.



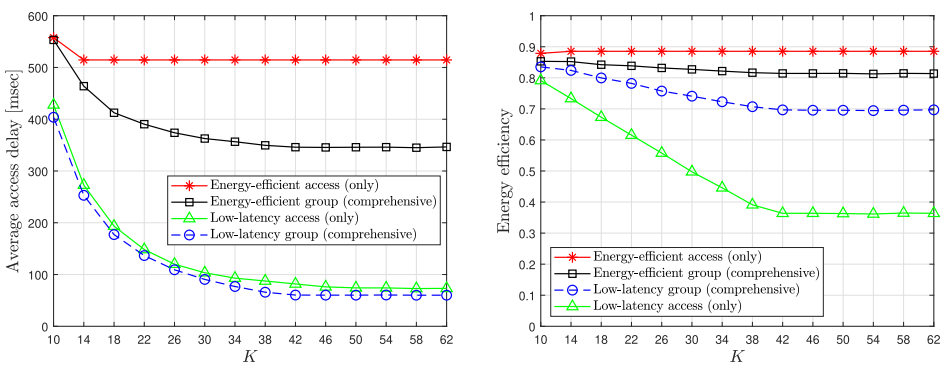
(a) Access throughput (b) Average access delay (c) Energy efficiency

Fig. 5. Performances of the proposed energy-efficient AC mechanism with the C-PCD technique for varying K.



(a) Number of access reattempts (b) Access delay (msec)

Fig. 6. Cumulative distribution function (CDF) of (a) the number of access reattempts and (b) the access delay in energy-efficient AC and low-latency AC when $K = 20$.



(a) Average access delay (msec) (b) Energy efficiency

Fig. 7. Performances of the proposed comprehensive AC mechanism with the C-PCD technique for varying K. The access ratio γ is set to 0.5.

$\theta = 0.6$, the proposed energy-efficient AC mechanism achieves higher EE than the target 0.6 when $K \leq 26$ since the proposed low-latency AC mechanism already achieves higher EE than 0.6 as shown in Fig. 4c. When $\theta = 0.6$ and $K \leq 38$, the proposed energy-efficiency AC mechanism outperforms the conventional AC mechanism in terms of all performance metrics such as access throughput, average access delay, and energy efficiency. In addition, when $\theta = 0.8$ and $K \leq 20$, the same performance tendency is observed.

Fig. 6 shows cumulative distribution functions (CDFs) of (a) the number of access reattempts and (b) the access delay of the proposed AC mechanisms when $K = 20$. The proposed energy-efficient AC mechanism with the target EE of 0.9 and 0.8 controls almost 89 and 78 percent of nodes to succeed in the RA in a single attempt. The average number of reattempts is equal to 0.12 and 0.27 for given EE targets of 0.9 and 0.8, respectively. Note that, with the conventional low-latency AC mechanism, the average number of reattempts is equal to 4.27 and only 21 percent of nodes succeeds in the RA in a single attempt. As shown in Fig. 6b, nodes experience longer access delay with the proposed energy-efficient AC mechanisms when $\theta = 0.9$, compared with both the proposed low-latency AC mechanism and the conventional low-latency AC mechanism. However, if the target EE is reduced from 0.9 to 0.8 with the proposed energy-efficient AC mechanism, then the access delay is significantly reduced from 1168 ms to 528 ms at the 90th percentile.

Figs. 7a and 7b show the average access delay and the energy efficiency of the proposed comprehensive AC mechanism with the C-PCD technique. Here, we set the access ratio

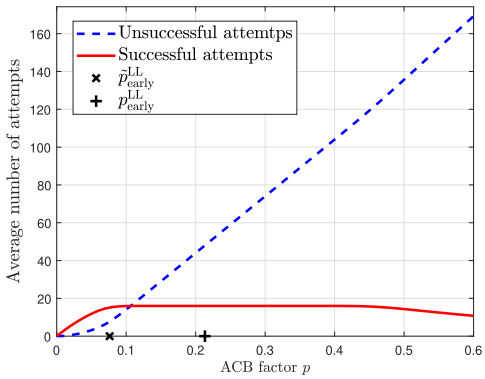


Fig. 8. Average numbers of successful and unsuccessful attempts of the proposed low-latency AC mechanism with the E-PCD technique for varying ACB factor p.

