Performance Analysis of Group Paging for Machine-Type Communications in LTE Networks

Chia-Hung Wei, Ray-Guang Cheng, *Senior Member, IEEE*, and Shiao-Li Tsao, *Member, IEEE*

Abstract—Machine-type communication (MTC) is a crucial service for next-generation cellular networks. Mass access to the network by MTC devices may result in the overload of radio access networks (RANs) and degrade the service quality of human-to-human communication. Group paging is one of the mechanisms proposed to alleviate the RAN-overload problem. This paper presents an analytical model based on a recursive contending-users estimation (RCE) method proposed in the literature to derive the performance metrics of collision probability, access success probability, average access delay, statistics of preamble transmissions, statistics of access delay, and utilization of random-access opportunities (RAOs) for group paging with various combinations of group sizes and reserved radio resources in a paging access interval. The optimal group size and required RAOs are subsequently derived based on the given target access success probability. Numerical results demonstrate that the proposed model can accurately estimate the performance of group paging.

Index Terms—Group paging, machine-type communications (MTC), overload control, random access.

I. INTRODUCTION

MACHINE-TYPE communication (MTC), which is also known as machine-to-machine communication (M2M) in the IEEE 802.16 Working Group, is a new service defined by the Third-Generation Partnership Project (3GPP) to facilitate machines communicating with each other over current cellular networks [2]. MTC usually involves a large number of MTC devices to support a wide range of applications, such as smart grid, road security, and consumer electronic devices. However, concurrent access to a radio network by mass MTC devices may result in intolerable delays, packet loss, or even service unavailability to current human-to-human (H2H) communication services. Hence, proper overload control mechanisms are required to guarantee network availability and quality of H2H services under heavy MTC load [2].

Overload control of an uplink random-access channel (RACH) in a radio access network (RAN) is one of the principle

C.-H. Wei and R.-G. Cheng are with the Department of Electronic and Computer Engineering, National Taiwan University of Science and Technology, Taipei 106, Taiwan (e-mail: crg@ieee.org).

S.-L. Tsao is with the Department of Computer Science, National Chiao Tung University, Hsinchu 300, Taiwan.

working items for 3GPP Long-Term Evolution (LTE) [2]. The purpose of RAN-overload control is to avoid RAN overload when mass MTC devices simultaneously contend for the RACH. From the perspective of the way MTC traffic is generated, RAN-overload control schemes can be categorized into push- and pull-based approaches [3]. In the push-based approach, MTC traffic is *pushed* from MTC devices to the network without any restriction until RAN overload is detected. In the pull-based approach, MTC traffic is *pulled* by the network, and thus, the network may properly control the MTC traffic load through paging and, thus, prevents RAN overload.

Access class barring (ACB), separated RACH resources for MTC, dynamic allocation of RACH resource, MTC-specific backoff scheme, and slotted access are examples of push-based RAN-overload control schemes [2]. In the ACB scheme, the network separates the MTC traffic into several access classes and assigns an ACB factor to each MTC access class. Each cell can control the channel access probability of a specific MTC access class by setting the ACB factor. In the separated RACH scheme, the network reserves dedicated RACH resources for the H2H and MTC devices to provide them with distinct access collision probabilities. In the dynamic allocation of the RACH resource scheme, the network dynamically allocates additional RACH resources for the MTC devices based on the predicted access load of MTC devices. The MTC-specific backoff scheme delays the random-access reattempts/attempts of MTC devices by assigning an MTC-specific random backoff procedure. In the slotted-access scheme, each MTC device is associated with dedicated access cycles/slots (similar to paging cycles/slots) through its identity. Each MTC device can transmit the random-access attempt only at its random-access slot. The advantages and disadvantages of various push-based RAN-overload control schemes are summarized in [4].

Paging and group paging are potential pull-based RAN-overload control schemes [5]. In LTE, a downlink paging channel is defined to transmit paging information to user equipment (UE), informing UE on system information changes and emergency notifications. The network may transmit a paging message to activate a specific UE at the UE's paging occasion. The paging occasion of each UE is determined according to its UE identity. The current paging mechanism that was originally designed for H2H services can only page up to 16 devices with a single paging message, and only two paging occasions are available per 10-ms radio frame [5]. Therefore, a base station (BS) must transmit multiple paging messages over a long period to activate a large number of MTC devices. Therefore, a group-paging mechanism that uses a single group-paging message to activate a group of MTC devices is proposed to address this

issue [2]. In group paging, an MTC device is assigned by a unique group identity (GID) after camping on a network and joining a group. All of the MTC devices in a group listen to the same paging channel at the same paging occasion derived from the GID [5]. The group of MTC devices shall simultaneously perform the standard LTE random-access procedure to access the network when they find their GID in a group-paging message. The MTC devices with random access that fail shall follow the standard LTE random backoff procedure to retransmit their random-access attempts during a paging access interval until the retry limitation exceeds. Note that the network may use the group-paging message to notify MTC devices of the paging access interval and the dedicated random-access resources reserved for group paging.

The first simulation study of group paging is given in [6]. The main performance metrics specified in 3GPP TR 37.868 [2], which include *collision probability*, *access success probability*, *statistics of number of preamble transmissions*, and *statistics of access delay*, were investigated. The preliminary study showed that group paging is a promising solution for RAN-overload control. However, it lacks a general rule for LTE operators to determine the proper number of reserved random-access resources and the group size. This paper aims to develop an analytical model to investigate the performance of group paging and to suggest related parameters. The analytical model needs to consider all of the implementation constraints specified in LTE [2]. The implementation constraints include the LTE random backoff procedure; the limited capacity of the downlink access grant channel, which results in failed random access even if the random-access attempts are not collided; the maximum number of retransmissions for failed random-access attempts; the exponential law for the power-ramping effect, which results in a time-varying detection probability for the random-access attempts; and the transmission delay of the message part considering the effect of a hybrid automatic retransmission request (HARQ) procedure.

The random-access architecture of LTE is similar to a multichannel slotted ALOHA system. Much research has been devoted to the slotted ALOHA system in single-channel [7], [8] or multichannel slotted systems [9]–[16]. In [7], the throughput of single-channel slotted ALOHA systems as a function of a constant offered load was presented. In [8], the relationship between the throughput and the average access delay of a finite-user single-channel slotted ALOHA system was investigated. For multichannel slotted ALOHA systems, the performance metrics of the throughput [9], [10], [13], average access delay [11]–[15], and collision and success probabilities [16] were also discussed. The purposes of these studies were to reduce access delay [9], adjust the design parameters to stabilize the channels [10]–[12], maximize the throughput [13], [14], or realize the tradeoff between the average throughput and the average access delay [15]. A finite-population system was considered in [11] and [12]. In [11], the stability and delay of finite-user slotted ALOHA systems with multipacket reception were investigated. In [12], Jelenkovic and Tan showed that the variability of packet sizes results in power law delays for finite-population ALOHA systems. In [14], Zhou *et al.* presented

closed-form expressions of throughput and access delay for orthogonal frequency-division multiple-access systems. The analysis was derived based on an assumption of a constant successful transmission probability. The collision and success probabilities defined from the perspective of a user and a RACH in LTE systems were discussed in [16]. The effect of the random backoff procedure in slotted ALOHA systems was considered in [7]–[9], [13], [14], and [16]. The effect of the time-varying preamble detection probability due to a fading channel was considered in [7]. The impact of the backoff window size on the average access delay was discussed in [8]. Access randomness in the time domain was considered in [8] and [9], access randomness in the frequency domain was considered in [13], and access randomness in the time, code, and frequency domains was considered in [14] and [16]. Most of the aforementioned studies focus on the uniform backoff policy. Both the binary exponential and uniform backoff policies were considered in [14]. The constraint of the maximum number of retransmissions in the random-access procedure was considered in [13] and [14].

