IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received March 11, 2019, accepted April 1, 2019, date of publication April 15, 2019, date of current version April 25, 2019.
Digital Object Identifier 10.1109/ACCESS.2019.2911089

Reliability Control Framework for Random Access of Massive IoT Devices

SO YI JUNG[1], (Student Member, IEEE), SUNG HYUNG LEE[2], AND JAE-HYUN KIM[1], (Member, IEEE)
[1]Department of Electrical and Computer Engineering, Ajou University, Suwon 16499, South Korea
[2]SOLVIT System Co., Ltd., Seoul 05854, South Korea

Corresponding author: Jae-Hyun Kim (jkim@ajou.ac.kr)

This work was supported in part by the National Research Foundation of Korea (NRF) through the Korea Government, Ministry of Science and ICT (MSIT), under Grant NRF-2017R1A2A2A05001404, and in part by the MSIT, South Korea, through the Information Technology Research Center (ITRC) support program under Grant IITP-2019-2017-0-01637 supervised by the Institute for Information and Communications Technology Promotion (IITP).

ABSTRACT The Internet of Things (IoT) devices with enhanced machine type communication (eMTC) technology require random access (RA) to transmit data. The success rate of data delivery in the eMTC depends on the probability of failure in the RA. Access class barring (ACB) can decrease the probability of failure in the RA procedure. However, it is hard to precisely predict the success rate of the RA with the ACB. In this paper, we aim to control the failure rate of the RA to the desired probability by designing a reliability control framework for the RA in the eMTC. The framework includes an algorithm that estimates the number of active devices in a cell from the number of undecoded preambles and the probability of preamble loss. The framework also consists of an algorithm to obtain the probability of RA failure by changing the number of preamble transmissions. In addition, the adaptive ACB factor decision algorithm adjusts the ACB factor based on observed state information. The performance of the proposed framework is evaluated using an RA simulator in the environments recommended by the third generation partnership project (3GPP). The results of the evaluation indicate that the number of preamble transmissions selected by our algorithm successfully determines the probability of RA failure. In addition, the simulation results suggest that the number of supportable devices decreases as a tradeoff for increased reliability due to decreased RA failure rate.

INDEX TERMS 3GPP, access class barring, congestion control, random access, reliability.

I. INTRODUCTION

The Internet of things (IoT) is considered to be a big opportunity for the Internet research community. IoT services can work with an enormous number of devices with communications capabilities [1]. Gartner expects 20.4 billion connected things by 2020, and IHS Markit announced that IoT devices worldwide will increase by 12 percent on average annually from 27 billion in 2017 to 125 billion in 2030. These expectations show that IoT is an enormous opportunity for cellular network service providers since numerous devices can be a new source of income.

Enhanced machine type communication (eMTC), which is sometimes referred to as long-term evolution-M (LTE-M) or category M1, is a new type of low-power wide area (LPWA) technology standard for cellular network service providers. eMTC is intended for data collection such as smart traffic management, smart logistics, and environmental monitoring. eMTC, which operates at a low speed using the 1.4 MHz bandwidth [2], includes improved power consumption reduction technology [3], and can provide a simple signaling procedure [4]. Therefore, eMTC can provide cost efficiency and low power consumption, which are required for IoT.

Rapid growth in the number of devices is a cause of network congestion [5]–[8]. Congestion degrades the performance of radio access networks (RANs) that provide the connection between devices and gateways of IoT systems [9]. IoT devices with eMTC technology require random access (RA) to acquire the opportunity to transmit data from/to the RAN [3], [10]. For a successful RA, an IoT device in the network requires successful transmissions of a preamble and a radio

The associate editor coordinating the review of this manuscript and approving it for publication was Zheng Chang.

2169-3536 © 2019 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

resource control (RRC) layer message [11]. The device can experience collisions of RRC layer messages when multiple devices select the same preamble. IoT devices will give up the transmission of data when the device experiences multiple failures of RA. Therefore, the reliability of data delivery between IoT devices and a remote server is dependent on the rate of RA collision.

The probability of a collision in RRC layer message is proportional to the number of devices that transmit their preamble in a given RA channel (RACH) [12]. Thus, the base station (BS) in a 3GPP network, such as evolved node b (eNB), can limit the number of contending devices per RACH. Based on this concept, 3GPP includes access class barring (ACB) in their standard [11], [13]. For the ACB, the BS decides and announces the probability to enter a contention in RACH, where the probability is also referred to as the ACB factor. The device obtains the announced ACB factor and randomly decides whether it participates in the contention in the upcoming RACH or defers the preamble transmission [14].

Previous studies for ACB have focused on the throughput, which is the number of devices that have experienced a successful RA per unit time [12], [15], [16]. Researchers have tried to find a good method to estimate the number of contending devices and the number of active devices, and have also tried to determine the ACB factor that maximizes throughput. However, throughput maximization does not ensure reliability. IoT users may not require very high level of reliability, but will still require a certain level of reliability. If we cannot ensure a certain level of reliability to these users, they will not want to use eMTC for IoT services. Therefore, we need a method to ensure a certain RA success rate by analyzing the RA procedure and ACB.

This study intends to provide a certain level of reliability for IoT services. The following are the contributions of this paper:

1) We propose a framework to provide a certain level of reliability in eMTC for application users.
2) The framework includes a model to estimate the state information from the number of undecoded preambles since the BS can count undecoded preambles in the eMTC without delay.
3) The framework also includes a search algorithm to select the maximum number of preamble transmissions in the RA procedure to achieve a desired probability of RA failure. The search algorithm is based on numerical analysis, and is corrected using the RA simulator to compensate for unavoidable error from the estimation and random changes in the contention.
4) The framework includes an adaptive estimation, and an ACB factor decision algorithm is also presented to configure the ACB related parameters adaptively from the observed state information.
5) We evaluate the framework by using an RA simulator. The evaluation includes validation for the estimation algorithm and an observation for reliability. In addition, we also evaluate the trade-offs to achieve a certain level of reliability from the simulation, such as the access delay, the number of preamble transmissions for a successful RA, and the maximum number of devices to provide a certain level of reliability.

The remainder of this paper is organized as follows: Section II introduces the background and related work for eMTC. In Section III, we provide the system model for this paper. Section IV introduces conventional throughput maximization in ACB. We propose our framework in Section V. We evaluate the proposed algorithm and perform a correction in Section VI. Finally, the conclusion is made in Section VII.

II. BACKGROUNDS

A. RANDOM ACCESS PROCEDURE IN EMTC

In the mobile network, an IoT device performs a RA procedure for connection through a BS to other network domains [3]. If a device has to transmit or receive data, the device requires an RA procedure in eMTC to establish or recover a connection between the core network and device.