γ to 0.5, which implies that 5000 nodes belong to the low-latency access group and the other 5000 nodes belong to the energy-efficient access group. In particular, we set the target energy efficiency θ to 0.9 for the energy-efficient group. Since the low-latency group and the energy-efficient group share all preamble and PUSCH resources together, the low-latency group under the proposed comprehensive AC shows lower average access delay than that of the low-latency access only (all 10000 nodes attempt low-latency access) as shown in Fig. 4b, and the energy-efficient group under the proposed comprehensive AC also shows lower average access delay than that of the energy-efficient access only (all 10000 nodes attempt energy-efficient access) as shown in Fig. 5 (b). In Fig. 7b, the energy-efficient group under the proposed comprehensive AC achieves the EE of around 0.8, which is lower than 0.9 of the energy-efficient access only as shown in Fig. 5c, whereas the low-latency group under the proposed comprehensive AC benefits considerably from the energy-efficient group in terms of EE, and thus, achieves higher EE than that of the low-latency access only as shown in Fig. 4c. According to the performance result, we observe that access delay is significantly sacrificed in order to achieve a very high energy efficiency during bursty traffic period.

6.2 With E-PCD Technique

It is worth noting that there exist no previous ACB technique that exploits the E-PCD techniques. Thus, with the E-PCD technique, we only evaluate the performance of the proposed AC mechanism by changing its own parameters or compare the proposed low-latency AC mechanism to the proposed energy-efficient AC mechanism.

Fig. 8 shows the number of successful and unsuccessful attempts of the proposed low-latency AC mechanism with the E-PCD technique, which are denoted by $\mathcal{S}(p,K)$ in (29) and $\mathcal{U}(p,K)$ in (30), respectively, when $K=16$ and $N=300$. If $p \le \tilde{p}^{\mathrm{LL}}_{\mathrm{early}}$, then $\mathcal{S}(p,K) > \mathcal{U}(p,K)$. However, $\mathcal{S}(p,K) < \mathcal{U}(p,K)$ as p approaches to $p^{\mathrm{LL}}_{\mathrm{early}}$. There exists a tradeoff relationship between the access throughput and the EE for varying the ACB factor. Hence, we define a novel ACB factor as

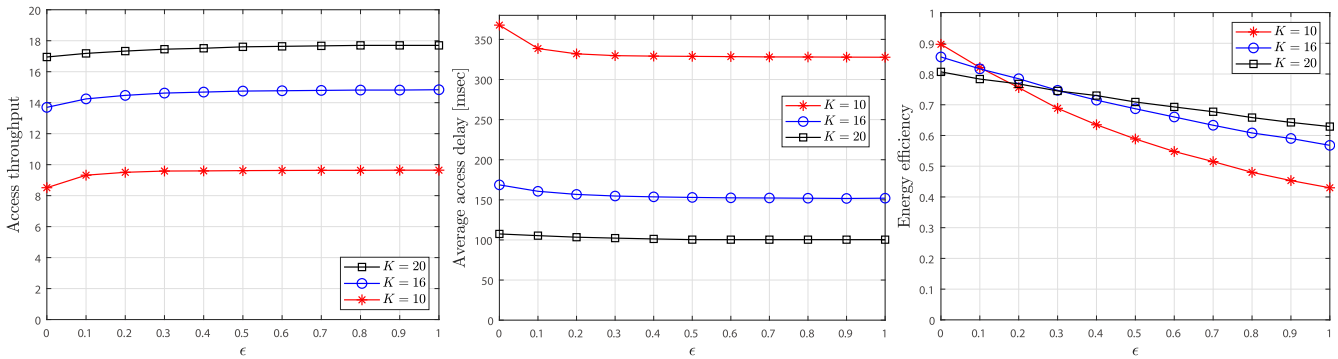
$$p = \epsilon p^{\mathrm{LL}}_{\mathrm{early}} + (1-\epsilon)\tilde{p}^{\mathrm{LL}}_{\mathrm{early}}, \tag{48}$$

for $0 \le \epsilon \le 1$.