Most of the performance evaluations of ALOHA systems focused on average value (throughput and/or access delay) analysis [9]–[11], [13], and the distributional property was only addressed in [12]. Existing studies normally assumed that new packet arrivals in a slot time follow a Bernoulli distribution [10], [11] or a Poisson distribution [7], [9], [12]–[15] with a constant rate, and thus, the combined rate of new and retransmission packets in a slot is a constant. In group paging, the number of MTC devices to be paged is known, and the MTC devices access the network in a highly synchronized manner once they are paged. However, the random-access attempts in each random-access slot are not fixed because no new arrival is generated and the number of MTC devices gradually decreases if any device successfully access the RACH. Moreover, the analytical model has to consider the system environment (for example, the power-ramping effect) and the distributional properties related performance metrics (that is, statistics of preamble transmissions and access delay) of LTE. Therefore, current analytical models cannot be directly applied to the performance analysis of group paging in LTE networks. This paper presents an analytical model to derive the performance metrics of group paging by considering all of the parameters defined in the LTE random-access procedure [17]. Similar to existing approaches, we use the known paging group size and the Poisson approximation model [1] to estimate the number of success and failed (or collided) devices in the first random-access slot. Different from current approaches, the numbers of contending devices in the successive random-access slots are then individually derived as a function of the number of failed devices, the time-domain backoff parameters, the limited capacity of the downlink access grant channel, and the time-varying detecting probability. The performance metrics of group paging are then derived based on the estimated number of success and failed devices obtained from all random-access slots. The remainder of this paper is organized as follows: The system model and the analytical model are described in Section II. Section III presents the numerical results. Finally, conclusions are offered in Section IV.
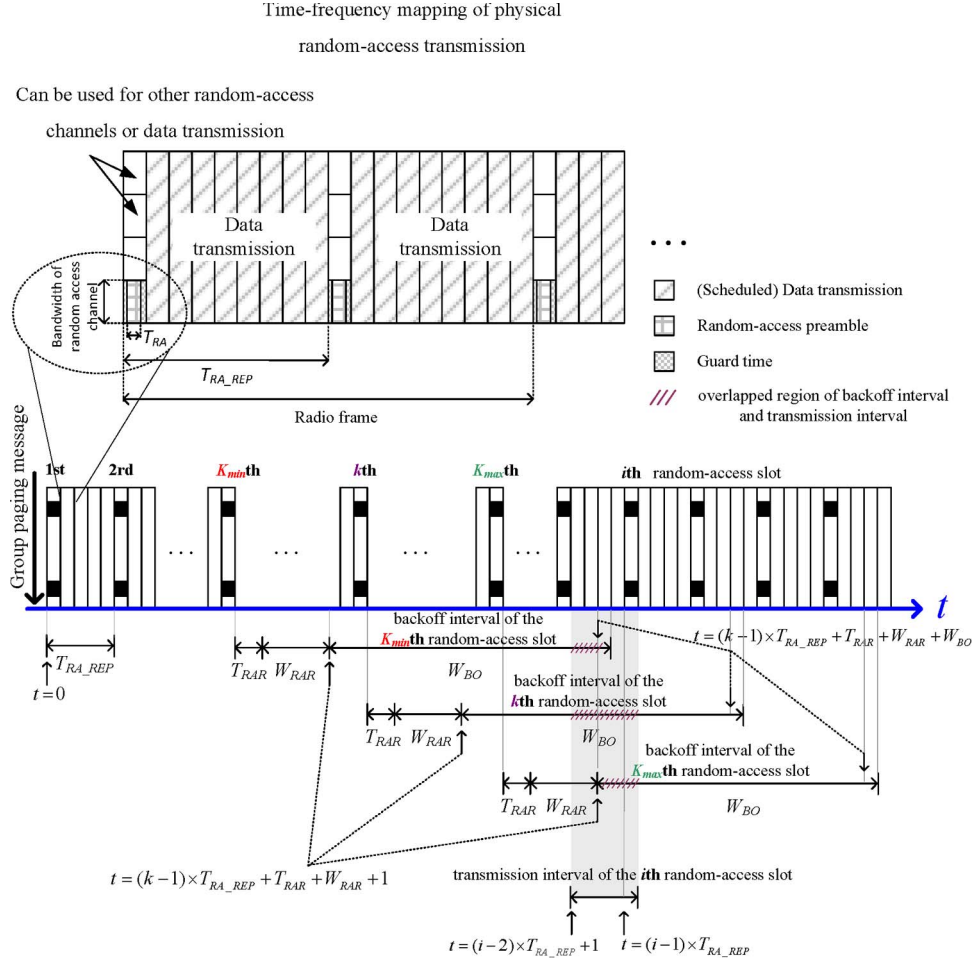
Fig. 1. Timing diagram of physical random-access transmission of LTE.

II. SYSTEM MODEL

This study considers a group of M MTC devices in a paging area containing K cells in an LTE network. Without loss of generality, we considered a case where the MTC devices are uniformly distributed in K cells, and thus, each cell has M/K MTC devices. It is assumed that each BS in the paging area reserves an amount of R dedicated random-access resources and sends a group-paging message containing a GID to page the M/K MTC devices simultaneously. Once the group-paging message is received, all M/K MTC devices follow the standard LTE random-access procedure to establish connections with the BS. The performance of random access during a paging access interval will be investigated. The paging access interval of group paging starts from the first random-access slot and ends at the $I_{\max}$th random-access slot, where $I_{\max}$ is the number of random-access slots reserved for group paging. Note that $I_{\max}$ depends on the maximum number of preamble transmissions of the random-access procedure and will be derived later in this paper.

Fig. 1 shows time-frequency mapping [18] and the timing diagram of physical random-access transmission of LTE. In LTE, time is divided into fix-length radio frames. Each radio frame consists of multiple subframes. Random-access transmissions are restricted to specific subframes [16], which are referred to as random-access slots in the rest of this paper.

In LTE, the random-access resource is determined in terms of random-access opportunities (RAOs). The total number of RAOs provided by a BS in a random-access slot is equal to the number of frequency bands in each random-access slot multiplied by the number of random-access preambles [17]. In LTE, one access window of length T_{RA} per $T_{\mathrm{RA\_REP}}$ period is allocated for random access [17], as shown in Fig. 1. $T_{\mathrm{RA\_REP}}$ is the interval between two successive random-access slots and can be obtained from the physical random access channel (PRACH) configuration index announced by the BS. For example, $T_{\mathrm{RA\_REP}} = 5$ radio frames [19] if PRACH configuration index 6 is used [2].

The timing diagram shown in Fig. 1 shows the behavior of an MTC device receiving a group-paging message. In this paper, the time axis starts from the first random-access slot in which all MTC devices send their first random-access attempts immediately after receiving the group-paging message from the BS. Hence, the time of the ith random-access slot is $(i - 1) \times T_{\mathrm{RA\_REP}}$, as shown in Fig. 1. Before going into the details, the LTE random-access procedure is first elaborated upon, as follows.

A. LTE Random-Access Procedure

Fig. 2 shows the LTE two-step random-access procedure [17], which separates the transmission of the random-access
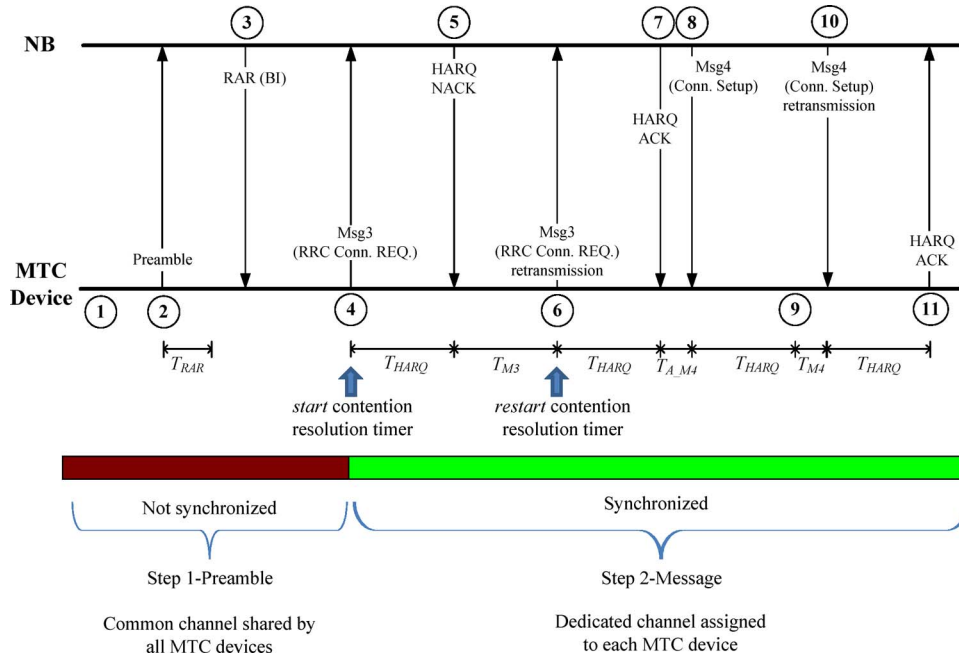
Fig. 2.   Two-step random access (the numbers describe different events during a random-access procedure) [21].

preamble and the associated signaling messages (Msg 3 and Msg 4) for a connection setup. The preamble is transmitted through a common channel shared by multiple UEs, and the signaling messages are transmitted in a dedicated channel specifically reserved for UE. In the first step, UE synchronizes to the downlink timing [(**1**) in Fig. 2], randomly selects a random-access preamble from a group of preambles reserved for the RACH, and transmits the random-access preamble in a randomly chosen random-access slot and a frequency band (**2**). The BS correlates the received signal with the set of possible sequences in a cell and transmits a response message indicating the detected preamble(s) (**3**). Each response message carries a medium access control (MAC) header and one or more MAC random-access responses (RARs). The MAC header may carry a *backoff indicator* (BI) (unit: subframe) to indicate the backoff parameter values for the collided or undetected UEs. In LTE, the range of *BI* is from 0 to 960 subframes [17]. Each RAR carries the identity (ID) of the preamble selected by the UE, the information to be used by the UE to adjust the uplink timing, and a dedicated uplink resource reserved for the UE to transmit the message [17].

Let T_{RAR} be the processing time required by the BS to detect the transmitted preambles (unit: subframe), W_{RAR} be the length of the random-access-response window (unit: subframe), W_{BO} be the length of the backoff window (unit: subframe) ($W_{\mathrm{BO}} = BI + 1$), and N_{PTmax} be the maximum number of preamble transmissions. As shown in Fig. 1, the random-access-response window starts at the end of the preamble transmission plus T_{RAR} [17] subframes, and the backoff window starts at the end of the random-access-response window. The UE should perform random backoff and retransmit its random-access attempt if it does not receive the response message within W_{RAR}. For each retransmitted random-access attempt, the UE must randomly choose a backoff counter from zero to BI, ramp up its transmission power, and transmit a newly selec-

ted random-access preamble in the next available random-access slot when the backoff counter decreases to zero. The process continues until N_{PTmax} preamble transmissions are reached.

