Dynamic Access Control With Resource Limitation for Group Paging-Based Cellular IoT Systems

Han Seung Jang, *Member, IEEE*, Bang Chul Jung, *Senior Member, IEEE*, and Dan Keun Sung, *Life Fellow, IEEE*

***Abstract*—In cellular Internet-of-Things (IoT) systems, system overload may occur during a random access (RA) procedure under a limited number of preamble resources and physical uplink shared channel (PUSCH) resources especially when there exist massive IoT devices in a cell. In order to resolve the system overload, the commercial system like 3GPP LTE adopted a group paging (GP)-based uplink access technique, but it has been known that the performance of the GP-based technique drastically degrades as the number of devices increases. In this paper, we first propose a dynamic access control (DAC) mechanism for the GP-based cellular massive IoT system, which dynamically adjusts RA-attempting probability by considering not only the number of available preambles but also the number of available PUSCH resources. We also intelligently combine the proposed DAC mechanism with an early preamble collision detection technique to further improve the RA performance of the cellular IoT system. Through extensive computer simulations, we show that the proposed DAC mechanism outperforms the conventional access control mechanisms, which consider only the number of available preamble resources, in terms of GP completion time, PUSCH resource efficiency, transmission efficiency, and energy efficiency.**

***Index Terms*—Cellular Internet-of-Things (IoT) systems, early preamble collision detection (e-CD), group paging (GP), machine-to-machine (M2M), physical uplink shared channel (PUSCH), preamble, random access (RA).**

I. Introduction

RECENTLY, wireless network operators have provided emerging cellular Internet-of-Things (IoT) and machine-to-machine (M2M) communication services since it is expected that the number of IoT devices reaches tens of billions in the future [1], [2]. In the cellular IoT systems, most IoT devices (machine nodes) tend to remain at a radio resource control (RRC) *idle* state in order to reduce battery power consumption and signaling burden required for maintaining the RRC connected state. When an IoT device has data to send, it initiates a random access (RA) procedure to make a connection with a base station, i.e., eNodeB in 3GPP LTE systems [3] and narrowband IoT systems [4], [5]. Fig. 1 shows the conventional RA procedure of the LTE network and its collision detection technique [6]. At the first step of the RA procedure, each node transmits a randomly selected preamble among available preambles on a physical random access channel (PRACH) slot [7]. If multiple nodes transmit the same preamble, the preamble collision occurs. However, the eNodeB cannot detect preamble collisions at the first step in the conventional RA procedure. At the second step of the RA procedure, the eNodeB sends the RA response (RAR) messages to the corresponding nodes by using a physical downlink control channel (PDCCH) or a physical downlink shared channel (PDSCH). The RAR message contains the information of the detected preamble, physical uplink shared channel (PUSCH) resource grant, and timing advance value. At the third step of the RA procedure, the node which received the RAR message transmits a connection request message through the allocated PUSCH resource by the RAR message. If multiple nodes, which used the same preamble at the first step, send connection request messages on the same PUSCH resource, then the eNodeB may fail to decode the multiple messages, and it finally recognizes the preamble collision based on the decoding failure. At the fourth step of RA procedure, the eNodeB transmits a contention resolution message to the node on an PDSCH for the nodes that succeeded in decoding of message at the third step. No contention resolution messages are delivered to the nodes that failed to decode the message at the third step.

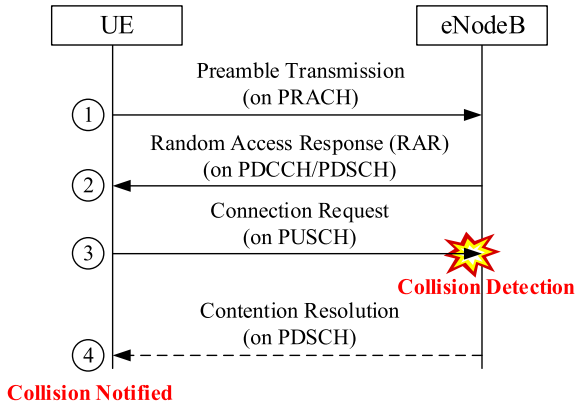


Fig. 1. Conventional RA procedure and collision detection technique.

Manuscript received February 19, 2018; revised August 26, 2018; accepted September 25, 2018. Date of publication October 1, 2018; date of current version January 16, 2019. This work was supported by the Basic Science Research Program through the NRF funded by the Ministry of Science and ICT under Grant NRF-2016R1A2B4014834. *(Corresponding author: Bang Chul Jung.)*

H. S. Jang is with the Information Systems Technology and Design Pillar, Singapore University of Technology and Design, Singapore 487372 (e-mail: hanseung_jang@sutd.edu.sg).

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

D. K. Sung is with the School of Electrical Engineering, Korea Advanced Institute of Science and Technology, Daejeon 34141, South Korea (e-mail: dksung@kaist.ac.kr).

Digital Object Identifier 10.1109/JIOT.2018.2873429

Unlike human-to-human communications, a severe RA overload may occur during the RA procedure since a large number of IoT devices attempt RAs at the same time in general cellular IoT systems. Hence, various techniques have been proposed to resolve the RA overload problem of the cellular IoT systems in [8]–[10]. Generally, there exist two types of RA overload control methods: 1) push-based and 2) pull-based control methods [11], [12]. In the push-based methods, IoT devices transmit data spontaneously and the eNodeB attempts to mitigate RA overload by exploiting various techniques, including access class barring (ACB), separating the RA channels, dynamic resource allocation, backoff-based schemes, etc. [13]. The ACB-based schemes inherently control the number of simultaneous accesses with an ACB factor ranging between 0 and 1 or different back-off times. In particular, dynamic ACB schemes have been actively investigated in [14]–[20]. On the other hand, in the pull-based methods, the eNodeB utilizes paging messages to activate certain devices when it requires information from them [21]. With the paging-based schemes, in general, the RA overload is more effectively controlled, compared to the push-based methods since it explicitly controls the access attempts from the IoT devices [21]. However, it becomes a significant burden to page each of devices individually as the number of IoT devices increases in the network.

In order to resolve this problem, a group paging (GP) technique was proposed in 3GPP systems, where the IoT devices are grouped together according to specific metrics, and each group is assigned a single identifier (ID) [22], [23]. When a certain group is paged by its group ID, all devices belonging to the group start the RA procedure at the same time in order to obtain RRC-connections. An analytical model was introduced to investigate the performance of the GP-based system, and a decision rule for both the number of reserved RA resources and the group size was proposed in [24]. A traffic scattering scheme was proposed to improve the energy efficiency (EE) of IoT devices in the GP-based LTE network, which prevents all devices from simultaneously attempting the RA procedure [25]. In addition, a dynamic preamble resource allocation scheme was proposed for the GP-based LTE-A network, which dynamically adjusts the number of preamble resources that are allocated to users in a certain group based on the estimated number of contending users in each RA slot [26]. Aforementioned schemes for the GP-based cellular networks considered only the preamble resources to alleviate the system overload problem. However, recently, it was reported that a lack of PUSCH and PDCCH resources is another bottleneck of the RA procedure [27], [28]. Moreover, the above schemes assumed that the preamble collisions are perfectly detected at the first step of the RA procedure, but the preamble collisions may be detected after packet decoding of the uplink data at the third step of the RA procedure in practice. Finally, the conventional schemes need to monitor/estimate the number of contending devices in each RA slot, but it is not feasible in general [20].

Hence, we propose a dynamic access control (DAC) mechanism for GP-based cellular IoT networks in this paper, which adjusts the RA-attempting probability (RAP) of the paged devices by considering the varying amount of preamble resources, RAR resources, and PUSCH resources. We apply the early collision detection (e-CD) technique [29] to the newly proposed DAC mechanism for estimating the number of contending devices, which further alleviates the RA overload, increases the PUSCH resources utilization, and improves the EE. It will be shown that the proposed DAC mechanism outperforms the conventional access control (C-AC) mechanism in terms of GP completion (GPC) time, PUSCH resource efficiency, transmission (TX) efficiency, and EE. It is worth noting that the proposed DAC mechanism can be applied to the GP based narrow band IoT systems [5]. The major contributions of this paper are summarized as follows.

1) We carefully consider the whole radio resources which are utilized in the RA procedure such as the number of available preambles, the number of available RAR messages, and the number of PUSCH resources in order to adaptively control RAP for efficiently accommodating a large number of devices in the GP-based IoT networks. It is trivial to apply the conventional ACB scheme that considers only the number of available preambles for access control (AC) to the GP-based IoT network.
2) We investigate the effect of the e-CD technique on the performance of the proposed DAC technique in the GP-based IoT networks even though most existing studies consider only the conventional preamble collision detection technique.
3) We mathematically analyze the proposed DAC mechanism in terms of various performance metrics, such as the GPC time, the PUSCH resource efficiency, the transmission efficiency, and the EE. In particular, the PUSCH resource efficiency, the transmission efficiency, and the EE are first analyzed in this paper for GP-based IoT networks to the best our knowledge.
4) The comprehensive performance results of the proposed DAC mechanism give insights on how to manage the radio resources, such as PRACH and PUSCH and how to efficiently satisfy the performance requirements of various traffic types in GP-based cellular IoT networks.