Fig. 9 shows (a) access throughput, (b) average access delay, and (c) energy efficiency of the proposed low-latency AC mechanism with the E-PCD technique for varying ϵ of the ACB factor in (48). When $\epsilon = 1$ and $\epsilon = 0$, the ACB factors becomes $p^{\mathrm{LL}}_{\mathrm{early}}$ in (26) and $\tilde{p}^{\mathrm{LL}}_{\mathrm{early}}$ in (28), respectively. As ϵ increases, the EE of the proposed mechanism decreases while the access throughput increases and the average access delay decreases, but they are saturated over some ϵ. All performances become improved as K increases. Furthermore, for a given K, we optimize ϵ so that the EE and the access throughput of the proposed low-latency AC mechanism are maximized while the average access delay is minimized, which is denoted by $\acute{\epsilon}$. For example, the access throughput is maximized and the average access delay is minimized for $\epsilon \in [0.6, 1]$ and the maximum EE is achieved when $\epsilon = 0.6$ for $\epsilon \in [0.6, 1]$ when $K = 16$. Thus, in this case, $\acute{\epsilon} = 0.6$. Then, for a given K, the ACB factor in the sense that the maximum EE and the maximum access throughput are achieved while the average access delay is minimized is obtained by

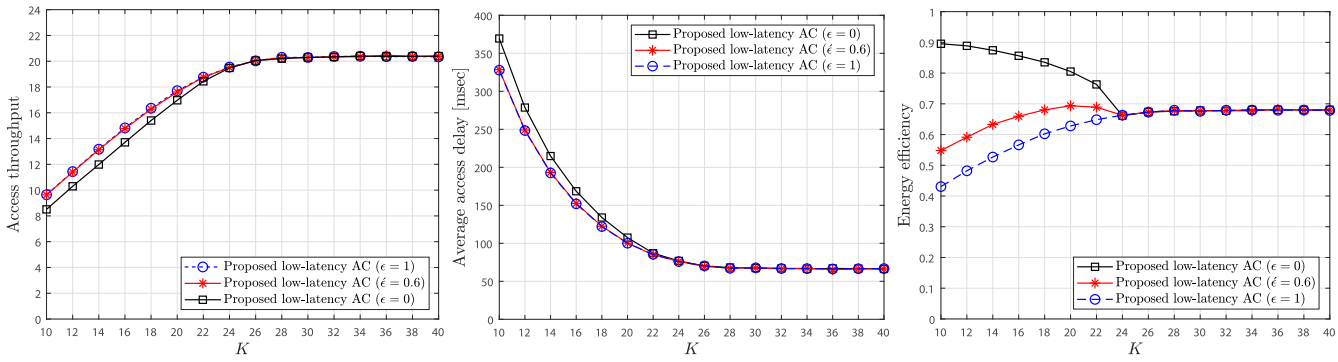
$$\acute{p}^{\mathrm{LL}}_{\mathrm{early}} = \acute{\epsilon} p^{\mathrm{LL}}_{\mathrm{early}} + (1-\acute{\epsilon})\tilde{p}^{\mathrm{LL}}_{\mathrm{early}}. \tag{49}$$

Fig. 10 shows (a) access throughput, (b) the average access delay, and (c) the energy efficiency of the proposed low-latency AC mechanism with the E-PCD technique for various ACB factors including $p^{\mathrm{LL}}_{\mathrm{early}}$ in (26) when $\epsilon = 0$, $\tilde{p}^{\mathrm{LL}}_{\mathrm{early}}$ in (28) when $\epsilon = 0$, and $\acute{p}^{\mathrm{LL}}_{\mathrm{early}}$ with the best $\acute{\epsilon}$ in (49). Even though the proposed low-latency AC mechanism with $\tilde{p}^{\mathrm{LL}}_{\mathrm{early}}$