Once the UE receives the response message from the BS and adjusts its uplink timing, the remaining signaling required for the connection setup is transmitted on the assigned dedicated uplink resource in a synchronized manner by using the same procedures as normal data transmission. Nonadaptive HARQ is subsequently enabled to protect the signaling exchange of the message. The UE, which successfully receives the RAR message, must send Msg3 carrying the UE ID and the "radio resource control connection request" message to the BS at the radio resource assigned by the BS [(**4**) in Fig. 2]. In response, the BS sends an HARQ acknowledgment (ACK) or negative-acknowledgment (NACK) after T_{HARQ} subframes. The BS waits for T_{A_M4} subframes and transmits Msg4 (**8**) after it replies an ACK indicating that Msg3 is successfully received (**7**). In contrast, the UE waits for T_{M3} subframes and retransmits Msg3 (**6**) if it receives an NACK (**5**). Similarly, the UE waits for T_{HARQ} subframes and sends an ACK to the BS if Msg4 is successfully received (**11**). The BS waits for T_{M4} subframes and retransmits Msg4 (**10**) if it does not receive an ACK for Msg4 (**9**). The HARQ retransmission of Msg3 and Msg4 can be up to N_{HARQ} times. The UE starts/restarts a contention resolution timer T_{CR} (unit: subframes) whenever it transmits Msg3. The UE declares a random-access failure and reverts to step 1 to retransmit its random-access attempt if the contention resolution timer expires. Note that Msg 3 and Msg 4 are used for carrying connection setup signaling messages and for contention resolution. In some cases, the BS may have a chance to decode the same preamble transmitted by multiple UE and reply a response message. The UE will transmit its own Msg3 on the same dedicated resource and then realize the random-access failure after the expiry of the contention resolution timer.

TABLE I
RANDOM-ACCESS RELATED SYSTEM PARAMETERS

Notations	Definitions	Settings in simulation
M/K	Average number of MTC devices in a cell	10~1000
W_{BO}	Backoff window size(unit: sub-frame)	$BI+1$
N_{UL}	Maximal number of MTC devices that can be acknowledged within the random-access response window	$N_{RAR} \times W_{RAR}$
T_{RA_REP}	Interval between two successive random-access slots (unit: sub-frame)	$T_{RA\_REP}=5$ if PRACH Configuration Index = 6
R	Total number of preambles in a random-access slot	54
N_{PTmax}	Maximal number of preamble transmission	10
N_{RAR}	Maximal number of RAR that can be carried in a response message	3
T_{CR}	Contention resolution timer (unit: sub-frame)	48
W_{RAR}	Length of the random-access response window (unit: sub-frame)	5
BI	Backoff indicator (unit: sub-frame)	20
p_f	HARQ transmissions for Msg3 and Msg4	10%
N_{HARQ}	Maximal number of HARQ transmission for Msg3 and Msg4	5
p_n	Preamble detection probability of the nth preamble transmission	$p_n = 1 - \dfrac{1}{e^n}$
T_{RAR}	Processing time required by a BS to detect the transmitted preamble (unit: sub-frame)	2
T_{HARQ}	Time interval required for receiving HARQ ACK (unit: sub-frame)	4
T_{M3}	Gap of Msg3 retransmission (unit: sub-frame)	4
$T_{A\,M4}$	Gap of monitor Msg4 (unit: sub-frame)	1
T_{M4}	Gap of Msg4 retransmission (unit: sub-frame)	1

TABLE II
VARIABLES USED IN THE ANALYTICAL MODEL

Notations	Variables	Setting
M_i	Total number of MTC devices that transmit their preambles in the ith random-access slot	Eq. (3)
$M_i[n]$	Number of contending MTC devices that transmit the nth preambles at the ith random-access slot	Eq. (5)
$M_{i,S}[n]$	Number of success MTC devices that transmit the nth preambles at the ith random-access slot and successfully finish the preamble transmission	Eq. (4)
$M_{i,F}[n]$	Number of failed MTC devices that transmit the nth preambles at the ith random-access slot but do not finish the preamble transmission	Eq. (5)
$P_{e,MSG}$	Error probability in transmitting the message part	Eq. (7)
$\alpha_{k,i}$	The portion of the backoff interval of the kth random-access slot ($K_{min} \le k \le K_{max} < i$) that overlaps with the transmission interval of the ith random-access slot	Eq. (10)
$\beta_{j,i}$	The portion of the MTC devices, which successfully transmit preambles in jth random-access slot ($J_{min} \le j \le J_{max} < i$), fail in message part transmission, and retransmit preambles in the ith random-access slot.	N/A
I_{max}	Number of random-access slots reserved for group paging	Eq. (2)
T_i	The access delay of the MTC devices which transmit preambles at ith random-access slot and successfully finish the preamble and message part transmission	Eq. (16)
$\overline{T_{MSG}}$	The average time required by an MTC device to successfully transmit the message part	Eq. (17)
P_C	The ratio between the number of occurrences when two or more MTC devices send random-access attempts with the same preamble and at the same frequency band and the overall number of RAOs (with or without random-access attempts) provided by the BS in the paging access interval in each cell	Eq. (13)
P_s	The probability to successfully complete the random-access procedure within the maximal number of preamble transmissions	Eq. (14)
$F(m)$	The CDF of the number of preamble transmissions to perform a random-access procedure for the successfully accessed MTC devices	Eq. (15)
$G(d)$	The CDF of the access delay	Eq. (18)
$\overline{D_a}$	The average access delay for the successfully accessed MTC devices	Eq. (19)
U	The ratio between the total number of successfully accessed MTC devices in the paging access interval and the total number of RAOs provided by the BS in the paging access interval in each cell.	Eq. (20)

Let R be the number of RAOs reserved by the BS in each random-access slot, N_{RAR} be the maximum number of RARs that can be carried in a response message, N_{UL} be the maximum number of MTC devices that can be acknowledged within the RAR window ($N_{\mathrm{UL}} = N_{\mathrm{RAR}} \times W_{\mathrm{RAR}}$), p_f be the HARQ retransmission probability for Msg3 and Msg4, and p_n be the preamble detection probability of the nth preamble transmission ($1 \le n \le N_{\mathrm{PTmax}}$). In LTE, p_n is expressed as [2]

$$p_n = 1 - \frac{1}{e^n}, \tag{1}$$

which models the power-ramping effect. Table I summarizes the random-access related system parameters defined in [2] and used in this paper.

When the group-paging message is received, all MTC devices should transmit their first preambles at the first random-access slot. The MTC devices should perform backoff and retransmit a new preamble up to ($N_{\mathrm{PTmax}} - 1$) times if the random access fail. For each preamble transmission, each MTC device may spend up to ($T_{\mathrm{RAR}} + W_{\mathrm{RAR}} + W_{\mathrm{BO}}$) subframes before retransmitting a new preamble. Hence, the number of random-access slots reserved for group paging (I_{max}) is expressed as

$$I_{\mathrm{max}} = 1 + (N_{\mathrm{PTmax}} - 1) \times \left\lceil \frac{T_{\mathrm{RAR}} + W_{\mathrm{RAR}} + W_{\mathrm{BO}}}{T_{\mathrm{RA\_REP}}} \right\rceil. \tag{2}$$

The paging access interval of group paging starts from the first random-access slot and ends at the I_{max}th random-access slot. In other words, the length of the paging access interval is $1 + (I_{\mathrm{max}} - 1) \times T_{\mathrm{RA\_REP}}$ subframes.

B. Analytical Model

In the following, an analytical model is presented to estimate the numbers of contending, success, and failed MTC devices in each random-access slot during a paging access interval. The performance metrics of group paging are then derived based on the estimated number of MTC devices. Table II summarizes the variables to be used in the proposed analytical model. Let

$M_i[n]$ be the number of contending MTC devices that transmit the nth preambles at the ith random-access slot, $M_{i,S}[n]$ be the number of success MTC devices that transmit the nth preambles at the ith random-access slot and successfully finish the preamble transmission (that is, the preambles are not collided, detected by the BS, and indicated in RARs), and $M_{i,F}[n]$ be the number of failed MTC devices that transmit the nth preambles at the ith random-access slot but do not finish the preamble transmission (that is, the preambles are collided; not collided and not detected by the BS; or not collided, detected by the BS, and not indicated in RARs). Let M_i be the total number of MTC devices that transmit their preambles in the ith random-access slot, which is the summation of all contending MTC devices. That is,

$$M_i = \sum_{n=1}^{N_{\mathrm{PTmax}}} M_i[n]. \tag{3}$$

In [1], we showed that the number of MTC devices with preamble transmissions that have not collided for M_i MTC devices and R RAOs at the ith random-access slot can be approximated by $M_i e^{-M_i/R}$. Among these, $M_i[n]/M_i$ of the MTC devices transmit their nth preambles that are detected with a probability of p_n. Therefore, $M_i[n]e^{-M_i/R}p_n$ MTC devices are detected. All of the detected MTC devices can receive the acknowledgment messages if the total number of detected MTC devices does not exceed N_{UL} (i.e., $\sum_{n=1}^{N_{\mathrm{PTmax}}} M_i[n]e^{-M_i/R}p_n \le N_{\mathrm{UL}}$). Otherwise, the BS randomly sends the acknowledgment messages to N_{UL} detected MTC devices. In this case, the number of MTC devices that can receive the acknowledgment messages is proportional to the number of detected MTC devices that belong to the same category. Hence, $M_{i,S}[n]$ can be determined as

$$M_{i,S}[n] = \begin{cases} M_i[n]e^{-\frac{M_i}{R}}p_n, & \text{if } \sum_{n=1}^{N_{\mathrm{PTmax}}} M_i[n] \\ & \times e^{-\frac{M_i}{R}}p_n \le N_{\mathrm{UL}} \\ \dfrac{M_i[n]e^{-\frac{M_i}{R}}p_n}{\sum\limits_{n=1}^{N_{\mathrm{PTmax}}} M_i[n]e^{-\frac{M_i}{R}}p_n} N_{\mathrm{UL}}, & \text{otherwise.} \end{cases} \tag{4}$$