The RA procedure mainly consists of a four-message handshake between the user equipment (UE) and the eNB. The device starts an RA procedure by selecting a preamble from the preamble pool and by transmitting the preamble in an upcoming RACH. For the RA procedure of the devices, the BS periodically allocates RACHs in the uplink band and announces the position using a system information block-2 (SIB2) message. If the BS decoded the transmitted preamble in RACH, it transmits a random access response (RAR, it is also referred as MSG2) message to devices that transmitted the decoded preamble. If the device receives the MSG2 corresponding to the transmitted preamble, the device sends a third message (MSG3) to the BS, where MSG3 is generally a control message from the radio resource control (RRC) layer in the device. If MSG3 is decoded in BS, the BS responds by transmitting a contention resolution message (MSG4). The procedure from the transmission of the preamble to the transmission of MSG4 is referred to as an RA procedure. The end of the reception of MSG4 means the success of an RA procedure for a device. Fig. 1 shows the RA procedure for the IoT devices using eMTC technology [3].

If the device cannot receive MSG2, which means that the transmitted preamble is not decoded in BS, or if it cannot receive MSG4, which means that the preamble is transmitted by multiple devices, the device goes into backoff. The backoff means waiting for a random time. After the end of the backoff, the device restarts the preamble selection. If the device experiences a number of preamble transmissions (q in Fig. 1) compared to the threshold ($Q_{\max}$ in Fig. 2), the device regards this as a failure of RA [3].

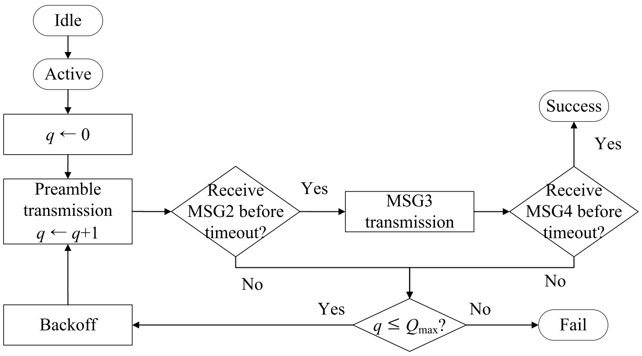


FIGURE 1. Random access procedure in eMTC.

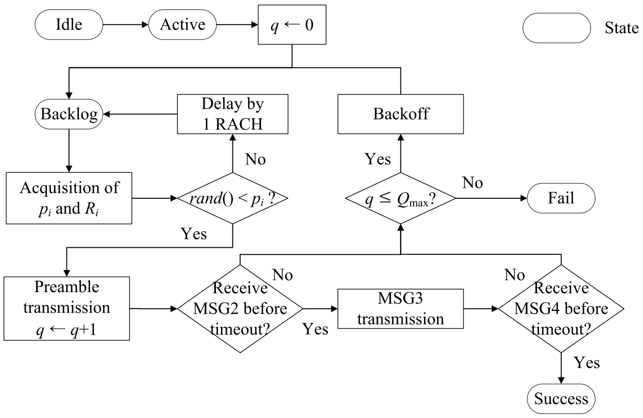


FIGURE 2. ACB and RA procedure model for devices.

B. ACCESS CLASS BARRING

3GPP recommends several solutions to overcome a traffic overload from IoT devices in RAN [13], [17]. Most of the research regarding congestion control in LTE-A focuses on ACB, especially for the UE individual ACB scaling, due to the limit on the number of preambles and resource blocks. Using ACB, an IoT device postpones its request with some probability. It effectively reduces the collision rate to control traffic overload when the resources for the random access is limited. The ACB spreads the device accesses through time by randomly delaying the beginning of the device access attempts according to a barring rate and a barring time. The objective of ACB is to reduce the number of simultaneous devices contending for access.

In ACB, the number of contending devices per RACH is controlled by announcing an ACB factor, where the ACB factor is a value that represents a probability. The ACB factor is a probability to start an RA procedure for a device that is determined by the BS. When an IoT device tries to initiate a transmission, it generates a random number between 0 and 1, and compares the generated number with the ACB factor broadcast by eNB.

Let p be the ACB factor and $rand()$ be a function that generates a real value in [0, 1]. The devices can transmit their preamble with probability p, or defer its preamble transmission for one RA period by probability $1 - p$. For the ACB, the device conducts an ACB trial, where the ACB trial is the comparison of p and $rand()$. If $rand()$ generates a value larger than the ACB factor, the device regards this as a failure of the ACB trial, waits for a certain amount of time, and then retries the ACB trial. Otherwise, the device regards this as a success of the ACB trial and starts the RA procedure. The ACB factor can be announced and obtained by the SIB2 message.

C. RELATED WORKS

The RACH procedure is similar to the frame slotted ALOHA or the multi-channel ALOHA, where the BS can estimate the number of MTC devices that send preambles in an RA slot for congestion control [18]. Liva [19] proposed irregular repetition slotted ALOHA which represents an improvement of the contention resolution diversity of slotted ALOHA. Purwita and Anwar [20] considered massive uncoordinated multiway relay networks applying coded RA for flexible topology changes.

The ACB was studied to maximize throughput or reduce delay rather than targeting the probability of failure. Duan *et al.* [21] presented an optimal ACB factor as the ratio of the number of preambles to the number of activated devices in the network. In addition, a heuristic algorithm was given to update the ACB factor in the BS where this algorithm assumes that the BS knows the number of users in the cell. In a later study by Duan *et al.* [12], they proposed a dynamic ACB (D-ACB) for both a fixed and adaptive number of preambles that includes an estimation using the number of successful and collided preambles. D-ACB aims to minimize the RA delay while maximizing the throughput.

He *et al.* [15] indicated that the number of activated devices can be estimated by the ratio of the number of contended devices in a RACH to the ACB factor used in the RACH. In addition, they proposed a dynamic ACB factor control algorithm based on the traffic arrival model in [13]. Tavana *et al.* [16] derived the ACB factor by predicting the number of activations in the next RACH. The prediction-based algorithm is improved by applying a Kalman filter. Wang and Wong [22] formulated an optimization problem to determine a closed-form optimal solution for the ACB factor, which maximized the expected number of MTC devices successfully served in each random access slot.

Moon and Lim [23] proposed an adaptive ACB and assumed that the distribution of the device arrival is known. Koseoglu [24] proposed an adaptive ACB scheme as the pricing based load control given the arrival distribution is known. Jin *et al.* [25] proposed a recursive pseudo-Bayesian ACB. Leyva-Mayorga *et al.* [26] proposed an ACB factor decision using the least-mean square algorithm and traversal filter to reduce the access delay.

Kalalas *et al.* [27] performed an evaluation of the reliability of the ACB for a fixed ACB factor over time. This evaluation indicates that the ACB can improve the reliability of the RA, and the level of reliability increases as the ACB factor decreases. However, Kalalas and others did not recommend parameters for ACB to achieve a certain level of reliability.

To the best of our knowledge, adaptive ACB schemes generally aim to provide a high throughput by focusing on a decision for an ACB factor. Since the eMTC assumes a limited number of preamble transmissions [3], [13], some devices can experience a failure in the RA, which can indicate a failure in the data transmission. The probability of failure can be critical to some application areas since the application users generally want a certain level of failure probability. Therefore, we need to revise the relation between the ACB and the RA failure probability to provide more reliable IoT services for application users.