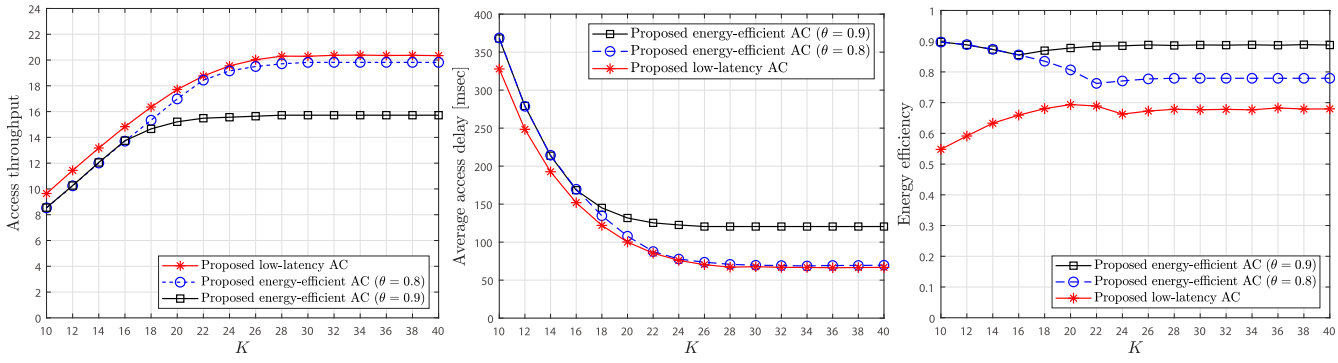
(a) Access throughput (b) Average access delay (c) Energy efficiency

Fig. 9. Performances of the proposed low-latency AC mechanism with the E-PCD technique for varying ϵ when $K = 10$, 16, and 20.



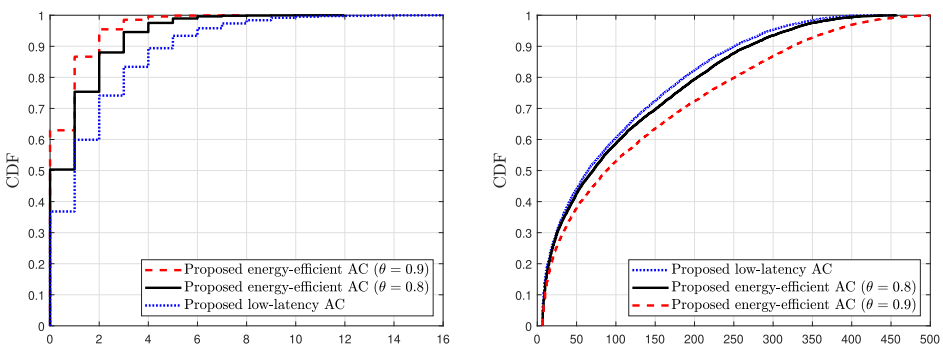
(a) Access throughput (b) Average access delay (c) Energy efficiency

Fig. 10. Performances of the proposed low-latency AC mechanism with the E-PCD technique for varying K.



(a) Access throughput (b) Average access delay (c) Energy efficiency

Fig. 11. Performances of the proposed energy-efficient AC mechanism with the E-PCD technique for varying K.



(a) Number of access reattempts (b) Access delay (msec)

Fig. 12. Cumulative distribution function (CDF) of (a) the number of access reattempts and (b) the access delay of the proposed AC mechanisms with the E-PCD technique when $K = 20$.

($\epsilon = 0$) results in a slightly longer average access delay when $K \leq 24$ than other cases, it significantly improves the EE. When $K > 24$, the proposed low-latency AC mechanism with the E-PCD technique utilizes only $p^{\mathsf{LL}}_{\mathsf{early}}$ since $\mathbb{E}[S|p^{\mathsf{LL}}_{\mathsf{early}} = \frac{M}{\nu}, \nu] \leq K$.

Fig. 11 shows (a) the access throughput, (b) the average access delay, and (c) the energy efficiency of the proposed energy-efficient AC mechanism with the E-PCD technique. As shown in the figure, the proposed energy-efficient AC mechanism effectively satisfies the target EE requirements even though there exist a little dissatisfaction for some K due to approximation error in obtaining the ACB factor.