The number of failed MTC devices is equal to the number of contending MTC devices minus the number of success MTC devices. That is

$$M_{i,F}[n]$$
$$= M_i[n] - M_{i,S}[n]$$
$$= \begin{cases} M_i[n]\left(1 - e^{-\frac{M_i}{R}}p_n\right), & \text{if } \sum_{n=1}^{N_{\mathrm{PTmax}}} M_i[n]e^{-\frac{M_i}{R}} \\ & \times p_n \le N_{\mathrm{UL}} \\ M_i[n]\left(1 - \dfrac{p_n}{\sum\limits_{n=1}^{N_{\mathrm{PTmax}}} M_i[n]p_n}\right) \\ \times N_{\mathrm{UL}}, & \text{otherwise.} \end{cases} \tag{5}$$

Note that $M_{i,S}[n]$ is the number of success MTC devices that transmit the nth preambles at the ith random-access slot and successfully finish the preamble transmission. These MTC devices will immediately transmit the messages through the dedicated channel indicated in the RAR.

The number of MTC devices that transmit the nth preambles at the ith random-access slot, i.e., $M_i[n]$, is expressed as

$$M_i[n] = \sum_{k=K_{\min}}^{K_{\max}} \alpha_{k,i} M_{k,F}[n-1]$$
$$+ \sum_{j=J_{\min}}^{J_{\max}} \beta_{j,i} p_{e,MSG} M_{j,S}[n-1]$$
$$\approx \sum_{k=K_{\min}}^{K_{\max}} \alpha_{k,i} M_{k,F}[n-1]. \tag{6}$$

Equation (6) comprises two parts: MTC devices with preamble transmissions that failed and those with message transmissions that failed. The first part of (6) ($M_{k,F}[n-1]$) represents a situation where the MTC devices transmit the $(n-1)$th preambles at the kth random-access slot, but they fail to complete the preamble transmission. Among these MTC devices with preamble transmissions that failed, $a_{k,i}$ of them will perform random backoff and retransmit the nth preambles at the ith random-access slot. $K_{\min}$ and $K_{\max}$ denote the minimal and maximal values of k, respectively. Hence, we accumulate the possible cases of k from $K_{\min}$ to $K_{\max}$ to obtain the number of MTC devices with preamble transmissions that failed. The second part of (6) ($P_{e,MSG} M_{j,S}[n-1]$) denotes a situation where the MTC devices transmit the $(n-1)$th preambles at the jth random-access slot, finish the preamble transmission, and fail in messages part (Msg 3 and Msg 4) transmission at error probability $P_{e,MSG}$. Among these MTC devices with message transmissions that failed, $\beta_{j,i}$ of them will perform random backoff and retransmit the nth preambles at the ith random-access slot. $J_{\min}$ and $J_{\max}$ denote the minimal and maximal values of j, respectively. Hence, we accumulate the possible cases of J from $J_{\min}$ to $J_{\max}$ to obtain the number of MTC devices with preamble transmissions that failed. Note that message transmission is failed if Mgs3 transmission exceeds N_{HARQ} times, or Msg3 transmission is a success but Mgs4 transmission exceeds N_{HARQ} times. Therefore, $P_{e,MSG}$ can be derived as

$$P_{e,MSG} = p_f^{N_{\mathrm{HARQ}}} + \sum_{j=0}^{N_{\mathrm{HARQ}}-1} p_f^j(1-p_f)p_f^{N_{\mathrm{HARQ}}}. \tag{7}$$

In LTE, N_{HARQ} is a constant of 5 [17], $p_f = 0.1$ is assumed in [2], and $\beta_{j,i} \le 1$, which obtains a small value for $P_{e,MSG}$. Hence, the second term in (6) is negligible.

$a_{k,i}$, $K_{\min}$, and $K_{\max}$ in (6) can be derived based on the timing diagram given in Fig. 1. The MTC device, in which preamble transmission is failed at the kth random-access slot, may retransmit a new preamble at the ith random-access slot only if the *backoff interval* of the kth random-access slot is overlapped with the *transmission interval* of the ith

random-access slot. Therefore, $a_{k,i}$ is the portion of the backoff interval of the kth random-access slot that overlaps with the transmission interval of the ith random-access slot ($k < i$). As shown in Fig. 1, the MTC devices that transmit their preambles at the kth random-access slot at time $(k-1) \times T_{\text{RA\_REP}}$ will recognize their random-access failure after $(T_{\text{RAR}} + W_{\text{RAR}})$ subframes. Each failed MTC device starts random backoff at time $(k-1) \times T_{\text{RA\_REP}} + (T_{\text{RAR}} + W_{\text{RAR}}) + 1$. Therefore, the backoff interval of the kth random-access slot starts from time $(k-1) \times T_{\text{RA\_REP}} + T_{\text{RAR}} + W_{\text{RAR}} + 1$ and ends at time $(k-1) \times T_{\text{RA\_REP}} + T_{\text{RAR}} + W_{\text{RAR}} + W_{\text{BO}}$. The MTC devices transmit their preambles at the ith random-access slot if their backoff counters reach zero during the interval between the $(i-1)$th random-access slot and the ith random-access slot. Therefore, the transmission interval of the ith random-access slot is $[(i-2) \times T_{\text{RA\_REP}} + 1, (i-1) \times T_{\text{RA\_REP}}]$.

The minimal value of k ($K_{\min}$) is obtained when the right boundary of the kth random-access slot backoff interval reaches the left boundary of the ith random-access slot transmission interval (i.e., $(K_{\min}-1) \times T_{\text{RA\_REP}} + T_{\text{RAR}} + W_{\text{RAR}} + W_{\text{BO}} \geq 1 + (i-2) \times T_{\text{RA\_REP}}$). Therefore, $K_{\min}$ is expressed as

$$K_{\min} = \left\lceil (i-1) + \frac{1 - (T_{\text{RAR}} + W_{\text{RAR}} + W_{\text{BO}})}{T_{\text{RA\_REP}}} \right\rceil. \quad (8)$$

The maximal value of k ($K_{\max}$) is obtained when the left boundary of the kth random-access slot backoff interval exceeds the right boundary of the ith random-access slot transmission interval (i.e., $(K_{\max}-1) \times T_{\text{RA\_REP}} + T_{\text{RAR}} + W_{\text{RAR}} + 1 \leq (i-1) \times T_{\text{RA\_REP}}$). Hence

$$K_{\max} = \left\lfloor i - \frac{T_{\text{RAR}} + W_{\text{RAR}} + 1}{T_{\text{RA\_REP}}} \right\rfloor. \quad (9)$$

$a_{k,i}$ can be determined based on k in three cases shown on the lower part in Fig. 1. In the first case, the right boundary of the backoff interval is within the transmission interval (i.e., $(i-2) \times T_{\text{RA\_REP}} + 1 \leq (k-1) \times T_{\text{RA\_REP}} + T_{\text{RAR}} + W_{\text{RAR}} + W_{\text{BO}} \leq (i-1) \times T_{\text{RA\_REP}}$). In this case, $K_{\min} \leq k \leq i - (T_{\text{RAR}} + W_{\text{RAR}} + W_{\text{BO}})/T_{\text{RA\_REP}}$ and the overlapped region start from the left boundary of the transmission interval and end at the right boundary of the backoff interval. In the second case, the transmission interval is fully overlapped with the backoff interval; thus, the length of the overlapped region is $T_{\text{RA\_REP}}$. In the third case, the left boundary of the backoff interval is within the transmission interval (i.e., $(i-2) \times T_{\text{RA\_REP}} + 1 \leq (k-1) \times T_{\text{RA\_REP}} + T_{\text{RAR}} + W_{\text{RAR}} + 1 \leq (i-1) \times T_{\text{RA\_REP}}$). In this case, $(i-$

$1) - (T_{\text{RAR}} + W_{\text{RAR}})/T_{\text{RA\_REP}} \leq k \leq K_{\max}$ and the overlapped region start from the left boundary of the backoff interval and end at the right boundary of the transmission interval. $a_{k,i}$ is the ratio between the overlapped region and the backoff interval and is expressed as in (10), shown at the bottom of the page.