III. SYSTEM MODEL

Consider a cell with a BS and M IoT devices in the coverage of the BS. The BS is connected to a server through a backbone network. To collect data from the M devices, the server requests a report from the M devices at a $t = 0$ subframe. Let T_A be the maximum activation time. A device activates at t-th subframe where $t \in [0, T_A]$ for the uplink data transmission. We assume that t is determined by a uniform distribution. Let I_A be the RACH interval. Let i be the index of RACH where i is a positive integer. The i-th RACH is allocated in the iI_A-th subframe. If a device is activated in a t_a-th subframe where $t_a \in [(i-1)I_A, iI_A - 1]$, its first RACH becomes the i-th RACH.

The BS can change the number of preambles per RACH ("pool size" in this paper) and the probability to enter contention ("ACB factor" in this paper). Let R_i be the pool size and p_i be the ACB factor for i-th RACH, respectively. Let $R_{\max}$ be the maximum pool size, i.e. $R_i \leq R_{\max}$. The BS announces R_i and p_i before the start of the i-th RACH.

An activated device conducts the acquisition, which means there is a reception of R_i and p_i. After the acquisition, the device conducts an ACB trial and generates a random real value in [0, 1] using a random number generation function, $rand()$, before the transmission of the preamble for the i-th RACH. If $rand() > p_i$, the device defers to the $(i+1)$-th RACH and returns to acquisition. Otherwise, the device starts the RA procedure.

The RA procedure starts with the selection of a preamble from the R_i preambles. The selected preamble is transmitted to the BS through a RACH. If the device tries to send the q-th preamble transmission, the probability of preamble transmission failure is $1/e^q$ in a single cell environment. The preamble decoding probability in the BS is as follow:

$$P_q = 1 - \frac{1}{e^q}, \tag{1}$$

where q is the number of preamble transmissions in a device [11]. If a preamble is decoded in the BS, the BS transmits MSG2 after T_{RAR} subframes for the preamble where T_{RAR} is the waiting time to start the RAR transmission. If a device receives MSG2 corresponding to its transmitted preamble, it transmits its MSG3 to the BS. The BS transmits MSG4 to the device as the response of MSG3, and the RA of the device is successfully completed.

If the transmitted preamble is not decoded, the device does backoff after $T_{RAR} + W_{RAR}$ subframes where W_{RAR} is the RAR window. In addition, the device also does a backoff when it experiences the collision. The collision is defined as the transmission of a preamble by multiple devices in a RACH [13]. We assume that the devices that transmitted the same preamble in a RACH can receive MSG2. These devices respond by transmitting MSG3 after T_{PROC} where T_{PROC} is the process time. However, we also assume that their MSG3s are not decoded in the BS since different messages are transmitted using the same time-frequency resource. We assume that these devices wait T_{MSG4} subframes from the time of MSG3 transmission and do a backoff [13].

For the backoff, the device selects a random integer in $[0, W_{BO}]$ where W_{BO} is the backoff indicator, and it waits for multiple subframes where the waiting time is equal to the random integer that was selected. Let $Q_{\max}$ be the maximum number of the preamble transmissions. If q is larger than $Q_{\max}$ before the backoff, the device regards this as a failure in the RA. Otherwise, we adopt the deferred first transmission (DFT) model: If q is equal or lower than $Q_{\max}$, the device return to the acquisition of p_i and R_i [12], [26]. Fig. 2 shows the flow chart for devices that include the ACB and RA procedure in our system model.

Let P_T be the desired access failure probability (AFP) for the application users. We assume that the BS knows P_T and can change $Q_{\max}$ or p_i to achieve AFP equal to or lower than P_T.

IV. THROUGHPUT MAXIMIZATION WITH FULL STATE INFORMATION

Let M_i, N_i, and S_i be the number of devices that are ready to conduct an ACB trial, the number of devices which pass the ACB trial, and the number of preambles selected by only one device in the i-th RACH, respectively. Suppose that the preamble is always decodable in the BS. For a given preamble k, let $D_k = 0, 1, \ldots$ be the case where the k-th preamble is chosen by none of the devices, by one device, and so on. Let $P(D_k = 1|N_i = n)$ be the probability that only one device selects the k-th preamble given that $N_i = n$. We have

$$P(D_k = 1|N_i = n) = \binom{n}{1} \frac{1}{R_i} \left(1 - \frac{1}{R_i}\right)^{n-1}. \tag{2}$$

Let $E[S_i|N_i = n]$ be the statistical expectation of the number of successful preamble transmissions in the i-th RACH given that $N_i = n$. We have

$$\begin{aligned} E[S_i|N_i = n] &= \sum_{k=1}^{R_i} P(D_k = 1|N_i = n) \\ &= R_i \binom{n}{1} \frac{1}{R_i} \left(1 - \frac{1}{R_i}\right)^{n-1}. \end{aligned} \tag{3}$$

Since $\frac{d}{dR_i} E[S_i|N_i = n] > 0$, we can conclude that R_i should be maximized to increase S_i. Let R^* be the optimal pool size.

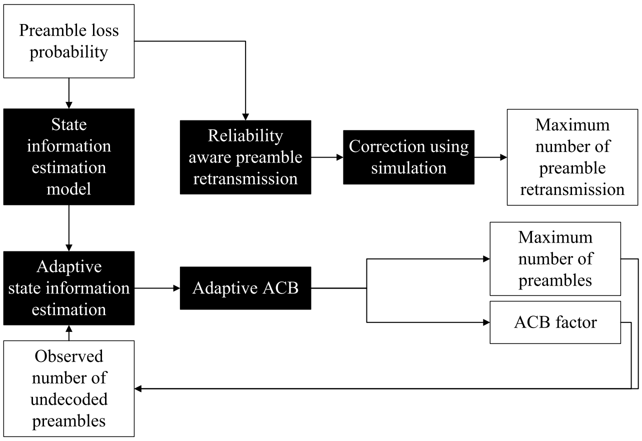


FIGURE 3. Reliability aware parameter configuration.