Fig. 12 shows the CDF of (a) the number of access reattempts and (b) the access delay of the proposed energy-efficient AC mechanism with $\theta = 0.9,\ 0.8$ and the proposed low-latency AC mechanism with the best $\acute{\epsilon}$. With the proposed energy-efficient AC mechanism, 63 percent and 50 percent of nodes succeed the RA in a single attempt and the average number of access reattempts is equal to 0.51 and 0.95 when $\theta = 0.9$ and 0.8, respectively. On the other hand, with the proposed low-latency AC mechanism, 35 percent of nodes succeed the RA in a single attempt and the average number of access reattempts is equal to 1.73. However, as shown in Fig. 12b, at the 90th percentile, the access delays of the proposed energy-efficient AC mechanism with the E-PCD

technique is equal to 325 ms and 270 ms for $\theta = 0.9$ and 0.8, respectively, while that of the proposed low-latency AC mechanism with the E-PCD technique is equal to 250 ms.

7 Conclusion

In this paper, we proposed a versatile access control mechanism for massive IoT networks, which effectively satisfies various QoS requirements: low-latency for mission-critical services and high energy efficiency (EE) for battery-powered sensor applications. We adopted the early collision detection technique as well as the conventional preamble collision detection technique at the first step of the random access (RA) procedure. For optimizing the access class barring factor of the proposed AC mechanism, we considered all radio resources which are involved in the overall RA procedure. The proposed low-latency AC mechanism achieves lowest access delay by aggressively allowing for IoT devices to attempt the RA at the cost of the decreased EE, while the proposed energy-efficient AC mechanism achieves a target EE by restricting attempts from the IoT devices at the cost of increased access delay. Compared to the conventional AC mechanism, the proposed low-latency AC mechanism further reduces the access delay, while improving the EE under diverse resource constraints. Furthermore, we proposed a comprehensive access control mechanism which can simultaneously support two groups of devices requiring low-latency and ultra-low power consumption, respectively.

Acknowledgments

This work was supported in part by the SUTD-ZJU Research Collaboration under Grant SUTD- ZJU/RES/01/2016, in part by the SUTD-ZJU Research Collaboration under Grant SUTD-ZJU/RES/05/2016, in part by the MSIT (Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program (IITP-2019-2017-0-01635) supervised by the IITP(Institute for Information & communications Technology Promotion), and in part by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2018R1C1B6008126).

References

[1] C. X. Mavromoustakis, G. Mastorakis, and J. M. Batalla, *Internet of Things (IoT) in 5G Mobile Technologies*. vol. 8. Berlin, Germany: Springer, 2016.

[2] 3GPP, *Study on Scenarios and Requirements for Next Generation Access Technologies*, 3GPP TR 38.913 V14.3.0, June 2017.

[3] ITU-R, *Minimum Requirements Related to Technical Performance for IMT-2020 Radio Interface(s)*, ITU-R M.2410–0, Nov. 2017.

[4] M. Li, R. Yu, P. Si, and Y. Zhang, "Energy-efficient machine-to-machine (M2M) communications in virtualized cellular networks with mobile edge computing (MEC)," *IEEE Trans. Mobile Comput.*, to be published, doi: 10.1109/TMC.2018.2865312.

[5] H. Li, K. Ota, and M. Dong, "Learning IoT in edge: Deep learning for the internet of things with edge computing," *IEEE Netw.*, vol. 32, no. 1, pp. 96–101, Jan./Feb. 2018.

[6] S.-C. Lin and K.-C. Chen, "Cognitive and opportunistic relay for QoS guarantees in machine-to-machine communications," *IEEE Trans. Mobile Comput.*, vol. 15, no. 3, pp. 599–609, Mar. 2016.

[7] J. Huang, C. Xing, S. Y. Shin, F. Hou, and C. Hsu, "Optimizing M2M communications and quality of services in the IoT for sustainable smart cities," *IEEE Trans. Sustain. Comput.*, vol. 3, no. 1, pp. 4–15, Jan.-Mar. 2018.