Initially, all of the M/K MTC devices transmit their first preambles at the first random-access slot. Hence, the initial conditions can be set by $M_1 = M_1[1] = M/K$ and $M_1[n] = 0$ for $n \neq 1$. Let $i = 1$ in (4); therefore, we derive

$$M_{1,S}[n] = \begin{cases} N_{\text{UL}}, & \text{if } n=1 \text{ and } \frac{M}{K}e^{-\frac{M/K}{R}}p_1 \geq N_{\text{UL}} \\ \frac{M}{K}e^{-\frac{M/k}{R}}p_1, & \text{if } n=1 \text{ and } \frac{M}{K}e^{-\frac{M/K}{R}}p_1 < N_{\text{UL}} \\ 0, & \text{if } n \neq 1. \end{cases} \quad (11)$$

Let $i = 1$ in (5), we obtain

$$M_{1,F}[n]$$
$$= \begin{cases} \frac{M}{K} - N_{\text{UL}}, & \text{if } n=1 \text{ and } \frac{M}{K}e^{-\frac{M/K}{R}}p_1 \geq N_{\text{UL}} \\ \frac{M}{K}\left(1 - e^{-\frac{M/K}{R}}p_1\right), & \text{if } n=1 \text{ and } \frac{M}{K}e^{-\frac{M/K}{R}}p_1 < N_{\text{UL}} \\ 0, & \text{if } n \neq 1. \end{cases} \quad (12)$$

For $i \geq 2$, $M_{i,S}[n]$, $M_{i,F}[n]$, and $M_i[n]$ can be recursively obtained from (4)–(6), respectively.

C. Performance Metrics

The collision probability, access success probability, statistics of number of preamble transmissions, statistics of access delay, average access delay, and the utilization of RAOs obtained in a paging access interval are chosen as the performance metrics in this paper to evaluate the performance of group paging. The collision probability, i.e., P_C, is defined as the ratio between the number of occurrences when two or more MTC devices send random-access attempts with the same preamble and at the same frequency band and the overall number of RAOs (with or without random-access attempts) reserved by the BS [2]. That is, P_C is the ratio between the total number of collided RAOs and the total number of reserved RAOs. In each random-access slot, the number of collided RAOs is equal to the number of reserved RAOs (R) minus the number of success RAOs ($M_i e^{-M_i/R}$) and the number of idle RAOs. In [1], it is shown that the number of idle RAOs at the ith random-access slot can be obtained from a one-shot random access with M_i MTC devices contending for R RAOs and can be approximated by $Re^{-M_i/R}$. The total number of collided RAOs in a paging

$$\alpha_{k,i} = \begin{cases} \frac{(k-1) \times T_{\text{RA\_REP}} + T_{\text{RAR}} + W_{\text{RAR}} + W_{\text{BO}} - (i-2) \times T_{\text{RA\_REP}}}{W_{\text{BO}}}, & \text{if } K_{\min} \leq k \leq i - \frac{T_{\text{RAR}} + W_{\text{RAR}} + W_{\text{BO}}}{T_{\text{RA\_REP}}} \\ \frac{T_{\text{RA\_REP}}}{W_{\text{BO}}}, & \text{if } i - \frac{T_{\text{RAR}} + W_{\text{RAR}} + W_{\text{BO}}}{T_{\text{RA\_REP}}} < k < (i-1) - \frac{T_{\text{RAR}} + W_{\text{RAR}}}{T_{\text{RA\_REP}}} \\ \frac{(i-1) \times T_{\text{RA\_REP}} - ((k-1) \times T_{\text{RA\_REP}} + T_{\text{RAR}} + W_{\text{RAR}})}{W_{\text{BO}}}, & \text{if } (i-1) - \frac{T_{\text{RAR}} + W_{\text{RAR}}}{T_{\text{RA\_REP}}} \leq k \leq K_{\max} \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

access interval is obtained from the sum of the number of collided RAOs from the first random-access slot to the $I_{\max}$th random-access slot. Hence, P_C is expressed as

$$P_C = \frac{\sum_{i=1}^{I_{\max}} \left(R - M_i e^{-\frac{M_i}{R}} - R e^{-\frac{M_i}{R}} \right)}{I_{\max} R}. \tag{13}$$

The access success probability, i.e., P_S, is the probability to successfully complete the random-access procedure within the maximum number of preamble transmissions [2]. That is, P_S is the ratio between the total number of successfully accessed MTC devices and the average group size in a cell. The number of successfully accessed MTC devices that transmit the nth preambles at the ith random-access slot is equal to $M_{i,S}[n](1 - P_{e,MSG})$. The total number of successfully accessed MTC devices is the sum of successfully accessed MTC devices in all random-access slots. Hence, P_S is expressed as

$$P_S = \frac{\sum_{i=1}^{I_{\max}} \sum_{n=1}^{N_{\mathrm{PTmax}}} M_{i,S}[n](1 - p_{e,MSG})}{M/K}$$

$$\approx \frac{K \sum_{i=1}^{I_{\max}} \sum_{n=1}^{N_{\mathrm{PTmax}}} M_{i,S}[n]}{M}. \tag{14}$$

Let m be the number of preamble transmissions to perform a random-access procedure for the successfully accessed MTC devices. The "statistics of number of preamble transmission" $(F(m))$ is the cumulative distribution function (cdf) of m [2]. $F(m)$ is the ratio between the number of successfully accessed MTC devices that transmit no more than m preambles and the total number of successfully accessed MTC devices. Hence, $F(m)$ is expressed as

$$F(m) = \frac{\sum_{i=1}^{I_{\max}} \sum_{k=1}^{m} M_{i,S}[k](1 - p_{e,MSG})}{\sum_{i=1}^{I_{\max}} \sum_{m=1}^{N_{\mathrm{PTmax}}} M_{i,S}[m](1 - p_{e,MSG})}$$

$$= \frac{\sum_{i=1}^{I_{\max}} \sum_{k=1}^{m} M_{i,S}[k]}{\sum_{i=1}^{I_{\max}} \sum_{m=1}^{N_{\mathrm{PTmax}}} M_{i,S}[m]}. \tag{15}$$

Let d be the access delay for each random-access procedure between the first random-access attempt and the completion of the random-access procedure for the successfully accessed MTC devices. Let $G(d)$ be the cdf of access delay, which is the ratio between the number of successfully accessed MTC devices observed before time d and the total number of successfully accessed MTC devices [2]. Let T_i be the access delay of MTC devices that transmit preambles at the ith random-access slot and successfully complete the preamble and message transmission. T_i comprises the time required to transmit the preamble (i.e., $(i-1)T_{\mathrm{RA\_REP}}$), receive the RAR (i.e., $T_{\mathrm{RAR}} + W_{\mathrm{RAR}}$), and transmit the message. That is

$$T_i = (i-1)T_{\mathrm{RA\_REP}} + (T_{\mathrm{RAR}} + W_{\mathrm{RAR}}) + \overline{T_{\mathrm{MSG}}} \tag{16}$$

where $\overline{T_{\mathrm{MSG}}}$ is the average time required by an MTC device to successfully transmit the message. Consider an MTC device that successfully transmits the message using u HARQ transmissions of Msg3 and v HARQ transmissions of Msg4 ($u \le N_{\mathrm{HARQ}}, v \le N_{\mathrm{HARQ}}$). The time required to transmit u Msg3 and v Msg4 is equal to $1 + (u-1)(T_{\mathrm{HARQ}} + T_{M3}) + T_{\mathrm{HARQ}}$ and $T_{A\_M4} + (v-1)(T_{\mathrm{HARQ}} + T_{M4}) + T_{\mathrm{HARQ}}$, respectively. The probability that an MTC device that successfully transmits the message using u HARQ transmissions of Msg3 and v HARQ transmissions of Msg4 is $(P_f)^{u-1}(1 - P_f)(P_f)^{v-1}(1 - P_f)$. Hence, $\overline{T_{\mathrm{MSG}}}$ is expressed as

$$\overline{T_{\mathrm{MSG}}} = \sum_{u=1}^{N_{\mathrm{HARQ}}} \sum_{v=1}^{N_{\mathrm{HARQ}}} P_f^{u-1}(1 - P_f) P_f^{u-1}(1 - P_f)$$
$$\times \big((1 + (u-1)(T_{M3} + T_{\mathrm{HARQ}}) + T_{\mathrm{HARQ}})$$
$$+ (T_{A\_M4} + (v-1)(T_{M4} + T_{\mathrm{HARQ}}) + T_{\mathrm{HARQ}})\big)$$
$$= \sum_{u=1}^{N_{\mathrm{HARQ}}} \sum_{v=1}^{N_{\mathrm{HARQ}}} (P_f)^{u+v-2}(1 - P_f)^2$$
$$\times \big((u-1)T_{M3} + (v-1)T_{M4}$$
$$+ (u+v)T_{\mathrm{HARQ}} + T_{A\_M4} + 1\big). \tag{17}$$