We have

$$R^* = R_{\max}. \tag{4}$$

Let $E[S_i|M_i = m]$ be the statistical expectation of S_i given that $M_i = m$. We have

$$\begin{aligned} E[S_i|M_i = m] &= \sum_{j=0}^{m} E[S_i|N_i = j]P[N_i = j|M_i = m] \\ &= \sum_{j=1}^{m} \binom{m}{j} p_i^j (1-p_i)^{m-j} j \left(1 - \frac{1}{R_i}\right)^{j-1} \\ &= m p_i \left(1 - \frac{p_i}{R_i}\right)^{m-1}. \end{aligned} \tag{5}$$

Let p^* be the optimal ACB factor. When $R_i \geq m$, $\frac{d}{dR_i}E[S_i|M_i = m] \geq 0$ which means that $p^* = 1$. When $R_i < m$, we can find the optimal ACB factor by setting $\frac{d}{dR_i}E[S_i|M_i = m] = 0$. In this case, $p^* = \frac{R_i}{m}$. Combined with the optimal pool size, we can obtain the throughput optimal ACB factor as follows:

$$p^* = \min\left(1, \frac{R_{\max}}{m}\right). \tag{6}$$

V. PROPOSED RELIABILITY CONTROL FRAMEWORK

The throughput maximization with full state information cannot be used directly in a real system due to a problem: M_i and N_i are unknown in the BS. In addition, the throughput maximization does not ensure a certain level of AFP in RA. In this section, we propose a framework to control the reliability of the RA in eMTC. This framework is summarized in Fig. 3 where the white boxes indicate the input and/or output, and the black boxes indicate the algorithms. We first present the state information estimation model based on the observed number of undecoded preambles and the preamble loss probability model in [13]. We then propose the reliability aware decision algorithm that determines the maximum number of preamble retransmissions from the numerical analysis. Since a perfect state information is not available in the BS, an error in the state information is inevitable but cannot be derived from numerical analysis. Thus, we use the simulation to correct the decision algorithm. We also present the adaptive state information estimation algorithm and the adaptive ACB algorithm based on the estimated state information.

A. STATE INFORMATION ESTIMATION MODEL

In this subsection, we present an estimation algorithm to estimate M_i and N_i from the number of undecoded preambles. The BS has difficulty in obtaining S_i or the number of collided preambles at the end of a RACH since the decoding of MSG3 is required to count S_i or the collided preambles in the eMTC. Fortunately, a number of undecoded preambles is available in the BS at the end of a RACH, thus the estimation method from the undecoded preambles can be useful for the BS.

Let U_i be the number of undecoded preambles in the i-th RACH. Let $E[U_i|N_i = n]$ be the statistical expectation of U_i given that $N_i = n$. If all transmitted preambles from N_i devices are decoded, we have

$$\begin{aligned} E[U_i|N_i = n] &= \sum_{k=1}^{R_i} \binom{n}{0} \left(\frac{1}{R_i}\right)^0 \left(1 - \frac{1}{R_i}\right)^n \\ &= R_i \left(1 - \frac{1}{R_i}\right)^n, \end{aligned} \tag{7}$$

where $R_i = R_{\max}$ from the throughput maximization. Let $f(U_i)$ be a function of U_i where

$$f(U_i) = \frac{\log(U_i/R_i)}{\log(1 - 1/R_i)}. \tag{8}$$

In (8), $f(U_i)$ is the function obtained by replacing $E[U_i|N_i = n]$ to U_i and n to $f(U_i)$ with a rearrangement of (7). For $U_i > 0$, the BS can use $f(U_i)$ as an estimation function to estimate N_i. Let $\hat{N}_i$ be the estimation of N_i. We have

$$\hat{N}_i = f(U_i) = \frac{\log(U_i/R_i)}{\log(1 - 1/R_i)}; \quad U_i > 0. \tag{9}$$

However, when $U_i = 0$, $f(U_i) = \infty$. To avoid this case, which is impossible in a real system, [28] recommends $\hat{N}_i = S_i + 2C_i$ where C_i is the number of collided preambles. Unfortunately, the BS cannot obtain S_i and C_i immediately after the completion of RACH in eMTC. Instead, we can expect $\hat{N}_i > f(1)$ for $U_i = 0$ and $\hat{N}_i - f(1) \geq f(1) - f(2)$. Therefore, we select $\hat{N}_i$ for $U_i = 0$ as follows:

$$\hat{N}_i = f(1) + \{f(1) - f(2)\} = 2f(1) - f(2); \quad U_i = 0. \tag{10}$$

The finalized form of the estimation function without considering the preamble loss probability can be summarized as follows:

$$\hat{N}_i = \begin{cases} f(U_i); & U_i > 0, \\ 2f(1) - f(2); & U_i = 0. \end{cases} \tag{11}$$

As in the simulation model presented by 3GPP [13], the preamble can be lost in the channel, which results in an error between the actual U_i and that in the BS. We need to compensate for the error from the preamble loss probability to produce a better estimation algorithm. The preamble loss probability is dependent on the number of preamble transmissions, but the BS cannot acquire the number of devices for each q in the practical system. We'll use the expectation instead of the probability corresponding to q. Let σ be a preamble loss correction factor to compensate for the error between the actual number of undecoded preambles and the observed number of undecoded preambles. Let $\tilde{N}_i$ be $\hat{N}_i$ after the correction of the preamble loss probability. Assume that σ of the preambles are undecoded incorrectly, we can set $\tilde{N}_i$ as follows:

$$\tilde{N}_i = \begin{cases} f((1-\sigma)U_i); & U_i > 0, \\ 2f(1) - f(2); & U_i = 0. \end{cases} \tag{12}$$

Let P_{S} be the probability that only one device selects a preamble. We have

$$P_{\mathrm{S}} = \frac{E[S_i|M_i = m]}{R_i} = \left(1 - \frac{p_i}{R_i}\right)^{m-1}. \tag{13}$$

Suppose that λ is the average number of devices that passed the first ACB check. Let N_q and $E[N_q]$ be the number of devices with the q-th preamble transmission that passed the ACB check and its statistical expectation, respectively. We have

$$E[N_q] = \begin{cases} \lambda; & q = 1, \\ \prod_{j=1}^{q-1} \lambda \left\{1 - P_j P_{\mathrm{S}}\right\}; & q > 1. \end{cases} \tag{14}$$

Let P_{PL} be the preamble loss probability given that a device transmits a preamble. P_{PL} can be derived as follows:

$$P_{\mathrm{PL}} = \frac{\sum_{q=1}^{Q_{\max}} (1 - P_q) E[N_q]}{\sum_{q=1}^{Q_{\max}} E[N_q]}. \tag{15}$$

Assume that the system controls N_i to optimize the throughput, which means that $E[N_i] = R_i$. A k-th preamble is not decoded in the BS when the preamble transmissions from D_k devices are lost given that the k-th preamble is transmitted. Therefore, σ can be obtained as follows:

$$\begin{aligned} \sigma &= \frac{\sum_{n=1}^{\infty} \sum_{x=1}^{n} P(D_k = x|N_i = n) P(N_i = n) (P_{\mathrm{PL}})^x}{\sum_{n=1}^{\infty} \sum_{x=1}^{n} P(D_k = x|N_i = n) P(N_i = n)} \\ &\simeq \frac{\sum_{x=1}^{E[N_i]} P(D_k = x|N_i = E[N_i]) (P_{\mathrm{PL}})^x}{\sum_{x=1}^{E[N_i]} P(D_k = x|N_i = E[N_i])} \\ &= \frac{\sum_{x=1}^{R_i} \binom{R_i}{x} \left(\frac{1}{R_i}\right)^x \left(1 - \frac{1}{R_i}\right)^{R_i - x} (P_{\mathrm{PL}})^x}{\sum_{x=1}^{R_i} \binom{R_i}{x} \left(\frac{1}{R_i}\right)^x \left(1 - \frac{1}{R_i}\right)^{R_i - x}}. \end{aligned} \tag{16}$$