The rest of this paper is organized as follows. In Section II, we describe a system model considered in this paper. In Section III, we propose a DAC mechanism for the GP-based cellular IoT network with a massive number of devices. The proposed DAC mechanism is mathematically analyzed and validated via extensive simulations in Sections IV and V, respectively. Finally, the conclusions are drawn in Section VI.

II. System Model

A. GP-Based Cellular IoT Networks

We consider a single cell IoT network consisting of a single eNodeB and a number of machine nodes (IoT devices). We assume that there exist N_G IoT groups and each group consists of N_{MN} machine nodes. Thus, there exist $N_G \cdot N_{\mathrm{MN}}$ machine nodes in a cell. For example, Fig. 2 shows the GP-based cellular IoT network, where there exist six groups in a cell and

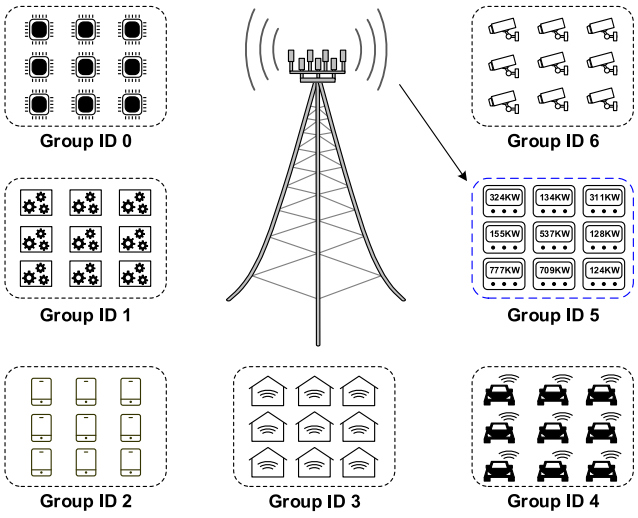


Fig. 2. GP-based IoT network. In this figure, the eNodeB pages the fifth group of smart meters.

nine machine nodes[1] in each group. In general, the group is set up according to its applications such as smart metering, smart transportation, e-health care, public safety, surveillance, remote maintenance, and control. We assume that all nodes in the same group are assigned to the same group identity (ID). Each group is paged with the group ID included in the paging message [21]. Upon receiving the paging message, all the nodes in the corresponding group, i.e., N_{MN} nodes, initiate RA procedures for RRC-connections. In Fig. 2, for example, the fifth group is paged by the eNodeB, and all machine nodes in the fifth group initiate the RA procedures simultaneously. In this paper, we focus on a single group operation without loss of generality since each group is exclusively controlled by the eNodeB with the GP signaling.

B. Access Control

Fig. 3 shows the overall procedure of the AC mechanism for the GP-based IoT network. The time interval of a group RA (GRA) opportunity for the paged group is denoted by T_{interval}. In a certain GRA opportunity, a dedicated PRACH is utilized for nodes in the group to transmit preambles, and the corresponding RAR messages are utilized to allocate resources to them.[2] The number of available preambles and the number of allocable PUSCH resources for each group are assumed to be M and V, respectively. In particular, the number of allocable PUSCH resources are decided by the minimum number of available RAR messages and available PUSCH resources. In this paper, we also assume that one PUSCH resource is allocated to a single node, and a small size data is delivered on the PUSCH resource in the third step of the RA procedure [30]. Then, up to V nodes can be accommodated to send data on the PUSCH resources.

The GRA overload is controlled by the AC. The eNodeB first notifies the RAP value of the ith GRA opportunity $p(i) \in [0, 1]$ at the beginning of the ith GRA opportunity by using the paging message. Then, each node generates a random number $q \in [0, 1]$ and compares q with $p(i)$. If $q \leq p(i)$ (*access check*), the machine node attempts an RA in the ith GRA opportunity. Otherwise, it defers an RA attempt to the $(i+1)$th GRA opportunity. The number of machine nodes which succeed in the RA in the ith GRA opportunity is assumed to be $N_{\mathsf{RA}}(i)$. Then, the number of backlogged nodes to attempt the RA at the beginning of the $(i+1)$th GRA opportunity is given by

$$N_{\mathsf{backlog}}(i+1) = N_{\mathsf{MN}} - \sum_{j=1}^{i} N_{\mathsf{RA}}(j). \tag{1}$$

C. Early Collision Detection

Fig. 4 shows the e-CD technique. If the e-CD technique is adopted at the eNodeB, then the eNodeB can detect the preamble collisions at the first step of the RA procedure [29], [31], [32]. Thus, the eNodeB sends the RAR messages only for the collision-free preambles at the second step of RA procedure while the eNodeB does not send the RAR messages for the collided preambles. In this paper, we consider both the conventional collision detection and e-CD techniques in the GP-based cellular IoT systems.

III. Dynamic Access Control Mechanisms With Group Paging

In this section, we propose two DAC mechanisms for the GP-based cellular IoT network.

1) *P-DAC 1:* Proposed DAC mechanism with the conventional collision detection technique.
2) *P-DAC 2:* Proposed DAC mechanism with the e-CD technique.

As shown in Fig. 3, in the ith GRA opportunity, the eNodeB calculates the RAP value $p(i)$, and then $p(i)$ is sent to machine nodes by using the paging message. Upon receiving the RAP value, each machine nodes in the corresponding group determines whether it attempts an RA or not. If it passes the access check, it attempts an RA, and then receives an RAR message containing a PUSCH resource grant. Finally, it transmits its own RA-step 3 data on the allocated PUSCH resource. Nodes which do not pass the access check, cannot receive their corresponding RAR messages (failure in PUSCH resource schedule), or are notified that their preambles collided, wait until the next GRA opportunity.

A. Proposed DAC Mechanism With Conventional Collision Detection Technique (P-DAC 1)

M and V denote the number of available preambles and the number of allocable PUSCH resources, respectively. In this mechanism, the eNodeB cannot detect preamble collisions at the first step of the RA procedure, and, thus, it allocates the PUSCH resources for all activated preambles regardless of collisions. Let n denote the number of machine nodes that

[1]For a representation simplicity, we show only nine machine nodes belonging to a single group. However, in practical systems, the number of machine nodes in a group is significantly larger. In case of smart metering, the number of smart meters in a group is expected to be 35 670 with a radius of 2 km in urban London [13].

[2]Since the machine nodes in a group utilize the dedicated PRACH, the eNodeB can exactly count the number of RA-success nodes and the number of backlogged nodes.

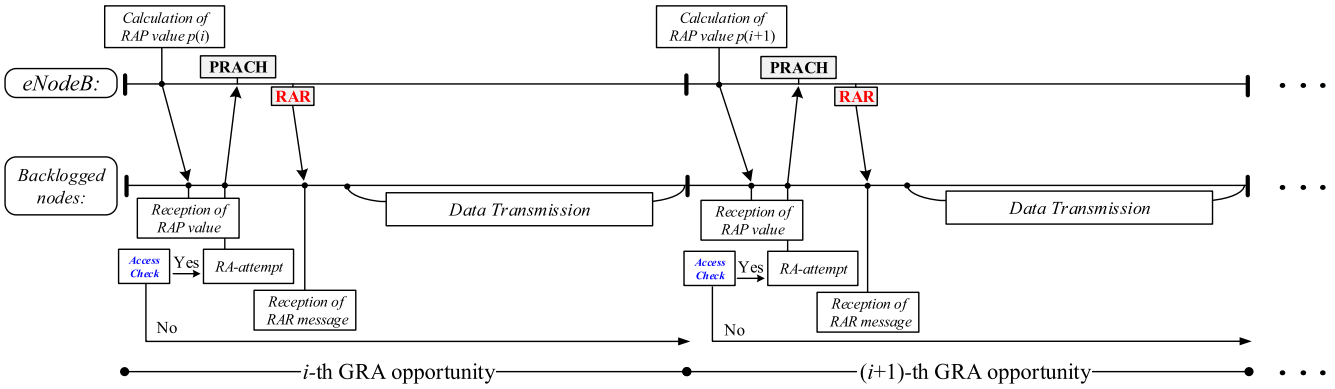


Fig. 3. Overall procedure of the AC mechanism for GP-based IoT networks.