[8] A. Sivanathan, H. H. Gharakheili, F. Loi, A. Radford, C. Wijenayake, A. Vishwanath, and V. Sivaraman, "Classifying IoT devices in smart environments using network traffic characteristics," *IEEE Trans. Mobile Comput.*, to be published, doi: 10.1109/TMC.2018.2866249.

[9] T. P. de Andrade, C. A. Astudillo, and N. L. da Fonseca, "Allocation of control resources for machine-to-machine and human-to-human communications over LTE/LTE-A networks," *IEEE Internet Things J.*, vol. 3, no. 3, pp. 366–377, Jun. 2016.

[10] T. P. de Andrade, C. A. Astudillo, L. R. Sekijima, and N. L. da Fonseca, "The random access procedure in long term evolution networks for the internet of things," *IEEE Commun. Mag.*, vol. 55, no. 3, pp. 124–131, Mar. 2017.

[11] N. Xia, H.-H. Chen, and C.-S. Yang, "Radio resource management in machine-to-machine communications-a survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 1, pp. 791–828, Jan.-Mar. 2018.

[12] S. Duan, V. Shah-Mansouri, and V. W. S. Wong, "Dynamic access class barring for M2M communications in LTE networks," in *Proc. IEEE GLOBECOM*, Dec. 2013, pp. 4747–4752.

[13] Z. Wang and V. W. Wong, "Joint access class barring and timing advance model for machine-type communications," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2014, pp. 2357–2362.

[14] A. Laya, L. Alonso, and J. Alonso-Zarate, "Is the random access channel of LTE and LTE-A suitable for M2M communications? A survey of alternatives," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 1, pp. 4–16, Jan.-Mar. 2014.

[15] M. Tavana, V. Shah-Mansouri, and V. W. Wong, "Congestion control for bursty M2M traffic in LTE networks," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2015, pp. 5815–5820.

[16] Z. Wang and V. W. Wong, "Optimal access class barring for stationary machine type communication devices with timing advance information," *IEEE Trans. Wireless Commun.*, vol. 14, no. 10, pp. 5374–5387, Oct. 2015.

[17] S. Duan, V. Shah-Mansouri, Z. Wang, and V. W. Wong, "D-ACB: Adaptive congestion control algorithm for bursty M2M traffic in LTE networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 12, pp. 9847–9861, Dec. 2016.

[18] H. Jin, W. Toor, B. C. Jung, and J. B. Seo, "Recursive Pseudo-Bayesian access class barring for M2M communications in LTE systems," *IEEE Trans. Veh. Technol.*, vol. 66, no. 9, pp. 8595–8599, Sep. 2017.

[19] S. Vural, N. Wang, P. Bucknell, G. Foster, R. Tafazolli, and J. Muller, "Dynamic preamble subset allocation for RAN slicing in 5G networks," *IEEE Access*, vol. 6, pp. 13 015–13 032, 2018.

[20] C. Y. Ho and C.-Y. Huang, "Energy-saving massive access control and resource allocation schemes for M2M communications in OFDMA cellular networks," *IEEE Wireless Commun. Lett.*, vol. 1, no. 3, pp. 209–212, Jun. 2012.

[21] H. S. Dhillon, H. C. Huang, H. Viswanathan, and R. A. Valenzuela, "Power-efficient system design for cellular-based machine-to-machine communications," *IEEE Trans. Wireless Commun.*, vol. 12, no. 11, pp. 5740–5753, Nov. 2013.

[22] A. Azari and G. Miao, "Energy efficient mac for cellular-based M2M communications," in *Proc IEEE Global Conf. Signal and Inform. Process.*, Dec. 2014, pp. 128–132.

[23] G. Miao, A. Azari, and T. Hwang, "E^2-MAC: Energy efficient medium access for massive M2M communications," *IEEE Trans. Commun.*, vol. 64, no. 11, pp. 4720–4735, Nov. 2016.

[24] A. Aijaz, M. Tshangini, M. R. Nakhai, X. Chu, and A.-H. Aghvami, "Energy-efficient uplink resource allocation in LTE networks with M2M/H2H co-existence under statistical QoS guarantees," *IEEE Trans. Commun.*, vol. 62, no. 7, pp. 2353–2365, Jul. 2014.