Algorithm 1 Decision for Q

procedure DECISION FOR $Q(P_T, Q_L, Q_U)$
 ▷ This procedure is called in BS at $t = 0$
 for $q = Q_L$ to Q_U **do**
  Derive P_F using $Q_{\max} = q$.
  if $P_F \le P_T$ **then**
   Return $Q = q$.
  end if
 end for
 Return as failure.
end procedure

Let $\hat{M}_i$ be the estimate of M_i. Let $E[N_i|M_i = m]$ be the statistical expectation of N_i given that $M_i = m$. From $E[N_i|M_i = m] = mp_i$, the BS can obtain $\hat{M}_i$ as follows:

$$\hat{M}_i = \frac{\tilde{N}_i}{p_i}. \tag{17}$$

B. RELIABILITY AWARE PREAMBLE RETRANSMISSION

We propose the selection of $Q_{\max}$ to achieve a desired AFP in this subsection. The throughput maximization does not ensure the AFP in the system to a certain level since the collision probability in the RA procedure and decoding failure probability can increase the AFP, although the throughput is maximized.

Let $P^*_{\mathrm{S},q}$ be the RA success probability with p^* and R^* in the q-th preamble transmission. Since $E[N_i|M_i = m] = mp_i = R_i$, we have

$$\begin{aligned} P^*_{\mathrm{S},q} &= \left(1 - \frac{1}{e^q}\right) \frac{E[S_i|M_i = m]}{E[N_i|M_i = m]} \\ &= \left(1 - \frac{1}{e^q}\right)\left(1 - \frac{1}{m}\right)^{m-1}. \end{aligned} \tag{18}$$

Let P^*_F be the failure probability of a device with p^* and R^*. We have

$$P^*_F = \prod_{q=1}^{Q_{\max}} (1 - P^*_{\mathrm{S},q}). \tag{19}$$

We need to change P^*_F to be lower than P_T. Since we cannot change the preamble detection probability, we need to increase $Q_{\max}$ or the $P^*_{\mathrm{S},q}$. However, the change in $P^*_{\mathrm{S},q}$ decreases the throughput, therefore we require a change in $Q_{\max}$.

Let's assume that m is very large since the ACB generally activates in this condition. Let $P_{S,q}$ be the RA success probability with the assumption, which is equal to

$$P_{\mathrm{S},q} = \lim_{m \to \infty} P^*_{\mathrm{S},q} = \left(1 - \frac{1}{e^q}\right) e^{-1}. \tag{20}$$

Let P_F be the AFP of a device with $P_{\mathrm{S},q}$, which is equal to

$$P_F = \prod_{q=1}^{Q_{\max}} (1 - P_{\mathrm{S},q}). \tag{21}$$

Algorithm 2 Reliability Aware Adaptive ACB Algorithm

procedure ADAPTIVEACB(U_i)
  ▷ This procedure is called in the BS at every end of RACH
  $\tilde{N}_i = \begin{cases} f((1-\sigma)U_i) = \dfrac{\log((1-\sigma)U_i/R_i)}{\log(1-1/R_i)}; & U_i > 0, \\ 2f(1) - f(2); & U_i = 0. \end{cases}$
  $\hat{M}_i = \tilde{N}_i / p_i$.
  $R_{i+1} = R_{\max}$.
  $p_{i+1} = \begin{cases} \dfrac{R_{i+1}}{\hat{M}_i}; & \hat{M}_i > R_{\max} \\ 1; & \hat{M}_i \le R_{\max} \end{cases}$
  Announce R_{i+1} and p_{i+1}.
end procedure

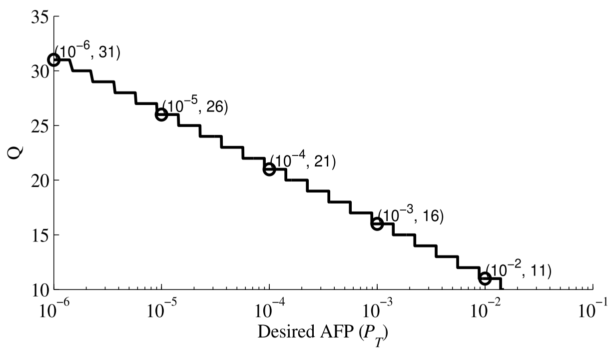


FIGURE 4. Q for desired AFP for $Q_L = 10$ and $Q_U = 40$.

Let Q be a minimum of $Q_{\max}$ to achieve $P_F \le P_T$. Q may be obtained by rearranging $P_F = P_T$. However, since the rearrangement is hard to summarize as a short form of the equation, instead we present a search algorithm to obtain Q for P_T. Let Q_L and Q_U be the arbitrary lower and upper limits for $Q_{\max}$, respectively. The decision algorithm for Q is represented in Algorithm 1.

Fig. 4 shows Q with respect to the desired AFP for $R_i = 32$. The RA procedure requires $Q_{\max}$ of 11, 16, 21, 26, and 31 to achieve an AFP lower than 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, and 10^{-6}, respectively. Note that, the correction for Q will be required since the error between N_i and $\hat{N}_i$, and that between M_i and $\hat{M}_i$ in the actual system will change the AFP. We'll correct Q in the evaluation of the framework.

C. ADAPTIVE ESTIMATION AND ACB

In this subsection, we present the adaptive estimation and ACB algorithm to adaptively update the ACB related parameters (R_i, and p_i).

Suppose that the preamble transmission in i-th RACH is completed. The BS can obtain $\tilde{N}_i$ and $\hat{M}_i$ from U_i. Regardless of $\hat{M}_i$, $R_{i+1} = R_{\max}$ for a high throughput and AFP. If $\hat{M}_i$ is lower than $R_{\max}$, the system does not require an ACB, and thus the device does not need to do the ACB trial: $p_{i+1} = 1$. Otherwise, the BS requires calculating p_{i+1} based on (6). Since the last observation shows $m = \hat{M}_i$,

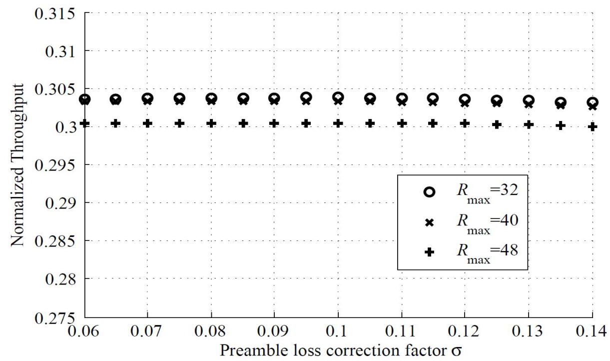


FIGURE 5. Normalized throughput vs. preamble loss correction factor σ.

we can select p_{i+1} as follows:

$$p_{i+1} = \frac{R_{\max}}{\hat{M}_i}. \tag{22}$$

The BS also determines $Q_{\max}$ from Algorithm 1. After determining the parameters, the BS announces new parameters by broadcasting a message to the devices. The adaptive access class barring algorithm with the estimation of state information is represented as a pseudo code in Algorithm 2.