For the successfully accessed MTC devices that complete their random-access procedure at time d, the average time they complete the preamble transmission is $d - (T_{\mathrm{RAR}} + W_{\mathrm{RAR}} + \overline{T_{\mathrm{MSG}}})$ or the $(\lfloor d - (T_{\mathrm{RAR}} + W_{\mathrm{RAR}} + \overline{T_{\mathrm{MSG}}})/T_{\mathrm{RA\_REP}} \rfloor + 1)$th random-access slot. $G(d)$ is estimated by

$$G(d)$$
$$= \frac{\sum_{i=1}^{\left\lfloor \frac{d - (T_{\mathrm{RAR}} + W_{\mathrm{RAR}} + \overline{T_{\mathrm{MSG}}})}{T_{\mathrm{RA\_REP}}} \right\rfloor + 1} \sum_{n=1}^{N_{\mathrm{PTmax}}} M_{i,S}[n](1 - p_{e,MSG})}{\sum_{i=1}^{I_{\max}} \sum_{n=1}^{N_{\mathrm{PTmax}}} M_{i,S}[n](1 - p_{e,MSG})}$$
$$= \frac{\sum_{i=1}^{\left\lfloor \frac{d - (T_{\mathrm{RAR}} + W_{\mathrm{RAR}} + \overline{T_{\mathrm{MSG}}})}{T_{\mathrm{RA\_REP}}} \right\rfloor + 1} \sum_{n=1}^{N_{\mathrm{PTmax}}} M_{i,S}[n]}{\sum_{i=1}^{I_{\max}} \sum_{n=1}^{N_{\mathrm{PTmax}}} M_{i,S}[n]}. \tag{18}$$

Let $\overline{D_a}$ be the average access delay for the successfully accessed MTC devices (unit: subframe). $\overline{D_a}$ is the total access delay for all successfully accessed MTC devices normalized by the total number of successfully accessed MTC devices. The total access delay for the successfully accessed MTC devices that transmit the last preambles at the ith random-access slot is $\sum_{n=1}^{N_{\mathrm{PTmax}}} M_{i,S}[n](1 - p_{e,MSG})T_i$. Hence, $\overline{D_a}$ is expressed as

$$\overline{D_a} = \frac{\sum_{i=1}^{I_{\max}} \sum_{n=1}^{N_{\mathrm{PTmax}}} M_{i,S}[n](1 - p_{e,MSG})T_i}{\sum_{i=1}^{I_{\max}} \sum_{n=1}^{N_{\mathrm{PTmax}}} M_{i,S}[n](1 - p_{e,MSG})}$$
$$= \frac{\sum_{i=1}^{I_{\max}} \sum_{n=1}^{N_{\mathrm{PTmax}}} M_{i,S}[n]T_i}{\sum_{i=1}^{I_{\max}} \sum_{n=1}^{N_{\mathrm{PTmax}}} M_{i,S}[n]}. \tag{19}$$

Finally, the utilization of RAOs (U) is defined as the ratio between the total number of successfully accessed MTC devices and the total number of reserved RAOs. The total number of successfully accessed MTC devices in the ith random-access slot is $\sum_{n=1}^{N_{\mathrm{PTmax}}} M_{i,S}[n](1-p_{e,MSG})$. U is expressed as

$$
\begin{aligned}
U &= \frac{\sum_{i=1}^{I_{\max}} \sum_{n=1}^{N_{\mathrm{PTmax}}} M_{i,S}[n](1-p_{e,MSG})}{I_{\max} R} \\
&\approx \frac{\sum_{i=1}^{I_{\max}} \sum_{n=1}^{N_{\mathrm{PTmax}}} M_{i,S}[n]}{I_{\max} R}.
\end{aligned}
\tag{20}
$$

III. NUMERICAL RESULTS

Computer simulations were conducted on top of a C-based simulator to verify the effectiveness of the proposed analytical model. The simulations were developed based on a Monte Carlo approach. In the simulations, each point represented the average value of 10^7 samples. We used a computer equipped with an Intel Core2 Quad central processing unit, a Q9500 2.83-GHz processor, and 6-GB random-access memory. The time used for calculating the performance metrics in all cases was less than 1 ms. Each sample was obtained by performing group paging in a paging access interval. In group paging, we may adjust design parameters of the average group size per cell (M/K), the reserved RAOs per random-access slot (R), the maximum number of preamble transmission (N_{PTmax}), and the backoff window size (W_{BO}) to meet the service quality constraints of the access success probability (P_S), the average access delay $(\overline{D_a})$, and/or the collision probability (P_C). For simplicity, W_{BO} was fixed in all of the simulations.

In the simulations, M/K MTC devices are assumed to initiate all random-access attempts. The background traffic caused by Poisson-distributed H2H UEs at an average rate (λ) of 0, 8, and 128 calls/s is considered [2]. The RACH parameters defined in [2, Tab. 2.2.1.1] and the processing latency specified in Table B.1.1.1-1 of TR 36.912 [20] were used as a baseline in the simulations. The settings of the parameters used in the simulation are summarized in Table I.

Two scenarios were investigated. *Scenario I* was used to verify the accuracy of the analytical model. We considered a case where a BS reserves 54 RAOs in each random-access slot $(R = 54)$ to page a group size of 10–1000 MTC devices per cell $(M/K = 10$ to $1000)$ with or without H2H traffic, and the results are shown in Figs. 3–7. Different values of R were investigated, but the results are not shown in the figures. The results show that the proposed analytical model can accurately estimate the performance metrics of group paging for various combinations of M/K, R, and N_{PTmax}. *Scenario II* was designed to study the effect of the design parameters and the tradeoffs among design strategies according to the given service quality constraints. Based on this study, a BS can properly select the design parameters of R and N_{PTmax} to minimize the total number of reserved RAOs according to the target group size (M/K) and the specified service quality constraints (P_S). The results are shown in Figs. 8–12.
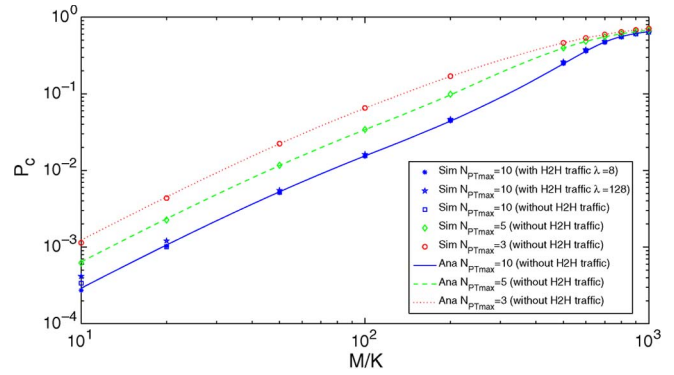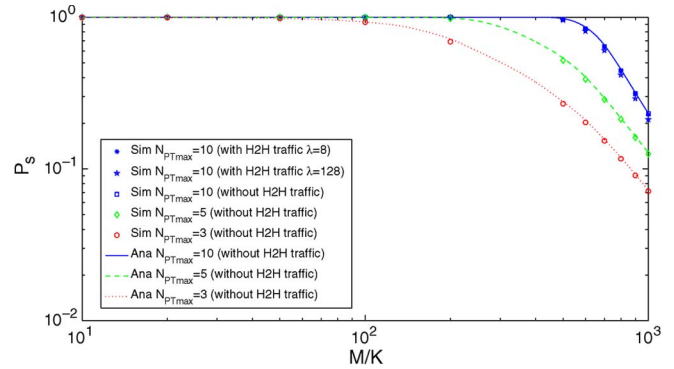

Fig. 3. Collision probability.
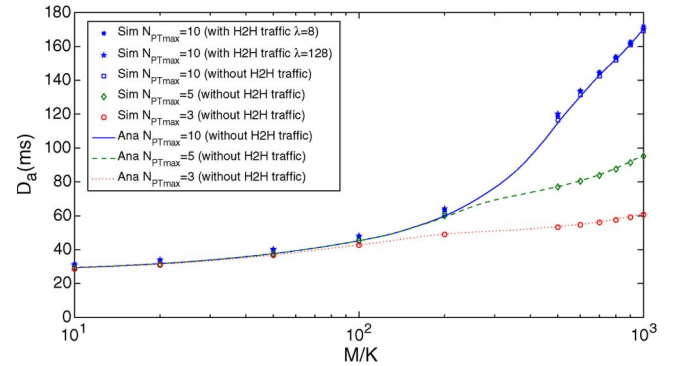

Fig. 4. Access success probability.

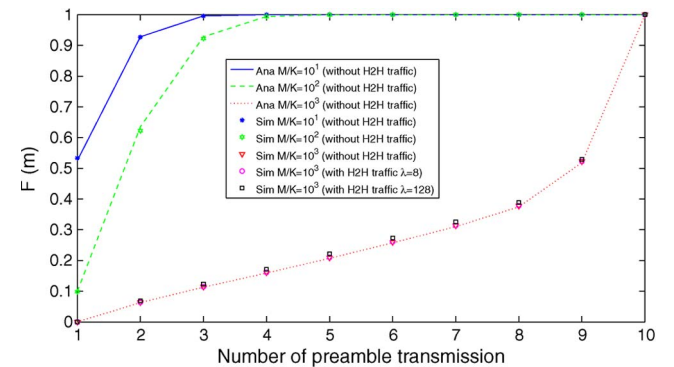Fig. 5. Average access delay.

Fig. 6. CDF of preamble transmission.