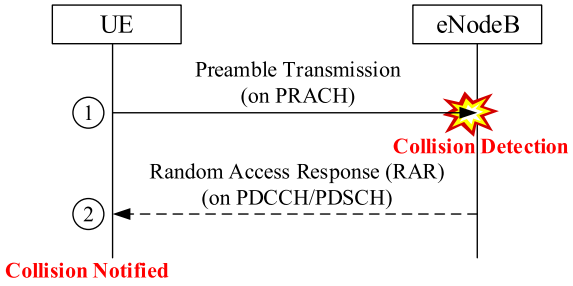


Fig. 4. e-CD technique.

generate the random number q such that $q \leq p(i)$ in the ith GRA opportunity. Thus, n machine nodes are assumed to pass the access check in the ith GRA opportunity. Then, for given n, the probability that an arbitrary preamble among M preambles is selected by only a *single* node is given by

$$P_{\mathsf{s}}(n) = n\left(\frac{1}{M}\right)\left(1-\frac{1}{M}\right)^{n-1} \tag{2}$$

and the average number of successful (collision-free) preamble transmissions is given by

$$N_{\mathsf{success}}(n) = P_s(n)M. \tag{3}$$

Similarly, for given n, the probability that an arbitrary preamble is activated is given by

$$P_{\mathsf{a}}(n) = 1-\left(1-\frac{1}{M}\right)^{n} \tag{4}$$

and the average number of activated preambles is given as

$$N_{\mathsf{active}}(n) = P_a(n)M. \tag{5}$$

Note that $P_{\mathsf{a}}(n)$ is the probability that at least one machine node selects a particular preamble, and it is activated. Thus, the activated preambles include both of collision-free preambles selected by a single node and collided preambles selected by two or more nodes.

Fig. 5 shows the average number of successful (collision-free) preamble transmissions, $N_{\mathsf{success}}(n)$, and the average number of activated preambles, $N_{\mathsf{active}}(n)$, for varying n when $M = 64$ and $V = 50$, 30, and 20. Basically, $N_{\mathsf{success}}(n)$ is maximized at $n^* = -1/\ln\alpha$ by solving $dN_{\mathsf{success}}(n)/dn = 0$,

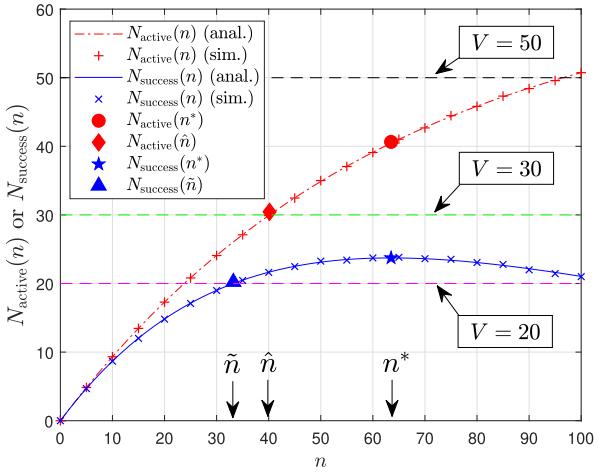


Fig. 5. Average number of successful (collision-free) preamble transmissions and the average number of activated preambles for varying n when $M = 64$ and $V = 50$, 30, and 20.

where $\alpha = (1 - 1/M)$. Since $N_{\mathsf{success}}(n)$ decreases when $n > n^*$, we only consider the case when $n \leq n^*$. Basically, the eNodeB should allocate PUSCH resources to each activated preamble for step 3 of RA procedure. However, when the allocable PUSCH resources are smaller than the number of activated preambles $N_{\mathsf{active}}(n)$, the eNodeB should choose V activated preambles among $N_{\mathsf{active}}(n)$. Thus, we should consider both of cases $N_{\mathsf{active}}(n) \leq V$ and $N_{\mathsf{active}}(n) > V$. When $n \leq n^*$ for given V and n, the average number of PUSCH resources assigned to the collision-free preambles is given by

$$V_{\mathsf{success}}(n, V) = \begin{cases} N_{\mathsf{success}}(n) & \text{if } N_{\mathsf{active}}(n^*) \leq V \\ V \times \frac{N_{\mathsf{success}}(n)}{N_{\mathsf{active}}(n)} & \text{if } N_{\mathsf{active}}(n^*) > V \end{cases} \tag{6}$$

where $V_{\mathsf{success}}(n, V)$ implies the average number of machine nodes that succeed in uplink data transmission via the RA procedure. When $N_{\mathsf{active}}(n^*) \leq V$ (e.g., $V = 50$ in Fig. 5), $N_{\mathsf{success}}(n^*)$ nodes succeed in uplink data transmission via the RA procedure by allowing only n^* nodes among the backlogged $N_{\mathsf{backlog}}(i)$ nodes to pass the access check with the following RAP value:

$$p^*(i) = \min\left\{-\frac{1}{N_{\mathsf{backlog}}(i)\ln\alpha}, 1\right\}. \tag{7}$$

On the other hand, when $N_{\text{active}}(n^*) > V$ (refer to $V = 30$ in Fig. 5), we can obtain $\hat{n} = \ln(1 - V/M)/\ln\alpha$ which maximizes $V_{\text{success}}(n, V)$ by solving $V = N_{\text{active}}(n)$. To allow only $\hat{n}$ nodes passing the access check among the backlogged $N_{\text{backlog}}(i)$ nodes, the RAP value $\hat{p}(i)$ is set to

$$\hat{p}(i) = \min\left\{\frac{\ln(1 - V/M)}{N_{\text{backlog}}(i)\ln\alpha}, 1\right\}. \tag{8}$$

Now, we have the following lemma.

Lemma 1: When $N_{\text{active}}(n^*) > V$, setting the RAP value to $\hat{p}(i)$ as in (8) achieves a larger number of RA-success nodes and wastes a smaller number of PUSCH resources on average sense than the C-AC mechanism which always generates $p^*(i)$ as in (7) without considering the number of allocable PUSCH resources.

Proof: When $N_{\text{active}}(n^*) > V$, the average number of RA-success nodes is given by

$$V_{\text{success}}(n, V) = \frac{V \cdot N_{\text{success}}(n)}{N_{\text{active}}(n)} = \frac{Vn\alpha^{(n-1)}}{(1 - \alpha^n)M} \tag{9}$$

whose derivative with respect to n is given by

$$\frac{dV_{\text{success}}(n, V)}{dn} = \frac{V\alpha^{(n-1)}M(1 - \alpha^n + n\ln\alpha)}{\{(1 - \alpha^n)M\}^2} < 0$$

since $V\alpha^{(n-1)}M/\{(1 - \alpha^n)M\}^2 > 0$ and $(1 - \alpha^n + n\ln\alpha) < 0$, where $n > 0$ and $0 < \alpha \leq 1$. It implies that $V_{\text{success}}(n, V)$ is a decreasing function, and, thus, $V_{\text{success}}(\hat{n}, V) > V_{\text{success}}(n^*, V)$, where $\hat{n} < n^*$. The average number of wasted PUSCH resources is calculated as $V\{1 - N_{\text{success}}(n)/N_{\text{active}}(n)\}$, whose derivative with respect to n is greater than 0. It implies that $V\{1 - N_{\text{success}}(n)/N_{\text{active}}(n)\}$ is an increasing function, and thus $V\{1 - N_{\text{success}}(\hat{n})/N_{\text{active}}(\hat{n})\} < V\{1 - N_{\text{success}}(n^*)/N_{\text{active}}(n^*)\}$, where $\hat{n} < n^*$. ∎

As described above, the RAP value needs to be carefully chosen by considering not only the number of available preambles M but also the number of allocable PUSCH resources V. In short, with the P-DAC 1, the eNodeB adaptively controls the RAP value. In specific, $p(i) = p^*(i)$ when $N_{\text{active}}(n^*) \leq V$, and $p(i) = \hat{p}(i)$ when $N_{\text{active}}(n^*) > V$. For given $p(i)$, each machine node determines its RA attempt on the ith GRA opportunity. At the end of the ith GRA opportunity, the eNodeB calculates the number of backlogged machine nodes for the $(i+1)$th GRA opportunity by using (1). If $N_{\text{backlog}}(i+1)$ is greater than 0, then the eNodeB computes the RAP value $p(i+1)$. If $N_{\text{backlog}}(i+1) = 0$, then all machine nodes in the group become in RRC-connected state, and the uplink data transmission for the group is completed.

B. Proposed DAC Mechanism With e-CD (P-DAC 2)

In this mechanism, we assume that eNodeB is equipped with the e-CD capability. The eNodeB can detect preamble collisions at the first step of the RA procedure, and thus it allocates the PUSCH resources only for the successful (collision-free) preambles. In theory, the PUSCH resources are not wasted at all. Thus, we consider only the average number of successful (collision-free) preambles, $N_{\text{success}}(n)$, for this mechanism. The average number of PUSCH resources assigned to the collision-free preambles, i.e., the average number of RA-success nodes, is given by

$$V_{\text{success}}(n, V) = \begin{cases} N_{\text{success}}(n^*), & \text{if } N_{\text{success}}(n^*) \leq V \\ V, & \text{if } N_{\text{success}}(n^*) > V. \end{cases}$$

When $N_{\text{success}}(n^*) \leq V$ (refer to $V = 30$ in Fig. 5), $N_{\text{success}}(n^*)$ nodes succeed in the RA procedure by allowing only n^* nodes among the backlogged $N_{\text{backlog}}(i)$ nodes to pass the access check with the same RAP value $p^*(i)$ as (7).