VI. PERFORMANCE EVALUATION

In this section, we present the results of the performance evaluation for the proposed reliability control framework. In Section VI-A, we present the environment for the performance evaluation. In the following sections, we'll present and analyze the results of the performance evaluation for the proposed framework. In addition, the correction is performed for the selection of $Q_{\max}$ to mitigate the gap between AFP in the numerical analysis and AFP in the simulation, where the gap occurs due to an inevitable error in the estimation model.

A. EVALUATION ENVIRONMENTS

In this subsection, we present the environment used for the performance evaluation. The activation time T_A is selected as 10,000 and the RACH interval I_A as 5 subframes. $R_{\max}$ is 32 in the evaluation. T_{RAR}, W_{RAR}, T_{MSG4}, and W_{BO} are 3, 5, 48, 20 subframes, respectively, as recommended by 3GPP [13]. T_{PROC} is 5 subframes as in [29]. If not specified, σ is equal to 0.0992 which is from (16) with a large number of $Q_{\max}$. The simulation model is developed using OPNET Modeler 14.5 based on the simulation model in 3GPP TR 37.868 [13]. More than 1500 simulations are performed for each point in the following figures. We confirmed that the standard error for each point is lower than 0.5% of the mean value.

B. THROUGHPUT

To check the parameter configuration for σ, we compared the throughput with respect to σ. Fig. 5 shows the normalized throughput with respect to σ where M is 30,000 and $Q_{\max}$ is 11. The normalized throughput is the ratio of the number of RA success per RACH to $R_{\max}$. The ACB can control

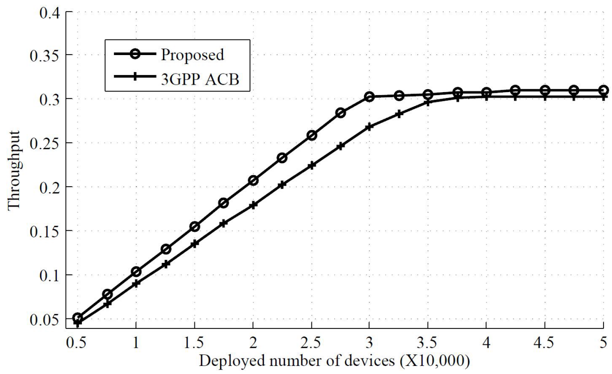


FIGURE 6. Throughput vs. M with 48 preambles.

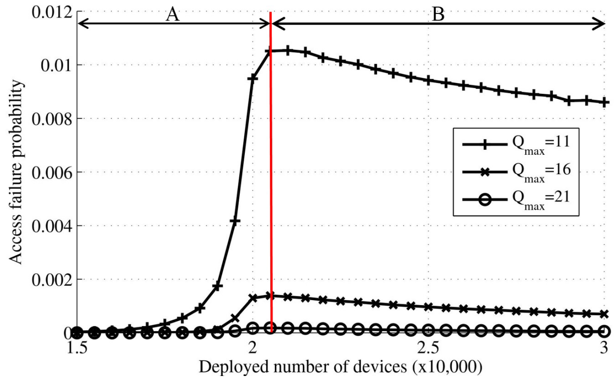


FIGURE 7. AFP vs. M from the simulation.

the number of simultaneous contending devices as $R_{\max}$. The collision probability and throughput are constant regardless of $R_{\max}$. The normalized throughput is maximized around $\sigma = 0.1$ as the numerical analysis expected since the BS requires the correct state information to adaptively select p_i.

We present a throughput compared with existing 3GPP ACB which is described in the concept of original ACB [13] with respect to M where the number of preambles is 48. We apply $\sigma = 0.1$ to the proposed algorithm based on the results of fig. 5 and $Q_{\max}$ is 11. Fig. 6 shows that the proposed algorithm has better performance when the number of deployed devices is between 5,000 and 35,000. Both throughputs converge to around 0.3 when $M > 35,000$.

C. ACCESS FAILURE PROBABILITY AND CORRECTION FOR Q

Fig. 7 shows the AFP with respect to M. The AFP with a very small number of devices (Region A) shows a very low AFP due to the low probability of collision. The devices do not need to do the ACB trial until the AFP reaches the desired value. The AFP with a large number of devices (Region B) shows values lower than P_T, which means that the derived Q is valid to achieve an AFP lower than P_T. However, the AFP for the middle range of M becomes much greater than the target AFP. In this case, two reasons cause the unexpected results: First, M_i is not enough to assume that $P_S = e^{-1}$.

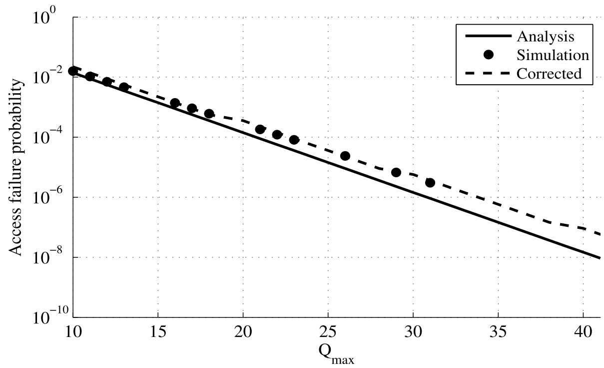


FIGURE 8. AFP from analysis, simulation, and by correction.

Second, the estimation method has inevitable error since the domain of the number of unused preambles (U_i) is smaller than that for the number of contending devices (N_i). Therefore, we can conclude that a correction for $Q_{\max}$ is required to achieve the desired AFP regardless of the number of devices in the cell.

Fig. 8 shows the AFP with respect to $Q_{\max}$. "Analysis" shows the P_F from (21), and "Simulation" shows the maximum of AFP measured from simulations corresponding to $Q_{\max}$. As shown in Fig. 7 and Fig. 8, some error occurs between the results in the analysis and the results in the simulation. Let Q_C be the Q with correction for mapping between Q from the numerical analysis and $Q_{\max}$ from the simulated result. From the result, we can obtain Q_C as follows:

$$Q_C = \left\lfloor \frac{10}{9} Q \right\rfloor, \tag{23}$$

where $\lfloor x \rfloor$ is the largest integer equal or lower than x. The line with "Corrected" in Fig. 8 shows P_F when Q_C is used as $Q_{\max}$. As shown in the figure, Q_C shows a better fit to the simulated results than Q from the analysis. From the results, we can expect that Q_C can be used to achieve the desired AFP. For example, $Q_{\max}$ of 12, 17, and 23 are recommended to achieve an AFP under 10^{-2}, 10^{-3}, and 10^{-4} regardless of the number of devices in the cell, respectively, where Q_C correspond to these values.