In *Scenario I*, the analytical and simulation results of P_C, P_S, $\overline{D_a}$, $F(m)$, and $G(d)$ are shown in Figs. 3–7, respectively. Symbols and lines in Figs. 3–7 are used to present simulation
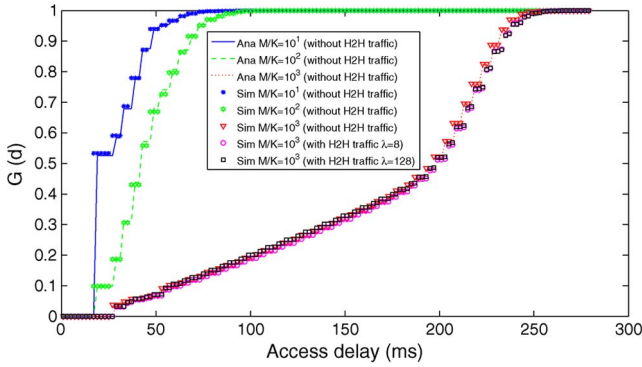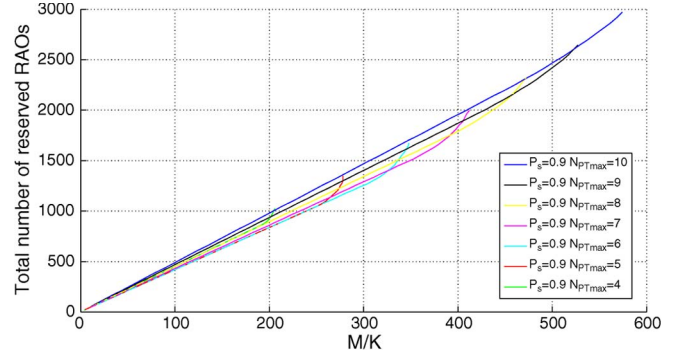
Fig. 7.   CDF of access delay.

Fig. 8.   Utilization of RAOs.

Fig. 9.   Surface indicating the feasible region of R and M/K to ensure $P_S = 0.9$ for $N_{\mathrm{PTmax}} = 10$.

Fig. 10.   Optimal edges of the feasible regions for various N_{PTmax} with $P_S = 0.9$.

and analytical results, respectively. The analytical results of P_C, P_S, $\overline{D_a}$, $F(m)$, and $G(d)$ were obtained based on (13), (14), (19), (15), and (18), respectively. In Figs. 3–5, various N_{PTmax} values were evaluated. In Figs. 6 and 7, $N_{\mathrm{PTmax}} = 10$ was used. In Figs. 3–7, all of the analytical results coincided with the

Fig. 11.   Optimal edges of the feasible regions indicating the required RAOs for various M/K with $P_S = 0.9$.
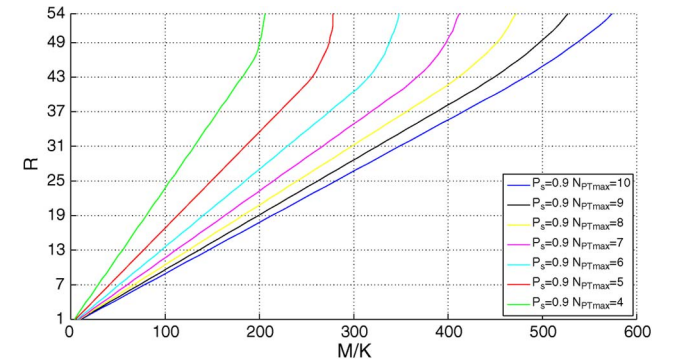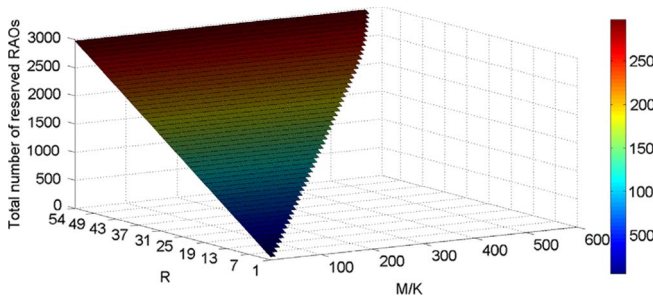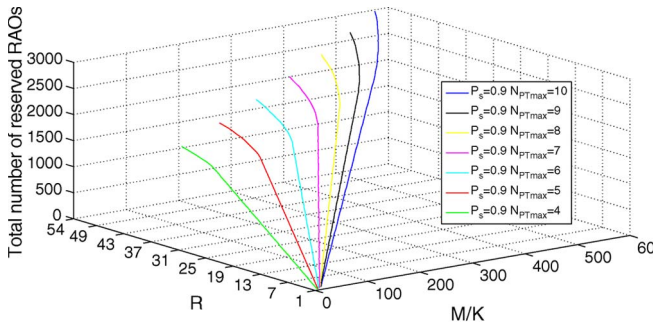

Fig. 12.   Optimal edges of the feasible regions indicating the R for various M/K with $P_S = 0.9$.

simulation results, which verified the accuracy of the analysis. Fig. 3 shows that collision probability P_C can be decreased by increasing N_{PTmax}. Fig. 4 shows access success probability P_S as a function of average group size M/K. As shown in Fig. 4, the BS can support a group size of up to 400 per cell ($M/K = 400$), and all of the MTC devices successfully completed the random-access procedure ($P_S = 1$) if $N_{\mathrm{PTmax}} = 10$. However, the group size reduced to 10 ($M/K = 10$) for a target access success probability of $P_S = 1$ if $N_{\mathrm{PTmax}} = 3$. Fig. 5 shows the average access delay for the successfully accessed MTC devices, i.e., $\overline{D_a}$. For smaller group size ($M/K \leq 50$), the average access delay is almost identical for different values of N_{PTmax}. As the group size increases, a smaller N_{PTmax} results in a lower average access delay because less MTC devices can finish their random-access procedures. Fig. 6 shows the cdf of the number of preamble transmissions, i.e., $F(m)$, for the successfully accessed MTC devices. The number of preamble transmissions increases in conjunction with the number of MTC devices per cell. In Fig. 6, the cdf value that is equal to 1 does not imply that all of the MTC devices complete the random-access procedure because only successfully accessed MTC devices are considered. Fig. 7 shows the cdf of access delay $G(d)$ for the successfully accessed MTC devices. Similarly, access delay increases in conjunction with the number of MTC devices per cell. The proposed analytical model maintains system validity at an extremely high H2H rate of 128 calls/s; thus, our assumption that the effect of H2H traffic can be ignored in the analysis is supported. The difference between the simulation and analytical results shown in Figs. 3–7 is because

of the error of the approximation model proposed in [1] that has been adopted to estimate the number of successful and collided devices in each random-access slot.

In *Scenario II*, the analytical results of group paging for various design parameters are shown in Figs. 8–12. Fig. 8 shows the utilization of RAOs as a function of R and M/K for $N_{\text{PTmax}} = 10$. This figure can be used to obtain the proper values of R and M/K to maximize U. The maximal value of U, as shown in Fig. 8, is approximately 0.2, which implies that, in the optimal case, a successfully accessed MTC device requires five RAOs to complete the random-access procedure. For a fixed value of M/K, U first increases and subsequently decreases when R increases. For smaller values of R, the utilization of RAOs (U) is low because only a small number of MTC devices can complete the random-access procedure due to high collisions. Therefore, U subsequently increases with R because the number of successful access MTC devices quickly increases. U approaches its maximal value when R reaches a specific value. A further increase in R results in a decreased value of U because more RAOs are reserved, but they are not used. Similarly, for a fixed value of R, U first increases and subsequently decreases when M/K increases. U increases in conjunction with M/K because the number of successful access MTC devices increases. U decreases after M/K exceeds a certain value because the number of successful access MTC devices rapidly decreases because of collision.

Fig. 9 shows a surface indicating the feasible region of the total number of reserved RAOs to ensure a 0.9 access success probability ($P_S = 0.9$) for various R and M/K with $N_{\text{PTmax}} = 10$. The total number of reserved RAOs is equal to $R \times I_{\max}$ and is proportional to R and N_{PTmax}. From (2) and (19), access delay is proportional to N_{PTmax}. Hence, a higher N_{PTmax} results in an increased access delay for the MTC devices. In general, the number of required RAOs is increased if R or M/K is increased. The minimal value of R required to support the maximal value of M/K is located at the right-hand side edge of the surface, as shown in Fig. 9, which is referred to as the *optimal edge*. The optimal edges of the feasible regions for various N_{PTmax} with $P_S = 0.9$ are shown in Fig. 10. Fig. 10 can be viewed as the 2-D figures shown in Figs. 11 and 12, which show the optimal edges of the feasible regions indicating the total number of reserved RAOs and R for various M/K to ensure $P_S = 0.9$, respectively. The optimal value of N_{PTmax} that minimizes the total number of reserved RAOs can be obtained in Fig. 11. A smaller value of N_{PTmax} will be chosen if several N_{PTmax} have the same total number of reserved RAOs. It is because access delay can be reduced if a smaller N_{PTmax} is chosen. In Fig. 11, it is shown that $N_{\text{PTmax}} = 4$ for M/K smaller than 7, $N_{\text{PTmax}} = 5$ is a good choice for M/K ranging from 7 to 250, $N_{\text{PTmax}} = 6$ is the appropriate choice for M/K ranging from 250 to 325, $N_{\text{PTmax}} = 7$ is a favorable option for M/K ranging from 325 to 390, $N_{\text{PTmax}} = 8$ is the efficient setting for M/K ranging from 390 to 458, $N_{\text{PTmax}} = 9$ is preferred for M/K ranging from 458 to 522, and $N_{\text{PTmax}} = 10$ should be set for M/K higher than 522. The optimal value of R can be subsequently obtained in Fig. 12 based on the target group size M/K and the optimal value of N_{PTmax} obtained in Fig. 11. Fig. 12

also shows that several N_{PTmax} may attain the same minimal total number of reserved RAOs. The price paid for choosing a smaller N_{PTmax} is to reserve a larger R for each random-access slot. The effect is similar to that pointed out in [13]. Consider the case in which the network reserves a number of preambles (R) to page a given workload (M/K) by using the same N_{PTmax}. Each group of MTC devices will be paged during a paging access interval of $I_{\max} \times T_{\text{RA\_REP}}$ subframes. For a network with multiple paging groups, each paging group must be paged in a paging access interval that is not overlapped with the others. The network may merge multiple paging groups into a new paging group to reduce the overall access delay. The network may page the new paging group consisting of H identical paging groups with a group size of $H \times M/K$ MTC devices/cell by using $H \times R$ preambles. As shown in Fig. 10, the number of reserved preambles is proportional to the offered workload for all N_{PTmax} until the workload reaches a certain value.