On the other hand, when $N_{\text{success}}(n^*) > V$ (refer to $V = 20$ in Fig. 5), we obtain $\tilde{n} = \mathcal{W}(V \cdot \alpha \cdot \ln\alpha)/\ln\alpha$ by solving $V = N_{\text{success}}(n)$, where $\mathcal{W}(\cdot)$ represents the Lambert-W function [33]. In order to allow for $\tilde{n}$ nodes to pass the access check among the backlogged $N_{\text{backlog}}(i)$ nodes, the RAP value $\tilde{p}(i)$ is set to

$$\tilde{p}(i) = \min\left\{\frac{\mathcal{W}(V \cdot \alpha \cdot \ln\alpha)}{N_{\text{backlog}}(i)\ln\alpha}, 1\right\}. \tag{10}$$

In short, with the P-DAC 2, the eNodeB also adaptively controls the RAP value. In specific, $p(i) = p^*(i)$ when $N_{\text{success}}(n^*) \leq V$, and $p(i) = \tilde{p}(i)$ when $N_{\text{success}}(n^*) > V$.

If the RAP value $\tilde{p}(i)$ is used instead of $p^*(i)$, then the average number of RA-success nodes becomes slightly smaller than V because the transmissions occur in a probabilistic manner. To show this, let p denote the RAP value 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 RAP value of p. The average number of successful RA-attempts is obtained as

$$S(p, V) = \sum\nolimits_{n=0}^{N} \min\{N_{\text{success}}(n), V\} \cdot \mathbb{B}_n^N(p) \tag{11}$$

where $S(\tilde{p}, V) < S(p^*, V)$ if $N_{\text{success}}(n) > V$. However, we need to also consider unsuccessful RA-attempts due to preamble collisions which cause the energy waste at machine nodes. The average number of unsuccessful RA-attempts is also obtained as

$$U(p, V) = \sum_{n=0}^{N} n\mathbb{B}_n^N(p) - S(p, V). \tag{12}$$

Hence, the RAP value needs to be carefully controlled between $\tilde{p}$ and p^* by considering both $S(p, V)$ and $U(p, V)$ in practice. We will further discuss a detailed setting for p in the section of numerical results.

The following lemma compares two proposed mechanisms: P-DAC 1 and P-DAC 2.

Lemma 2: When the number of activated preambles N_{active} is greater than that of PUSCH resources V, i.e., $N_{\text{active}} > V$, the average number of RA-success nodes in the P-DAC 2 is larger than or equal to the average number of RA-success nodes in the P-DAC 1.

Proof: Let N_{success} and N_{active} denote the number of successful preambles and the number of activated preambles, respectively, where $N_{\text{success}} \leq N_{\text{active}}$. When $N_{\text{active}} > V$, the P-DAC 1 yields $V \cdot N_{\text{success}}/N_{\text{active}}$ RA-success nodes, while the P-DAC 2 yields $\min\{N_{\text{success}}, V\}$ RA-success nodes.

When $N_{\text{success}} < V$, the difference of the number of RA-success nodes between the P-DAC 2 and the P-DAC 1 is

$$\begin{aligned}&\min\{N_{\text{success}}, V\} - V \cdot N_{\text{success}}/N_{\text{active}} \\ &\quad = N_{\text{success}}(1 - V/N_{\text{active}}) > 0.\end{aligned}$$

Furthermore, when $N_{\text{success}} \geq V$, the difference is

$$\begin{aligned}&\min\{N_{\text{success}}, V\} - V \cdot N_{\text{success}}/N_{\text{active}} \\ &\quad = V(1 - N_{\text{success}}/N_{\text{active}}) \geq 0.\end{aligned}$$

Therefore, the average number of RA-success nodes in the P-DAC 2 is always larger than or equal to the number of RA-success nodes in the P-DAC 1. ■

IV. Performance Analysis

In this section, we investigate four performance metrics to evaluate the proposed DAC mechanisms: GPC time, PUSCH resource efficiency, transmission efficiency, and EE.

A. Group Paging Completion Time

GPC time is defined as the total spent time for all N_{MN} nodes in a group to complete the overall RA procedure. It is calculated by

$$T_{\text{GP}} = \frac{T_{\text{interval}} \cdot N_{\text{MN}}}{\overline{N}_{\text{RA}}} \tag{13}$$

where T_{interval} denotes the time duration of a GRA opportunity and $\overline{N}_{\text{RA}}$ denotes the average number of RA-success nodes in the GRA opportunity. If $N_{\text{backlog}}(i) \geq M$, then $N_{\text{RA}}(i-1) \approx N_{\text{RA}}(i)$. Assuming $N_{\text{MN}} \gg M$, the average number of RA-success nodes in a GRA opportunity for given V can be approximated as

$$\begin{aligned}\overline{N}_{\text{RA}} \approx N_{\text{RA}}(i) = &\sum_{n=0}^{N_{\text{backlog}}(i)} V_{\text{success}}(n, V)\binom{N_{\text{backlog}}(i)}{n} \\ &\times p(i)^n\{1-p(i)\}^{(N_{\text{backlog}}(i)-n)}\end{aligned} \tag{14}$$

where $p(i)$ represents the RAP value in the ith GRA opportunity.

B. PUSCH Resource Efficiency

PUSCH resource efficiency is defined as the ratio of the number of PUSCH resources allocated to the successful (collision-free) preambles to the total number of allocated PUSCH resources. The PUSCH resource efficiency of the C-AC mechanism is given by

$$\eta_{\text{PUSCH}}^{\text{conv}} = \frac{N_{\text{success}}(n^*)}{N_{\text{active}}(n^*)} \tag{15}$$

where $n^* = -1/\ln\alpha$ and $\alpha = (1 - 1/M)$. Note that the PUSCH resource efficiency does not depend on the number of PUSCH resources V in the C-AC mechanism.

For the P-DAC 1, the PUSCH resource efficiency is given by

$$\eta_{\text{PUSCH}}^{\text{DAC1}} = \begin{cases} \frac{N_{\text{success}}(n^*)}{N_{\text{active}}(n^*)}, & \text{if } N_{\text{active}}(n^*) \leq V \\ \frac{N_{\text{success}}(\hat{n})}{N_{\text{active}}(\hat{n})}, & \text{if } N_{\text{active}}(n^*) > V \end{cases} \tag{16}$$

where $\hat{n} = \ln(1 - V/M)/\ln\alpha$. Contrary to the C-AC mechanism, the PUSCH resource utilization of the P-DAC 1 depends on the number of available PUSCH resources V.

For the P-DAC 2, the PUSCH resource efficiency is given by

$$\eta_{\text{PUSCH}}^{\text{DAC2}} = \begin{cases} \frac{N_{\text{success}}(n^*)}{N_{\text{success}}(n^*)} = 1, & \text{if } N_{\text{success}}(n^*) \leq V \\ \frac{N_{\text{success}}(\tilde{n})}{V} = 1, & \text{if } N_{\text{success}}(n^*) > V \end{cases} \tag{17}$$

where $\tilde{n} = \mathcal{W}(V \cdot \alpha \cdot \ln\alpha)/\ln\alpha$ and $N_{\text{success}}(\tilde{n}) = V$. It is worth noting that the PUSCH resource efficiency of the P-DAC 2 is always equal to 1 regardless of the number of available PUSCH resources V due to capability of the e-CD.

C. Transmission Efficiency

TX efficiency is defined as the ratio of the number of successful uplink TXs to the total number of uplink TXs during the overall RA procedure at a machine node. It is given by

$$\eta_{\text{tx}} = \frac{2}{\overline{T}(n)} \tag{18}$$

where 2 in the numerator represents the two uplink transmissions for the successful RA at the first and the third steps and $\overline{T}(n)$ denotes the average of the total number of uplink TXs at the machine node until the successful RA.