D. DELAY AND THE NUMBER OF TRANSMISSIONS

Fig. 9 shows the average delay of the devices for three different $Q_{\max}$ where the delay is the time between the activation of a device and successful access (if the access of a device failed, the delay for the device is not collected). The average delay increases as $Q_{\max}$ and M increase. As shown in this figure, M M is more critical to the delay compared to $Q_{\max}$. Fig. 10 shows the average number of preamble transmissions for the successful devices. The number of transmissions converges to around 3.5 and slightly increases as $Q_{\max}$ increases. From the convergence of the number of transmissions, we can expect that the delay increment above 20,000 deployed devices is due to the ACB trial failure where each failure postpones the end of the access by I_A subframes.

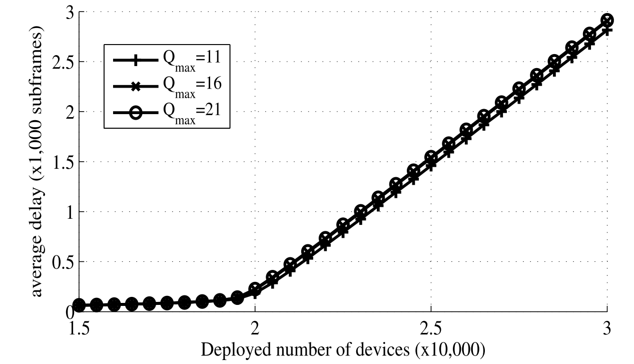


FIGURE 9. Average access delay vs. *M*.

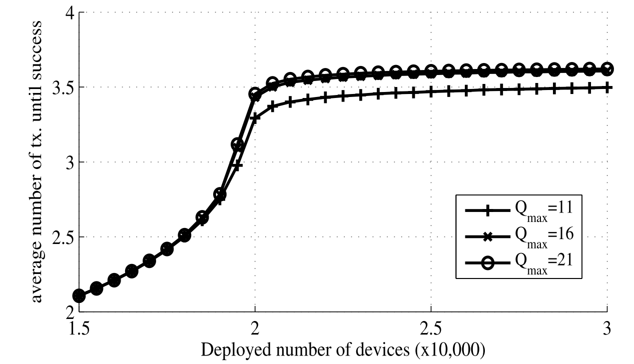


FIGURE 10. Average number of preamble transmissions vs. *M* for successful devices.

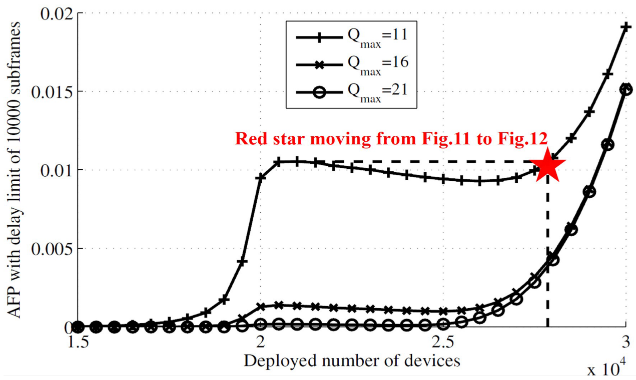


FIGURE 11. AFP vs. *M* with delay limit of 10,000 subframes.

E. CAPACITY FOR DELAY LIMITED APPLICATIONS

Suppose that the limit of delay is given as 10,000 subframes since 1 subframe corresponds to 1 ms and 10 seconds is the 3GPP target for delay limits for IoT services [13]. With the delay limit, if the device experiences a delay larger than 10,000 subframes, it results in an RA failure even if the devices received MSG4. Fig. 11 shows the AFP when the limit of the delay is given as 10,000 subframes for three different $Q_{\max}$. The ACB scheme is used or not at peak point. When the number of devices increases more than 20,000,

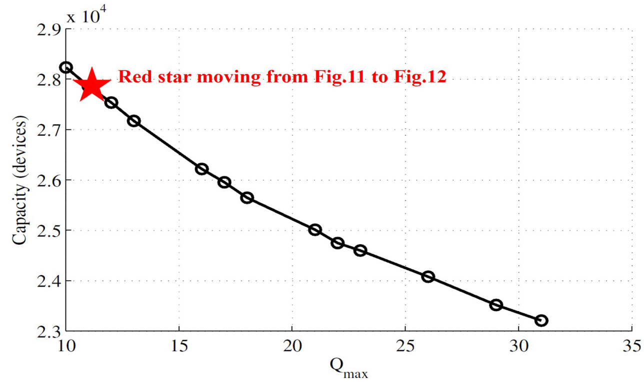


FIGURE 12. Capacity vs. $Q_{\max}$ with delay limit of 10,000 subframes.

the AFP slighty decrease to be satisfied desired access failure probability. When the number of devices increases to more than 25,000, the AFP suddenly rises due to the large number of deferences, where the rising AFP is due to the increased the delay from ACB.

Let the capacity be the maximum M which satisfies the AFP lower than the first peak in Fig. 11. For example, the capacity for $Q_{\max} = 11$ can be obtained as the dotted line: drawing a horizontal line over the local maximum, finding a point where the horizontal line intersects the AFP, and taking the x-axis value of the point as the capacity. We can figure out that proposed algorithm can accommodate 27,862 devices for $Q_{\max} = 11$ (AFP $= 10^{-2}$) from a red star moving from Fig. 11 to Fig. 12. As shown in this figure, we can obtain the capacity with respect to $Q_{\max}$. The capacity decreases almost linearly as $Q_{\max}$ increases, which means the number of devices should decrease to increase the reliability of the service when the delay limit is given.

VII. CONCLUSION

This paper reviews the RA procedure and the ACB of eMTC in terms of reliability. We propose a reliability control framework to provide the desired level of reliability for massive numbers of IoT devices using eMTC. The framework includes an estimation model for the state information of the cell where the estimation is performed by using the number of undecoded preambles. The adaptive estimation and ACB factor decision algorithm based on the estimation model are also presented in this paper. The estimation model considers the information available in the BS and the preamble loss probability in the channel. In addition, the framework includes an algorithm to obtain the maximum number of preamble transmissions according to the desired access failure probability. This paper can conclude that $Q_C = \left\lfloor \frac{10}{9} Q \right\rfloor$ can be the maximum number of preamble transmission to achieve the desired AFP. The results of the performance evaluation show that the parameters configured by our framework can provide the desired AFP of the RA procedure. Our evaluation indicates that our proposed framework can be used to provide the desired AFP for IoT users in eMTC.

REFERENCES

[1] W. Ejaz *et al.*, "Internet of Things (IoT) in 5G wireless communications," *IEEE Access*, vol. 4, pp. 10310–10314, 2016.

[2] *Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2*, document TS 36.300 V15.2.0, 3GPP, Jun. 2018.

[3] *Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Medium Access Control (MAC) Protocol Specification*, document TS 36.321 V15.2.0, 3GPP, Jul. 2018.

[4] *Technical Specification Group Core Network and Terminals; Non-Access-Stratum (NAS) Protocol for Evolved Packet System (EPS); Stage 3*, document TS 24.301 V15.3.0, 3GPP, Jun. 2018.