[25] M. A. Mehaseb, Y. Gadallah, A. Elhamy, and H. Elhennawy, "Classification of LTE uplink scheduling techniques: An M2M perspective," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1310–1335, Apr.-Jun. 2016.

[26] D. Zhang, Z. Zhou, S. Mumtaz, J. Rodriguez, and T. Sato, "One integrated energy efficiency proposal for 5G IoT communications," *IEEE Internet Things J.*, vol. 3, no. 6, pp. 1346–1354, Dec. 2016.

[27] Z. Yang, W. Xu, Y. Pan, C. Pan, and M. Chen, "Energy efficient resource allocation in machine-to-machine communications with multiple access and energy harvesting for IoT," *IEEE Internet Things J.*, vol. 5, no. 1, pp. 229–245, Feb. 2018.

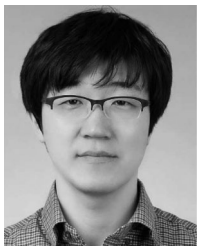
[28] A. Azari and G. Miao, "Network lifetime maximization for cellular-based M2M networks," *IEEE Access*, vol. 5, pp. 18 927–18 940, 2017.

[29] N. M. Balasubramanya, L. Lampe, G. Vos, and S. Bennett, "On timing reacquisition and enhanced primary synchronization signal (ePSS) design for energy efficient 3GPP LTE MTC," *IEEE Trans. Mobile Comput.*, vol. 16, no. 8, pp. 2292–2305, Aug. 2017.

[30] H. S. Jang, S. M. Kim, H.-S Park, and D. K. Sung, "An early preamble collision detection scheme for cellular M2M random access," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 5974–5984, Jul. 2017.
[31] H. S. Jang, S. M. Kim, H.-S Park, and D. K. Sung, "A preamble collision resolution scheme via tagged preambles for cellular IoT/M2M communications," *IEEE Trans. Veh. Technol.*, vol. 67, no. 2, pp. 1825–1829, Feb. 2018.
[32] 3GPP, *Study on RAN Improvements for Machine-Type Communications*, 3GPP TR 37.868 V11.0.0, Oct. 2011.
[33] 3GPP, *Medium Access Control (MAC) Protocol Specification (Release 14)*, 3GPP TS 36.321 V14.6.0, Mar. 2018.
[34] S. Sesia, I. Toufik, and M. Baker, *LTE - The UMTS Long Term Evolution: From Theory to Practice*. Hoboken, NJ, USA: Wiley, 2009.
[35] H. S. Jang, B. C. Jung, and D. K. Sung, "Dynamic access control with resource limitation for group paging-based cellular IoT systems," *IEEE Internet Things J.*, vol. 5, no. 6, pp. 5065–5075, Dec. 2018.
[36] N. Zhang, G. Kang, J. Wang, Y. Guo, and F. Labeau, "Resource allocation in a new random access for M2M communications," *IEEE Commun. Lett.*, vol. 19, no. 5, pp. 843–846, May 2015.
[37] D. Magrin, C. Pielli, C. Stefanovic, and M. Zorzi, "Enabling LTE RACH collision multiplicity detection via machine learning," arXiv:1805.11482, 2018.
[38] 3GPP, *[70bis#11]-LTE: MTC LTE Simulations*, 3GPP 3GPP TSG-RAN WG2 meeting# 71 R2–104663, Aug. 2010.
[39] W. Feller, *An Introduction to Probability Theory and Its Application*. Hoboken, NJ, USA: Wiley, 1968.
[40] R. Rivest, "Network control by bayesian broadcast," *IEEE Trans. Inf. Theory*, vol. IT-33, no. 3, pp. 323–328, May 1987.
[41] R. M. Corless, G. H. Gonnet, D. E. Hare, D. J. Jeffrey, and D. E. Knuth, "On the LambertW function," *Advances Comput. Math.*, vol. 5, no. 1, pp. 329–359, 1996.