In order to compute $\overline{T}(n)$ for the C-AC mechanism and the P-DAC 1, we need to obtain the probabilities of three different events on the preamble collision and the PUSCH resource allocation in a single RA attempt

$$\begin{aligned}P_{\text{event1}}^{\text{DAC1}}(n) &= \Pr\big[\text{PUSCH unallocated}|n\big] \\ &= 1 - \min\{1, V/N_{\text{active}}(n)\} \\ P_{\text{event2}}^{\text{DAC1}}(n) &= \Pr\big[\text{Preamble collision, PUSCH allocated}|n\big] \\ &= \left\{1 - \alpha^{(n-1)}\right\} \times \min\{1, V/N_{\text{active}}(n)\} \\ P_{\text{event3}}^{\text{DAC1}}(n) &= \Pr\big[\text{No preamble collision, PUSCH allocated}|n\big] \\ &= \alpha^{(n-1)} \times \min\{1, V/N_{\text{active}}(n)\}\end{aligned}$$

where $\alpha = (1 - 1/M)$ and recall that n denotes the number of RA-attempting nodes in a certain GRA opportunity. Note that $P_{\text{event3}}^{\text{DAC1}}(n)$ represents the successful RA probability. Based on the above three probabilities, we can obtain the average number of the total uplink transmissions until the successful RA. Considering the first event, the machine node performs a *single* transmission, i.e., $T_{\text{event1}}^{\text{DAC1}} = 1$, at the first step of the RA procedure and it does not transmit data at the third step of the RA procedure for the GRA opportunity because the PUSCH resource is not allocated for the machine node. Considering the second event, the machine node performs *two* transmissions at the first and third steps of the RA procedure, i.e., $T_{\text{event2}}^{\text{DAC1}} = 2$. The data sent at the third step of the RA procedure, however, is not decoded successfully due to multiple machine nodes send data with the same PUSCH resource. As for the third event, the machine node also performs *two* transmissions, i.e., $T_{\text{event3}}^{\text{DAC1}} = 2$, at the first and third steps of the RA procedure, and the RA attempt completes.

On the other hand, the P-DAC 2 adopts the e-CD, and, thus, the above probabilities of three different events are differently given as

$$\begin{aligned}P_{\text{event1}}^{\text{DAC2}}(n) &= \Pr\big[\text{Preamble collision}|n\big] \\ &= 1 - \alpha^{(n-1)}\end{aligned}$$

$$P^{\mathsf{DAC2}}_{\mathsf{event2}}(n) = \Pr[\text{No preamble collision, PUSCH unallocated}|n] = \alpha^{(n-1)} \times (1 - \min\{1, V/N_{\mathsf{success}}(n)\})$$

$$P^{\mathsf{DAC2}}_{\mathsf{event3}}(n) = \Pr[\text{No preamble collision, PUSCH allocated}|n] = \alpha^{(n-1)} \times \min\{1, V/N_{\mathsf{success}}(n)\}$$

where the probability of successful PUSCH allocation is calculated based on $N_{\mathsf{success}}(n)$ and the corresponding numbers of transmissions are given by $T^{\mathsf{DAC2}}_{\mathsf{event1}} = T^{\mathsf{DAC2}}_{\mathsf{event2}} = 1$ and $T^{\mathsf{DAC2}}_{\mathsf{event3}} = 2$, respectively.

Let N_{attempt} denote the total number RA attempts until the RA attempt succeeds. Then, there exist $(N_{\mathsf{attempt}} - 1)$ failed RA attempts before the successful RA attempt which is the N_{attempt}th RA attempt. Since the RA attempt fails due to the first or the second uplink transmissions of the machine node during the RA procedure, total $2^{(N_{\mathsf{attempt}}-1)}$ cases can exist for the successful RA. Let $N_{k,j}$ denote the number of occurrences for the event k in the jth case, where $N_{\mathsf{attempt}} - 1 = \sum_{k=1}^{2} N_{k,j}$, $0 \le N_{k,j} \le N_{\mathsf{attempt}} - 1$, $j = 1, \ldots, 2^{N_{\mathsf{attempt}}-1}$. Thus, the average of total number of uplink transmissions for the P-DAC x (x = 1 or 2) is given by

$$\overline{T}^{\mathsf{DACx}}(n) = \frac{P^{\mathsf{DACx}}_{\mathsf{event3}}(n) T^{\mathsf{DACx}}_{\mathsf{event3}} + \sum_{m=2}^{\infty} \sum_{j=1}^{2^{m-1}} A^{\mathsf{DACx}}_{j} B^{\mathsf{DACx}}_{j}}{P^{\mathsf{DACx}}_{\mathsf{event3}}(n) + \sum_{m=2}^{\infty} \sum_{j=1}^{2^{m-1}} A^{\mathsf{DACx}}_{j}} \tag{19}$$

where $A^{\mathsf{DACx}}_{j} = P^{\mathsf{DACx}}_{\mathsf{event3}}(n) \prod_{k=1}^{2} \{P^{\mathsf{DACx}}_{\mathsf{event}k}(n)\}^{N_{k,j}}$ and $B^{\mathsf{DACx}}_{j} = T^{\mathsf{DACx}}_{\mathsf{event3}} + \sum_{k=1}^{2} T^{\mathsf{DACx}}_{\mathsf{event}k} N^{j}_{k}$.

D. Energy Efficiency

EE is defined as the ratio of the energy consumption for the successful uplink TXs to the energy consumption for the all uplink TXs during the RA procedure at a machine node, which is given by

$$\eta_{\mathsf{ee}} = \frac{E_{\mathsf{S1}} + E_{\mathsf{S3}}}{\overline{E}(n)} \tag{20}$$

where E_{S1} and E_{S3} denote the energy consumption for preamble transmission at the first step and the data transmission at the third step, respectively. Furthermore, $\overline{E}(n)$ denotes the average energy consumption for all uplink TXs until the RA completes.

To calculate the EE, we need to replace the number of TXs in (19) with the corresponding energy consumption for each event. The energy consumption for the three events in the C-AC mechanism and the P-DAC 1 is equal to $E^{\mathsf{DAC1}}_{\mathsf{event1}} = E_{\mathsf{S1}}$, $E^{\mathsf{DAC1}}_{\mathsf{event2}} = E_{\mathsf{S1}} + E_{\mathsf{S3}}$, and $E^{\mathsf{DAC1}}_{\mathsf{event3}} = E_{\mathsf{S1}} + E_{\mathsf{S3}}$, respectively. On the other hand, the energy consumption for three events in the P-DAC 2 is equal to $E^{\mathsf{DAC2}}_{\mathsf{event1}} = E_{\mathsf{S1}}$, $E^{\mathsf{DAC2}}_{\mathsf{event2}} = E_{\mathsf{S1}}$, and $E^{\mathsf{DAC3}}_{\mathsf{event2}} = E_{\mathsf{S1}} + E_{\mathsf{S3}}$, respectively.

Hence, the average energy consumption for the P-DAC x (x = 1 or 2) is given by

$$\overline{E}^{\mathsf{DACx}}(n) = \frac{P^{\mathsf{DACx}}_{\mathsf{event3}}(n) E^{\mathsf{DACx}}_{\mathsf{event3}} + \sum_{m=2}^{\infty} \sum_{j=1}^{2^{m-1}} A^{\mathsf{DACx}}_{j} C^{\mathsf{DACx}}_{j}}{P^{\mathsf{DACx}}_{\mathsf{event3}}(n) + \sum_{m=2}^{\infty} \sum_{j=1}^{2^{m-1}} A^{\mathsf{DACx}}_{j}} \tag{21}$$

where $C^{\mathsf{DACx}}_{j} = E^{\mathsf{DACx}}_{\mathsf{event3}} + \sum_{k=1}^{2} E^{\mathsf{DACx}}_{\mathsf{event}k} N^{j}_{k}$.

TABLE I
SIMULATION PARAMETERS

Parameters	Values
Number of machine nodes in a group (N_{MN})	10000
Number of allocable PUSCH resources for GP (V)	$10 - 50$
Time interval of a GRA opportunity (T_{interval})	50 ms
Number of available preamble resources (M)	64
Energy consumption ratio between RA steps 1 and 3 (β)	$0.5 - 5$

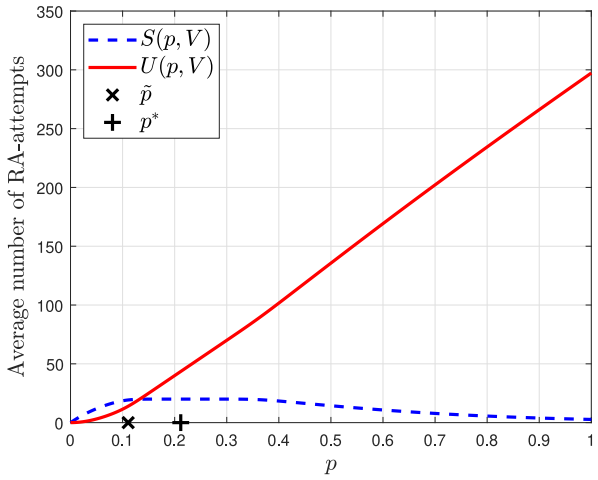


Fig. 6. Average number of successful and unsuccessful RA-attempts of the P-DAC2 for varying RAP value p.

V. NUMERICAL RESULTS

In this section, we evaluate the performance of the proposed DAC mechanisms for the GP-based cellular IoT system in terms of GPC time, PUSCH resource efficiency η_{PUSCH}, TX efficiency η_{tx}, and EE η_{ee}, compared with the C-AC mechanism [34]. The C-AC mechanism considers only the available number of preamble resources when computing the RAP value. In this section, we consider another conventional AC mechanism called the C-AC with e-CD as a baseline technique, which exactly estimate the successful preamble transmissions at the first step of the RA procedure by utilizing the e-CD capability [24]–[26]. However, the C-AC with e-CD still considers only the available number of preamble resources like the C-AC mechanism. Table I lists the simulation parameters. V PUSCH resources are fixedly allocable for the group of nodes on each GRA opportunity.