IV. CONCLUSION

Group paging is one of the approaches proposed for alleviating the RAN-overload problem resulting from the simultaneous access of mass MTC devices. This paper has presented analytical models to estimate the collision probability, access success probability, average access delay, cdf of preamble transmissions, cdf of access delay, and utilization of RAOs of group paging in a paging access interval. All of the factors specified in the 3GPP evaluation methodology [2], such as the power-ramping effect, backoff window size, limited number of UL grants per RAR, and the HARQ of message transmissions, have been considered in the proposed model. Numerical results demonstrate that the proposed model can accurately predict the performance metrics. A possible approach to utilize the proposed model to determine the optimal group size and required radio resource based on the given target access success probability has been also presented. The results demonstrate that the tradeoffs and design strategies of group paging can be achieved by properly adjusting the design parameters for various paging cells (K) and workload (M). The analytical model is applicable to other RANs adopting a similar random-access procedure such as WiMAX and can be used by network operators to dimension the required resources in RACHs.

REFERENCES

[1] C. H. Wei, R. G. Cheng, and S. L. Tsao, "Modeling and estimation of one-shot random access for finite-user multichannel slotted ALOHA systems," *IEEE Commun. Lett.*, vol. 16, no. 8, pp. 1196–1199, Aug. 2012.

[2] "RAN improvements for machine-type communications," Third-Gen. Partnership Proj., Sophia-Antipolis Cedex, France, 3GPP TR 37.868, Aug. 2011.

[3] "Comparing push and pull based approaches for MTC," Third-Gen. Partnership Proj., Sophia-Antipolis Cedex, France, 3GPP R2-104873, Aug 2010, Inst. Inf. Ind. (III), Coiler Corp., RAN2#71.

[4] "Evaluation on push based RAN overload control schemes," Third-Gen. Partnership Proj., Sophia-Antipolis Cedex, France, 3GPP R2-112071, Apr. 2011, RAN2#73bis.

[5] "Pull Based RAN Overload Control," Third-Gen. Partnership Proj., Sophia-Antipolis Cedex, France, 3GPP R2-104870, Aug. 2010, Huawei and China Unicom, RAN2#71.

[6] "Further Analysis of Group Paging for MTC," Third-Gen. Partnership Proj., Sophia-Antipolis Cedex, France, 3GPP R2-113198, May 2011, ITRI, RAN2#74.

[7] J. C. Arnbak and W. Van Blitterswijk, "Capacity of slotted ALOHA in Rayleigh-fading channels," *IEEE J. Sel. Areas Commun.*, vol. 5, no. 2, pp. 261–269, Feb. 1987.

[8] L. Kleinrock and S. S. Lam, "Packet switching in a multi-access broadcast channel: Performance evaluation," *IEEE Trans. Commun.*, vol. COM-23, no. 4, pp. 410–423, Apr. 1975.

[9] I. Rubin, "Group random-access discipline for multi-access broadcast channels," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 5, pp. 578–592, Sep. 1978.

[10] H. H. Tan and H. Wang, "Performance of multiple parallel slotted ALOHA channels," in *Proc. INFOCOM*, Mar. 1987, pp. 931–940.

[11] V. Naware, G. Mergen, and L. Tong, "Stability and delay of finite user slotted ALOHA with multipacket reception," *IEEE Trans. Inf. Theory*, vol. 5, no. 7, pp. 2636–2656, Jul. 2005.

[12] P. R. Jelenkovic and J. Tan, "Stability of finite population Aloha with variable packets," Tech. Rep. EE2009-02-20, 2009, arXiv:0902.4481v2.

[13] Y. J. Choi, S. Park, and S. Bahk, "Multichannel random access in OFDMA wireless network," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 603–613, Mar. 2006.

[14] P. Zhou, H. Hu, H. Wang, and H. H. Chen, "An efficient random access scheme for OFDMA systems with implicit message transmission," *IEEE Trans. Wireless Commun.*, vol. 7, no. 7, pp. 2790–2797, Jul. 2008.

[15] L. Kleinrock and F. Tobagi, "Packet switching in radio channels, Part I—Carrier sense multiple-access modes and their throughput—Delay characteristics," *IEEE Trans. Commun.*, vol. COM-23, no. 12, pp. 1400–1416, Dec. 1975.

[16] R. G. Cheng, C. H. Wei, S. L. Tsao, and F. C. Ren, "RACH collision probability for machine-type communications," in *Proc. IEEE 75th Veh. Technol. Conf.*, May 2012, pp. 1–5.

[17] "Evolved universal terrestrial radio access (E-UTRA): Medium access control (MAC) protocol specification," Third-Gen. Partnership Proj., Sophia-Antipolis Cedex, France, 3GPP TS 36.321, Jun. 2010, ver. 9.3.0.

[18] "LTE Random-Access Capacity and Collision Probability," Third-Gen. Partnership Proj., Sophia-Antipolis Cedex, France, 3GPP R1-061369, May 2006, Ericsson, RAN1#45.

[19] "Evolved universal terrestrial radio access (E-UTRA): Physical channels and modulation," Third-Gen. Partnership Proj., Sophia-Antipolis Cedex, France, 3GPP TS 36.211, Jun. 2011, ver. 10.2.0.

[20] "Feasibility study for further advancements for E-UTRA (LTE-advanced)," Third-Gen. Partnership Proj., Sophia-Antipolis Cedex, France, 3GPP TR 36.912, Mar. 2011, ver. 10.0.0.

[21] "E-UTRA random access," Third-Gen. Partnership Proj., Sophia-Antipolis Cedex, France, 3GPP R1-060584, Feb. 2006, Ericsson, RAN1#44.

Chia-Hung Wei received the B.S. degree in computer and communication engineering from National Kaohsiung First University of Science and Technology, Kaohsiung, Taiwan, the M.S. degree in communication engineering from Tatung University, Taipei, Taiwan, and the Ph.D. degree from the Department of Electronic and Computer Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan.

From 2008 to 2010, he was with the Center of Wireless Broadband Technology, Tatung University, where he was an assistant researcher. His research interests include machine-type communications, multichannel slotted aloha, and performance analysis of Fourth-Generation cellular networks.

Ray-Guang Cheng (S'94–M'97–SM'07) received the B.E., M.E., and Ph.D. degrees in communication engineering from National Chiao Tung University, Taiwan, in 1991, 1993, and 1996, respectively.

From 1997 to 2000, he was a Researcher and a project leader with Advance Technology Center, Computer and Communication Laboratories, Industrial Technology Research Institute (ITRI), Taiwan. From 2000 to 2003, he was a senior manager with the R&D division of BenQ Mobile System Inc., Taiwan. He is currently a Professor with the Department of Electronic Engineering, National Taiwan University of Science and Technology, Taiwan. His research interests include multi-hop wireless networks and machine-to-machine communications.

Dr. Cheng is a senior member of IEEE communications society and the Phi Tau Phi scholastic honor society. He has published more than 70 international journal and conference papers and holds 15 U.S. patents. He led the Third-Generation Protocol project, and his team was named Top Research Team of the Year by ITRI in 2000. He received the Best Industrial-based Paper Award from Ministry of Education in 1998; the Advanced Technologies Award from the Ministry of Economic Affairs in 2000; and the Teaching and Research Awards from NTUST in 2006 and 2009, respectively.

Shiao-Li Tsao (M'04) received the Ph.D. degree in engineering science from National Cheng Kung University, Tainan, Taiwan, in 1999.

In 1998, he was a Visiting Scholar with Bell Laboratories—Alcatel-Lucent, New Providence, NJ, USA. He was a Visiting Professor with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada, in 2007 and with the Department of Computer Science, Swiss Federal Institute of Technology Zurich (ETH Zürich), Zürich, Switzerland, from 2010 to 2013. From 1999 to 2003, he was a Researcher and a Section Manager with Computer and Communications Research Laboratories, Industrial Technology Research Institute (ITRI). He is currently an Associate Professor with the Department of Computer Science, National Chiao Tung University, Hsinchu, Taiwan. He has published more than 95 international journal and conference papers. He is the holder of 18 U.S. patents. His research interests include embedded software and system and mobile communication and wireless networks.

Dr. Tsao received the Research Achievement Award from ITRI in 2000 and 2004, the Young Engineer Award from the Chinese Institute of Electrical Engineering in 2007, the Outstanding Teaching Award from National Chiao Tung University, and the K. T. Li Outstanding Young Scholar Award from the Association for Computing Machinery Taipei/Taiwan Chapter in 2008.