Han Seung Jang (S'14–M'18) received the BS degree in electronics and computer engineering from Chonnam National University, Gwangju, Korea, in 2012, and the MS and PhD degrees in electrical engineering from the Korea Advanced Institute for Science and Technology (KAIST), Daejoen, Korea, in 2014 and 2017, respectively. From May 2018 to February 2019, he was a post-doctoral fellow with the Information Systems Technology and Design (ISTD) Pillar, Singapore University of Technology and Design (SUTD), Singapore. He was also a post-doctoral fellow from September 2017 to April 2018 with Chungnam National University, Daejoen, Korea. He is currently an assistant professor with the School of Electrical, Electronic Communication, and Computer Engineering, Chonnam National University, Yeosu, Korea. His research interests include cellular Internet-of-things (IoT)/machine-to-machine (M2M) communications, machine learning, smart grid, and energy ICT. He is a member of the IEEE.

Hu Jin (S'07–M'12–SM'18) received the BE degree in electronic engineering and information science from the University of Science and Technology of China, Hefei, China, in 2004, and the MS and PhD degrees in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2006 and 2011, respectively. From 2011 to 2013, he was a post-doctoral fellow with The University of British Columbia, Vancouver, BC, Canada. From 2013 to 2014, he was a research professor with Gyeongsang National University, Tongyeong, South Korea. Since 2014, he has been with the Division of Electrical Engineering, Hanyang University, Ansan, South Korea, where he is currently an associate professor. His research interests include medium-access control and radio resource management for random access networks and scheduling systems considering advanced signal processing and queueing performance. He is a senior member of the IEEE.



Bang Chul Jung (S'02–M'08–SM'14) received the BS degree in electronics engineering from Ajou University, Suwon, Korea, in 2002, and the MS and PhD degrees in electrical & computer engineering from KAIST, Daejeon, Korea, in 2004 and 2008, respectively. He was a senior researcher/research professor with the KAIST Institute for Information Technology Convergence, Daejeon, Korea, from January 2009 to February 2010. From March 2010 to August 2015, he was a Faculty of Gyeongsang National University. He is currently an associate professor in the Department of Electronics Engineering, Chungnam National University, Daejeon, Korea. He has served as associate editor of *IEICE Transactions on Fundamentals of Electronics, Communications, and Computer Sciences* since 2018. His research interests include wireless communication systems, internet-of-things (IoT) communications, statistical signal processing, information theory, interference management, random access, radio resource management, cooperative relaying techniques, in-network computation, and mobile computing. He was the recipient of several awards including the Fifth IEEE Communication Society Asia-Pacific Outstanding Young Researcher Award in 2011 and he has been selected as a winner of Haedong Young Scholar Award in 2015, which is sponsored by the Haedong foundation and given by Korea Institute of Communication and Information Science (KICS), and the Best Paper Award of KICS Journal in 2018, and several Best Paper Awards of KICS Conferences in 2017/2018. He is a senior member of the IEEE.

Tony Q.S. Quek (S'98-M'08-SM'12-F'18) received the BE and ME degrees in electrical and electronics engineering from the Tokyo Institute of Technology, Tokyo, Japan, in 1998 and 2000, respectively, and the PhD degree in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, MA, in 2008. Currently, he is a tenured associate professor with the Singapore University of Technology and Design (SUTD). He also serves as the Acting Head of ISTD Pillar and the deputy director of the SUTD-ZJU IDEA. His current research topics include wireless communications and networking, internet-of-things, network intelligence, wireless security, and big data processing. He has been actively involved in organizing and chairing sessions, and has served as a member of the technical program committee as well as symposium chairs in a number of international conferences. He is currently an elected member of IEEE Signal Processing Society SPCOM Technical Committee. He was an executive editorial committee member and an editor for transactions and journals.. He is a co-author of the books *Small Cell Networks: Deployment, PHY Techniques, and Resource Allocation* and the book *Cloud Radio Access Networks: Principles, Technologies, and Applications* by Cambridge University Press in 2013 and 2017, respectively. Dr. Quek was honored with several awards including the 2017 CTTC Early Achievement Award, the 2017 IEEE ComSoc AP Outstanding Paper Award, and the 2016-2018 Clarivate Analytics Highly Cited Researcher. He is a distinguished lecturer of the IEEE Communications Society and a fellow of the IEEE.