Fig. 6 shows the number of successful and unsuccessful RA-attempts of the P-DAC2, which are denoted by $S(p, V)$ in (11) and $U(p, V)$ in (12), respectively, when $V = 20$ and the number of backlogged nodes is 300. If $p \le \tilde{p}$, then $S(p, V) > U(p, V)$. However, $S(p, V) < U(p, V)$ as p approaches to p^*. There exists a tradeoff relationship between the GPC time and the EE for varying the RAP value. Hence, we define a novel RAP value as

$$p = \epsilon p^* + (1 - \epsilon)\tilde{p} \tag{22}$$

for $0 \le \epsilon \le 1$. Fig. 7 shows (a) the GPC time and (b) the EE of the P-DAC2 for varying ϵ of the RAP value in (22). When $\epsilon = 1$ and $\epsilon = 0$, the RAP value becomes p^* in (7) and $\tilde{p}$

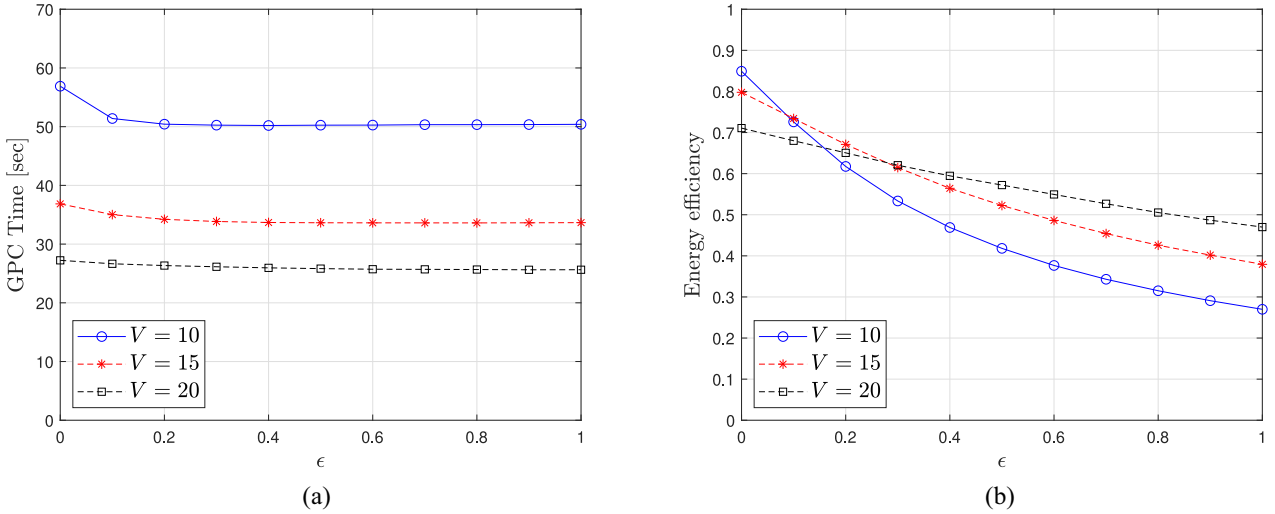


Fig. 7. Performances of the P-DAC2 for varying ϵ. (a) GPC time. (b) EE.

in (10), respectively. As ϵ increases, the GPC time decreases but it is saturated over some ϵ, whereas the EE gradually decreases. Furthermore, for a given V, we optimize ϵ so that the GPC time of the P-DAC2 is minimized and the EE is maximized, which is denoted by $\acute{\epsilon}$. For example, the GPC time is minimized for $\epsilon \in [0.4, 1]$ when $V = 15$. Thus, in this case, $\acute{\epsilon} = 0.4$. Then, for a given V, the optimal RAP value in the sense that the minimum GPC time while the EE is maximized is obtained by

$$\acute{p} = \acute{\epsilon} p^* + \left(1 - \acute{\epsilon}\right)\tilde{p}. \tag{23}$$

In what follows, the P-DAC2 mechanism utilizes the RAP values of $\acute{p}$.

Fig. 8 shows the GPC time of the proposed DAC mechanisms for varying the number of allocable PUSCH resources (V). When V is larger than 40 (i.e., sufficient PUSCH resources are allocable for the GP), three mechanisms C-AC, P-DAC 1, and P-DAC 2 result in a similar GPC time of 21 s. On the other hand, as V decreases, the GPC time of the P-DAC 1 is much shorter than that of the C-AC. Since the RAP value of the C-AC is adjusted by only considering the number of available preamble resources, the number of RA-attempting nodes which pass the access check is not properly controlled when the number of allocable PUSCH resources is insufficient. In contrast, the P-DAC 1 adjusts the RAP value by considering both the number of available preambles and allocable PUSCH resources so that more PUSCH resources are allocated to the collision-free preambles among the activated preambles. With the capability of the e-CD, the P-DAC 2 intelligently adjusts the RAP value by considering the expected number of collision-free preambles, the number of available preambles, and the number of allocable PUSCH resources so that the PUSCH resources are allocated to only the collision-free preambles among the activated preambles. Thus, the P-DAC 2 results in the best GPC time among the mechanisms. Thanks to the e-CD, the C-AC with e-CD exactly allocates the PUSCH resources to only the collision-free preambles, and, thus, both

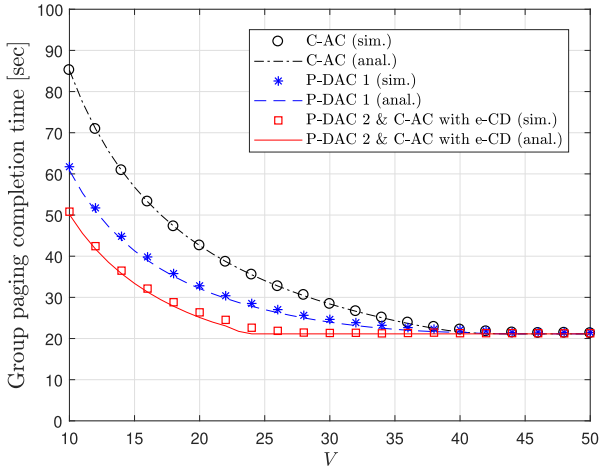


Fig. 8. GPC time of the proposed DAC mechanisms for varying the number of PUSCH resources (V).

the C-AC with e-CD and the P-DAC 2 yield the similar GPC time as shown in Fig. 8.

Fig. 9 shows the PUSCH resource efficiency according to V. Since the C-AC does not take into account V, its PUSCH resource efficiency remains constant regardless of V, which is equal to 0.58 in Fig. 9. Considering that $N_{\mathrm{MN}} = 10\,000$, it implies that 17 241 PUSCH resources are used, but 7241 PUSCH resources among them are wasted. The P-DAC 1 results in a higher PUSCH resource efficiency when $V \leq 40$ than that of the C-AC, but the PUSCH resource efficiency of the P-DAC 1 decreases as V increases. With the P-DAC 1, interestingly, there exist a tradeoff between the GPC time and the number of wasted PUSCH resources. To be specific, more PUSCH resources are wasted in order to reduce the GPC time. For example, when $V = 10$, 812 PUSCH resources are wasted for the GPC time of 61.74 s. When $V = 20$, 2030 PUSCH resources are wasted for the GPC time of 32.14. In addition, the P-DAC 2 and the C-AC with e-CD achieve the best

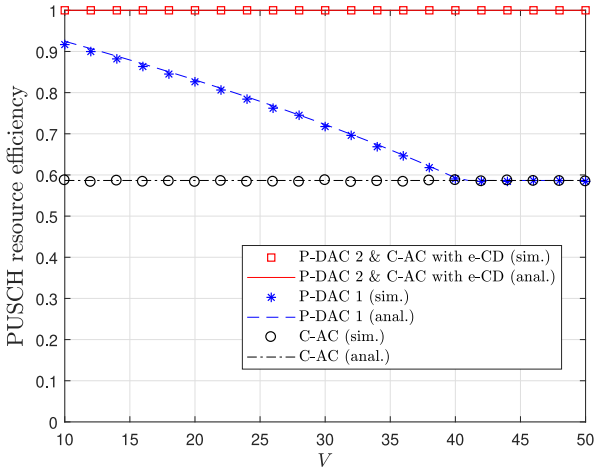


Fig. 9. PUSCH resource efficiency for varying the number of available PUSCH resources (V).