[5] S. Vural, N. Wang, P. Bucknell, G. Foster, R. Tafazolli, and J. Müller, "Dynamic preamble subset allocation for RAN slicing in 5G networks," *IEEE Access*, vol. 6, pp. 13015–13032, 2018.

[6] L. Ferdouse, A. Anpalagan, and S. Misra, "Congestion and overload control techniques in massive M2M systems: A survey," *Trans. Emerg. Telecommun. Technol.*, vol. 28, no. 2, p. e2936, Mar. 2015. doi: 10.1002/ett.2936.

[7] A. Ksentini, Y. Hadjadj-Aoul, and T. Taleb, "Cellular-based machine-to-machine: Overload control," *IEEE Netw.*, vol. 26, no. 6, pp. 54–60, Nov. 2012.

[8] A. G. Gotsis, A. S. Lioumpas, and A. Alexiou, "M2M scheduling over LTE: Challenges and new perspectives," *IEEE Veh. Technol. Mag.*, vol. 7, no. 3, pp. 34–39, Sep. 2012.

[9] Google Cloud. (Nov. 2017). *Overview of Internet of Things*. [Online]. Available: https://cloud.google.com/solutions/iot-overview

[10] T. P. C. de Andrade, C. A. Astudillo, L. R. Sekijima, and N. L. S. 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] *Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC); Protocol Specification*, document TS 36.331 V15.2.2, 3GPP, Jun. 2018.

[12] 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.

[13] *Technical Specification Group Radio Access Network; Study on RAN Improvements for Machine-Type Communications;* document TS 37.868 V11.0.0, 3GPP, Sep. 2011.

[14] S.-H. Lee, S.-Y. Jung, and J.-H. Kim, "Dynamic resource allocation of random access for MTC devices," *ETRI J.*, vol. 39, no. 4, pp. 546–557, Aug. 2017.

[15] H. He, Q. Du, H. Song, W. Li, Y. Wang, and P. Ren, "Traffic-aware ACB scheme for massive access in machine-to-machine networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, London, U.K., Jun. 2015, pp. 617–622.

[16] M. Tavana, V. Shah-Mansouri, and V. W. S. Wong, "Congestion control for bursty M2M traffic in LTE networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, London, U.K., Jun. 2015, pp. 5815–5820.

[17] C. Anton-Haro and M. Dohler, *Machine-to-machine (M2M) Communications: Architecture, Performance and Applications*. Amsterdam, The Netherlands: Elsevier, 2014.

[18] J. Choi, "On the adaptive determination of the number of preambles in RACH for MTC," *IEEE Commun. Lett.*, vol. 20, no. 7, pp. 1385–1388, Jul. 2016.

[19] G. Liva, "Graph-based analysis and optimization of contention resolution diversity slotted ALOHA," *IEEE Trans. Commun.*, vol. 59, no. 2, pp. 477–487, Feb. 2011.

[20] A. A. Purwita and K. Anwar, "Massive multiway relay networks applying coded random access," *IEEE Trans. Commun.*, vol. 64, no. 10, pp. 4134–4146, Oct. 2016.

[21] S. Duan, V. Shah-Mansouri, and V. W. S. Wong, "Dynamic access class barring for M2M communications in LTE networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Atlanta, GA, USA, Dec. 2013, pp. 4747–4752.

[22] Z. Wang and V. W. S. 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.

[23] J. Moon and Y. Lim, "Adaptive access class barring for machine-type communications in LTE-A," in *Proc. 8th Int. Conf. Ubiquitous Future Netw. (ICUFN)*, Vienna, Austria, Jul. 2016, pp. 398–402.

[24] M. Koseoglu, "Pricing-based load control of M2M traffic for the LTE-A random access channel," *IEEE Trans. Commun.*, vol. 65, no. 3, pp. 1353–1365, Mar. 2017.

[25] 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.

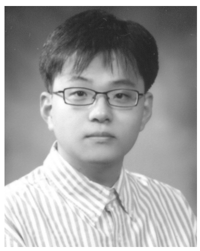
[26] I. Leyva-Mayorga, M.-A. R. Hernandez, V. Pla, J. Martinez-Bauset, and L. Tello-Oquendo, "An adaptive access class barring scheme for handling massive M2M communications in LTE-A," in *Proc. 23th Eur. Wireless Conf.*, Dresden, Germany, May 2017, pp. 1–6.

[27] C. Kalalas and J. Alonso-Zarate, "Reliability analysis of the random access channel of LTE with access class barring for smart grid monitoring traffic," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Paris, France, May 2017, pp. 724–730.

[28] G. Khandelwal, A. Yener, K. Lee, and S. Serbetli, "ASAP : A MAC protocol for dense and time constrained RFID systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Istanbul, Turkey, vol. 9, Jun. 2006, pp. 4028–4033.

[29] *Technical Specification Group Radio Access Network; Feasibility Study for Further Advancements for E-UTRA (LTE-Advanced)*, document TS 36.912 V15.0.0, 3GPP, Sep. 2018.

SO YI JUNG received the B.S. and M.S. degrees from the Department of Electrical and Computer Engineering, Ajou University, Suwon, South Korea, in 2013 and 2015, respectively, where she is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering. From 2015 to 2016, she was the Head Researcher in electrical and energy research with the Korea Testing and Research Institute. Her current research interests include random access protocols, network optimization for the NB-IoT, mobile edge caching, and cellular-V2X communication.



SUNG HYUNG LEE received the B.S., M.S., and Ph.D. degrees in electrical engineering from the Department of Electrical and Computer Engineering, Ajou University, Suwon, South Korea, in 2007, 2009, and 2017, respectively. His current research interests include network simulations, random access protocols, and protocol design for military network systems.

JAE-HYUN KIM received the B.S., M.S., and Ph.D. degrees in computer science and engineering from Hanyang University, Ansan, South Korea, in 1991, 1993, and 1996, respectively. In 1996, he was with the Communication Research Laboratory, Tokyo, Japan, as a Visiting Scholar. From 1997 to 1998, he was a Postdoctoral Fellow of the Department of Electrical Engineering, University of California at Los Angeles. From 1998 to 2003, he was a Member of Technical Staff with the Performance Modeling and QoS Management Department, Bell Laboratories, Lucent Technologies, Holmdel, NJ, USA. He has been with the Department of Electrical and Computer Engineering, Ajou University, Suwon, South Korea, as a Professor, since 2003. He is the Center Chief of Satellite Information Convergence Application Services Research Center (SICAS) supported by the Institute for Information and Communications Technology Promotion in South Korea. He is the Executive Director of the Korea Institute of Communication and Information Sciences (KICS). His research interests include medium access control protocols, QoS issues, cross layer optimization for wireless communication, and satellite communication. He is a member of KICS, the Institute of Electronics and Information Engineers, and the Korea Information Science Society. He has been the Chair of the Smart City Committee of 5G Forum in South Korea, since 2018.