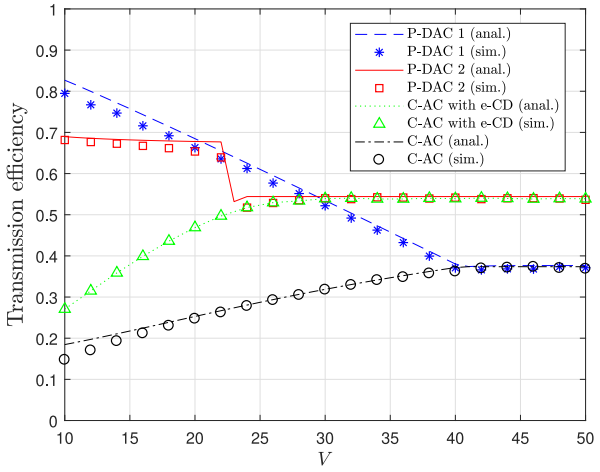


Fig. 10. Transmission efficiency for varying the number of available PUSCH resources (V).

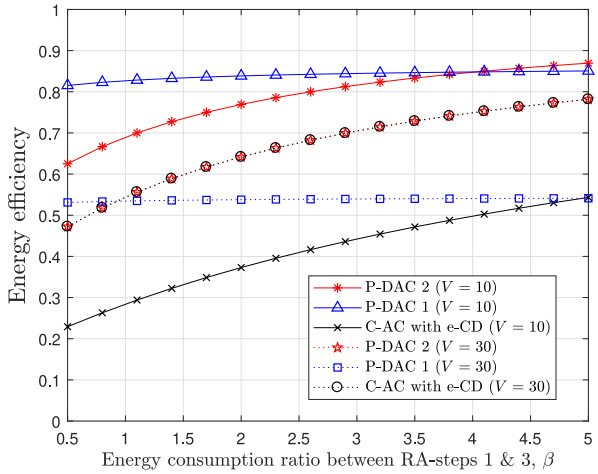


Fig. 11. EE for varying β when $V = 10$ and $V = 30$.

PUSCH resource efficiency performance among all mechanisms by utilizing the capability of e-CD, which is equal to 1 regardless of V.

Fig. 10 shows the transmission efficiency of the proposed DAC mechanisms for varying V. The P-DAC 1 yields a higher transmission efficiency than other mechanisms when the available PUSCH resources are relatively small (i.e., $V < 30$). Similar to results in Fig. 8, the C-AC and the P-DAC 1 result in the identical transmission efficiency (i.e., 0.37) when $V > 40$. The P-DAC 2 achieves a lower transmission efficiency than the P-DAC 1 when $V < 30$, but it outperforms the P-DAC1 and the C-AC and its transmission efficiency remained constant, i.e., 0.54 when $V \geq 30$. In particular, the transmission efficiency of P-DAC2 shows different tendencies in the insufficient resource regime and the sufficient resource regime, respectively, which is separated at $N_{\mathsf{success}}(n^*)$. It is mainly due to a successful resource allocation probability, $\min\{1, V/N_{\mathsf{success}}(n)\}$, in the transmission efficiency. With $M = 64$, the transmission efficiency of the P-DAC2 has a step approximately at $V = 23.7299$ since $N_{\mathsf{success}}(n^*) = 23.7299$, which is the separation point for the resource sufficiency. In short, the proposed mechanisms, the P-DAC 1 and the P-DAC 2, outperform the C-AC mechanisms regardless of e-CD especially for small values of V since the proposed mechanisms consider the limitation of allocable PUSCH resources. It is worth noting that the mathematical analyses on GPC time, PUSCH resource efficiency, and transmission efficiency are well matched with the computer simulation results, which implies the analytical results are correct.

We now evaluate the EE of the proposed DAC mechanisms. Let β denote the ratio of energy consumption between the first and the third steps of the RA procedure, i.e., $E_{\mathsf{S3}} = \beta E_{\mathsf{S1}}$. The ratio β varies according to the number of allocated resources for the third RA procedure [35]. For the clear comparison, we consider two cases in this section: $V = 10$ (insufficient PUSCH resource) and $V = 30$ (sufficient PUSCH resources). Fig. 11 shows the EE as β varies. When $V = 10$, the P-DAC 1 results in the highest EE while the C-AC with e-CD results in the lowest EE over all values of β. In contrast, when $V = 30$, the P-DAC 2 and the C-AC with e-CD achieve the same EE, while the P-DAC 1 results in constant EE for all values of β. In addition, it is shown that the EE of the P-DAC 1 and the P-DAC 2 mechanisms decrease as V increases, while that of C-AC increases.

Table II summarizes the relationship among performance metrics for satisfying the GPC time requirement. To be specific, Table II compares the required number of PUSCH resources (V), PUSCH resource efficiency (η_{PUSCH}), transmission efficiency (η_{tx}), and EE (η_{ee}) of four mechanisms for satisfying the GPC times. We assume that $N_{\mathsf{MN}} = 10\,000$ and $\beta = 1.5$. For fair comparison, we first compare the performance of the P-DAC 1 with those of the C-AC. Basically, more PUSCH resources are required to achieve a shorter GPC time in all mechanisms. When the target GPC time is equal to 20 s, both mechanisms yield the same performance. On the other hand, when the target GPC time is larger than 20 s, the P-DAC 1 outperforms the C-AC in terms of all performance metrics. In particular, a tradeoff exists between the GPC time and other metrics in the P-DAC 1. Then, we compare the performance of the P-DAC 2 with those of the C-AC with e-CD. Both mechanisms require the same

TABLE II
PERFORMANCE COMPARISON OF FOUR MECHANISMS: THE C-AC, THE P-DAC 1, THE C-AC WITH E-CD, AND THE P-DAC 2

GPC time	C-AC				P-DAC 1				C-AC with e-CD				P-DAC 2			
	V	η_{PUSCH}	η_{tx}	η_{ee}	V	η_{PUSCH}	η_{tx}	η_{ee}	V	η_{PUSCH}	η_{tx}	η_{ee}	V	η_{PUSCH}	η_{tx}	η_{ee}
20 seconds	40	$\frac{10000}{17241}$	$\frac{2}{5.40}$	0.37	40	$\frac{10000}{17241}$	$\frac{2}{5.40}$	0.38	24	$\frac{10000}{10000}$	$\frac{2}{3.867}$	0.60	24	$\frac{10000}{10000}$	$\frac{2}{3.68}$	0.60
40 seconds	21	$\frac{10000}{17241}$	$\frac{2}{7.85}$	0.27	16	$\frac{10000}{11582}$	$\frac{2}{2.79}$	0.74	12	$\frac{10000}{10000}$	$\frac{2}{6.333}$	0.37	12	$\frac{10000}{10000}$	$\frac{2}{2.96}$	0.73
60 seconds	14	$\frac{10000}{17241}$	$\frac{2}{10.4}$	0.23	10	$\frac{10000}{10814}$	$\frac{2}{2.51}$	0.83	8	$\frac{10000}{10000}$	$\frac{2}{9.024}$	0.29	8	$\frac{10000}{10000}$	$\frac{2}{2.85}$	0.74

number of PUSCH resources and achieves the same PUSCH resource efficiency. However, the P-DAC 2 achieves much better transmission efficiency and EE. It is worth noting that EE is one of the most important performance metric for IoT devices (machine nodes).

VI. CONCLUSION

In this paper, we proposed two DAC mechanisms which effectively control the number of RA-attempting nodes by adjusting the RAP values, considering both the number of preamble resources and the number of allocable PUSCH resources, while the conventional AC mechanisms consider only the number of preamble resources. We also mathematically analyzed the proposed DAC mechanisms and showed that the analytical results are well matched with the computer simulations. It was shown through extensive computer simulations that the proposed DAC mechanisms, the P-DAC 1 and the P-DAC 2, achieve better performance in terms of GPC time, PUSCH resource efficiency, TX efficiency, and EE, compared with the C-AC mechanisms especially when the number of allocable PUSCH resources is insufficient. Thanks to the capability of e-CD, the P-DAC 2 mechanism can fully utilize the PUSCH resources without wastes, and it significantly improves the EE. The comprehensive results of the proposed DAC mechanisms give network operators insights on how to satisfy the performance requirements of various cellular IoT/M2M applications.

REFERENCES

[1] Technology & Communications Research Team, *The Internet of Things: A Study in Hype, Reality, Disruption, and Growth*, Raymond James U.S. Res., St. Petersburg, FL, USA, Jan. 2014.

[2] 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.

[3] E. Dahlman, S. Parkvall, and J. Skold, *4G: LTE/LTE-Advanced for Mobile Broadband*. New York, NY, USA: Academic, 2013.

[4] J. Chen *et al.*, "Narrowband Internet of Things: Implementations and applications," *IEEE Internet Things J.*, vol. 4, no. 6, pp. 2309–2314, Dec. 2017.

[5] Y. Sun, F. Tong, Z. Zhang, and S. He, "Throughput modeling and analysis of random access in narrowband Internet of Things," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 1485–1493, Jun. 2018.

[6] M. Centenaro, L. Vangelista, S. Saur, A. Weber, and V. Braun, "Comparison of collision-free and contention-based radio access protocols for the Internet of Things," *IEEE Trans. Commun.*, vol. 65, no. 9, pp. 3832–3846, Sep. 2017.

[7] *Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation, Release 8, V8.1.0*, 3GPP, Sophia Antipolis, France, Rep. TR 36.211, Nov. 2007.

[8] M. Hasan, E. Hossain, and D. Niyato, "Random access for machine-to-machine communication in LTE-Advanced networks: Issues and approaches," *IEEE Commun. Mag.*, vol. 51, no. 6, pp. 86–93, Jun. 2013.

[9] M.-Y. Cheng, G.-Y. Lin, H.-Y. Wei, and A. C.-C. Hsu, "Overload control for machine-type-communications in LTE-Advanced system," *IEEE Commun. Mag.*, vol. 50, no. 6, pp. 38–45, Jun. 2012.

[10] G.-Y. Lin and H.-Y. Wei, "Auction-based random access load control for time-dependent machine-to-machine communications," *IEEE Internet Things J.*, vol. 3, no. 5, pp. 658–672, Oct. 2016.

[11] 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, 1st Quart., 2014.

[12] M.-J. Shih, G.-Y. Lin, and H.-Y. Wei, "Two paradigms in cellular Internet-of-Things access for energy-harvesting machine-to-machine devices: Push-based versus pull-based," *IET Wireless Sensor Syst.*, vol. 6, no. 4, pp. 121–129, Aug. 2016.

[13] *Study on RAN Improvements for Machine-Type Communications, V11.0.0*, 3GPP, Sophia Antipolis, France, Rep. TR 37.868, Oct. 2011.

[14] 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.

[15] H. He *et al.*, "Traffic-aware ACB scheme for massive access in machine-to-machine networks," in *Proc. IEEE ICC*, Jun. 2015, pp. 617–622.

[16] J.-P. Cheng, C.-H. Lee, and T.-M. Lin, "Prioritized random access with dynamic access barring for RAN overload in 3GPP LTE-A networks," in *Proc. IEEE GLOBECOM Workshops*, Dec. 2011, pp. 368–372.

[17] T.-M. Lin, C.-H. Lee, J.-P. Cheng, and W.-T. Chen, "PRADA: Prioritized random access with dynamic access barring for MTC in 3GPP LTE-A networks," *IEEE Trans. Veh. Technol.*, vol. 63, no. 5, pp. 2467–2472, Jun. 2014.

[18] 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.

[19] S. Duan, V. Shah-Mansouri, Z. Wang, and V. W. S. 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.

[20] H. Jin, W. T. 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.

[21] *Pull Based RAN Overload Control*, document TSG-RAN WG2 Meeting #71, 3GPP, Sophia Antipolis, France, Aug. 2010.

[22] *Group Paging for MTC Devices*, document TSG-RAN2 WG2 Meeting #70, 3GPP, Sophia Antipolis, France, Jun. 2010.

[23] *Further Analysis of Group Paging for MTC*, document TSG-RAN2 WG2 Meeting #74, 3GPP, Sophia Antipolis, France, May 2011.

[24] C.-H. Wei, R.-G. Cheng, and S.-L. Tsao, "Performance analysis of group paging for machine-type communications in LTE networks," *IEEE Trans. Veh. Technol.*, vol. 62, no. 7, pp. 3371–3382, Sep. 2013.

[25] O. Arouk, A. Ksentini, and T. Taleb, "Group paging-based energy saving for massive MTC accesses in LTE and beyond networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 5, pp. 1086–1102, May 2016.

[26] R.-G. Cheng, F. M. Al-Taee, J. Chen, and C.-H. Wei, "A dynamic resource allocation scheme for group paging in LTE-advanced networks," *IEEE Internet Things J.*, vol. 2, no. 5, pp. 427–434, Oct. 2015.

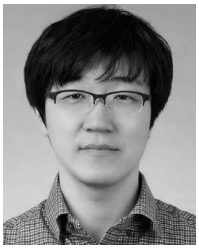
[27] H. S. Jang, H.-S. Park, and D. K. Sung, "A non-orthogonal resource allocation scheme in spatial group based random access for cellular M2M communications," *IEEE Trans. Veh. Technol.*, vol. 66, no. 5, pp. 4496–4500, May 2017.

[28] T. P. de Andrade, C. A. Astudillo, and N. L. S. 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.

[29] H. S. Jang, S. M. Kim, H.-S. Park, and D. K. Sung, "An early preamble collision detection scheme based on tagged preambles for cellular M2M random access," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 5974–5984, Jul. 2017.
[30] D. T. Wiriaatmadja and K. W. Choi, "Hybrid random access and data transmission protocol for machine-to-machine communications in cellular networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, pp. 33–46, Jan. 2015.
[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] 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.
[33] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, "On the Lambertw function," *Adv. Comput. Math.*, vol. 5, no. 1, pp. 329–359, Dec. 1996.
[34] S. Sesia, I. Toufik, and M. Baker, *LTE—The UMTS Long Term Evolution From Theory to Practice*. Chichester, U.K.: Wiley, 2009.
[35] *LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Layer Procedures, Release 12, V12.4.0*, 3GPP Standard TS 36.213, Dec. 2015.

Han Seung Jang (S'14–M'18) received the B.S. degree in electronics and computer engineering from Chonnam National University, Gwangju, South Korea, in 2012, and the M.S. and Ph.D. degrees in electrical engineering from the Korea Advanced Institute for Science and Technology, Daejoen, South Korea, in 2014 and 2017, respectively.

Since 2018, he has been a Post-Doctoral Fellow with the Information Systems Technology and Design Pillar, Singapore University of Technology and Design, Singapore. He was also a Post-Doctoral Fellow with Chungnam National University, Daejeon, from 2017 to 2018. His current research interests include cellular Internet-of-Things/machine-to-machine communications, machine learning, smart grid, and energy ICT.



Bang Chul Jung (S'02–M'08–SM'14) received the B.S. degree in electronics engineering from Ajou University, Suwon, South Korea, in 2002, and the M.S. and Ph.D. degrees in electrical and computer engineering from the Korea Advanced Institute for Science and Technology (KAIST), Daejeon, South Korea, in 2004 and 2008, respectively.

He was a Senior Researcher/Research Professor with the KAIST Institute for Information Technology Convergence, Daejeon, from 2009 to 2010. From 2010 to 2015, he was a Faculty Member of Gyeongsang National University, Jinju, South Korea. He is currently an Associate Professor with the Department of Electronics Engineering, Chungnam National University, Daejeon. His current research interests include wireless communication systems, Internet-of-Things communications, statistical signal processing, information theory, interference management, random access, radio resource management, cooperative relaying techniques, in-network computation, and mobile computing.

Dr. Jung was a recipient of the Fifth IEEE Communication Society Asia–Pacific Outstanding Young Researcher Award in 2011, the Bronze Prize of Intel Student Paper Contest in 2005, the First Prize of KAIST's Invention Idea Contest in 2008, the Bronze Prize of Samsung Humantech Paper Contest in 2009, the Best Paper Award in Spring Conference of Korea Institute of Information and Communication Engineering in 2015, the Best Paper Award in the Institute of Electronics and Information Engineering Symposium in 2017, and the Best Paper Award in Winter Conference of Korea Institute of Communication and Information Science (KICS) in 2018. He has been selected as a winner of Haedong Young Scholar Award in 2015, which is sponsored by the Haedong foundation and given by KICS. He has been serving as an Associate Editor for the *IEICE Transactions on Fundamentals of Electronics, Communications, and Computer Sciences* since 2018.

Dan Keun Sung (S'80–M'86–SM'00–F'15–LF'18) received the B.S. degree in electronics engineering from Seoul National University, Seoul, South Korea, in 1975 and the M.S. and Ph.D. degrees in electrical and computer engineering from the University of Texas at Austin, Austin, TX, USA, in 1982 and 1986, respectively.

Since 1986, he has been with the faculty of the Korea Advanced Institute of Science and Technology (KAIST), Daejon, South Korea, where he is currently a Professor Emeritus with the School of Electrical Engineering. From 1996 to 1999, he was the Director of the Satellite Technology Research Center, KAIST. His current research interests include mobile communication systems and networks with special interest in resource management, cellular M2M, smart grid, energy networks, energy ICT, WLANs, WPANs, traffic control in wireless and wired networks, performance and reliability of communication systems, and microsatellites.

Dr. Sung was a recipient of the 1992 National Order of Merits, the Dongbaek, the 1997 Research Achievement Award, the 2000 Academic Excellence Award, the 2004 Scientist of the Month from the Ministry of Science and Technology and the Korea Science and Engineering Foundation, the 2013 Haedong Academic Grand Award from the Korean Institute of Communications and Information Sciences, and the 2017 National Order of Merits, the Okjo Medal. He had served as the Division Editor of the *Journal of Communications and Networks* from 1998 to 2007. He also had served as an Editor for *IEEE Communications Magazine* from 2002 to 2011. He is a fellow of the Korean Academy of Science Technology and a Life Fellow of the National Academy of Engineering of Korea.