# A Hint-based Random Access Protocol for mMTC in 5G Mobile Network

(Invited Paper)

Yining Xu[†], Yi Ren[‡], Kate Ching-Ju Lin[†] and Yu-Chee Tseng[†]
[†]Department of Computer Science, College of Computer Science, National Chiao Tung University, Hsinchu, Taiwan
[‡]School of Computing Science, University of East Anglia, Norwich, UK
xuynktw@gmail.com, e.ren@uea.ac.uk, {katelin, yctseng}@cs.nctu.edu.tw,

***Abstract*—With the increasing popularity of machine-type communication (MTC) devices, several new challenges are encountered by the legacy long term evolution (LTE) system. One critical issue is that a massive number of MTC devices trying to conduct random access procedures may cause significant collisions and long delays. In this work, we present a new random access mechanism by splitting the contention-based preambles in LTE into two logically disjoint parts, one for the user equipment (UE) being paged and the other for the UEs not being paged. Since the IDs of paged UEs are known by the base station, a novel hash-based random access, which we call *hint*, is possible. The main idea is to pre-allocate preambles to paged UEs in a contention-free manner and confines non-paged UEs to contend in a separate region. We further build a mathematical model to find the optimal ratio of pre-allocated preambles. Extensive simulations are conducted to validate our results.**

***Index Terms*—Communication Protocol, Internet of Things, LTE, Machine-Type Communication (MTC), Wireless Network**

## I. Introduction

The number of cellular Internet of things (IoT) connections is expected to reach 3.5 billion in 2023—increasing with an annual growth rate of 30% [1]. Connecting to such massive IoT devices imposes new challenges for current cellular networks. To handle the challenge, new massive IoT cellular technologies, e.g., Narrowband IoT (NB-IoT) and Cat-M1, are driving the deployment of cellular IoT applications. Also, 3GPP has developed new radio access technologies based on long term evolution-advance (LTE-A) to support massive connections (up to 50,000) within an eNodeB (eNB) [2].

To upload data, a user equipment (UE) needs to attach and establish time synchronization with an eNB. A random access (RA) procedure is used to help UEs attach to an eNB. In the procedure, UEs need to contend for a limited number of preambles (e.g., 64 in LTE-A) to win attaching opportunities. However, when the number of contending UEs is large, collisions may lead to low RA success probability. Further, these failed UEs would continue trying in the following rounds, leading long access delays. The goal of this work is to reduce such collisions when UEs contend for preambles.

Reducing RA collisions has been studied through preamble allocation (e.g., [3]–[5]) and access barring (e.g., [6]–[10]). Preamble allocation schemes split RA preambles for different access purposes. However, the splitting is static irrespective of the network conditions. Access barring schemes limit contending UEs in a probabilistic manner. The barring factor is decided based on these aspects: the number of collisions in the radio access network (RAN), the collision information in the core network, and the channel statistical occupancy rate. Although the current contention level can be reduced, the effect is propagated to later rounds. Recently, Hint-based protocols (e.g., [11]–[15]) are proposed for reducing collisions. Perhaps the closest work to ours was studied in [15]. In this work, we further enhanced [15] by setting an access barring mechanism to guarantee the success probability of non-paged UEs.

In this paper, we propose a new Hint-based scheme to improve the RA success probability. We divide the preambles into two categories, one for paged UEs and the other for non-paged UEs. By pre-assigning some preambles to paged UEs and setting access barring for non-paged UEs, we make the RA success probability higher. Our scheme makes use of the paged UEs' IDs which are known by an eNB in advance. Meanwhile, our scheme follows the principle of fairness by trying to balance the success probability of paged and non-paged UEs. It does not sacrifice success probability of some UEs to improve that of the others. Our scheme relies on hashing UEs' IDs to their assigned preamble indices using a broadcast message from the eNB. Both mathematical analysis and simulation results are presented to prove the effectiveness of our scheme. According to our evaluation, the RA success probability can be increased by 10%∼40% and the average access delay can be decreased by about 50%.

The rest of the paper is organized as follows. Section II reviews the related work. Section III introduces some preliminaries. We present the proposed scheme in Section IV, followed by mathematical analysis in Section V. Performance evaluation is shown in Section VI. Section VII offers conclusions.

## II. Related Work

How to reduce RA collisions has been intensively studied. We review these works below.

The first type is the preamble allocation approaches (e.g., [3]–[5]), where RA preambles are divided into several different classes for different purposes. The authors in [3] proposed two RA preamble separate schemes. The first approach divides RA preambles into two subsets; one subset is reserved for H2H devices, and the other is allocated to machine-type communication (MTC) devices. The second approach also splits RA preambles into two subsets: one subset for H2H devices only, but the other subset is shared by both H2H devices and MTC devices. The scheme helps to diminish the negative impacts of MTC devices on H2H devices. However, since the available resources are significantly reduced for MTC devices, the performance tends to be worse if MTC traffic load is high. In dynamic RA resource allocation approaches, eNBs can dynamically allocate RA resources based on traffic load. The authors in [4] proposed to dynamically allocate RA resources between preambles and data. An optimization problem is to maximize the M2M data throughput where the average RA delay is less than a given upper bound. However, the dynamic allocation algorithm is based on the prediction of data traffic, which is not accurate in some scenarios. In [5], a generalized resource allocation scheme to support access prioritization in a multichannel slotted ALOHA system is presented. These schemes [3]–[5] do reduce RA collision. However, the splitting is static and can not adapt to traffic loads.

The second type is access barring. Its basic idea is to define a barring factor as a threshold and broadcast it over a cell. Once an eNB is overload, the eNB broadcast the barring factors (usually a threshold and a barring timer). According to the barring factors, a UE generates a random number to decide whether to contend or delay a while specified by the barring timer. 3GPP defines extended access barring (EAB) [6] further considering delay-tolerant applications and delay-constrained applications. EAB gives priorities to delay-constrained UEs, however, it may cause longer access delays to some devices.

The configuration strategies of barring factors are intensively studied by research communities (e.g., [7]–[10]). To facilitate devices escaping from continuous congestion, [7] proposes a cooperative barring for global stabilization and access load sharing. The authors of [8] proposed a drift-based backlog estimation scheme for reducing RA collisions in MTC. The scheme defines a provably stable control procedure for the barring factor in the sense that it maintains a finite number of unserviced users in the system. Based on the work [8], reference [9] introduces an iterative algorithm to update the barring factors, which yields near-optimal performance and a reduction in the total service time compared with [8]. The authors of [10] proposed two queueing model based algorithms to adjust the barring factors and backoff timers to control RA congestion. Our work differs from the above schemes in that we reserve some preambles for paged UEs to guarantee success, but still maintain the flexibility for the remaining UEs for RA process.

The Hint protocols open a new research direction for reducing collisions (e.g., [11]–[15]). The Hint protocols leverage a tiny broadcast message (i.e., *hint*) to deliver encoded information. The *hint* allows devices to "decode" their transmission slots, which significantly reduces transmission and contention overheads. A series of Hint-based frameworks is firstly proposed in [11], [12] based on the assumption that each data has the same size. In [13], a Chinese reminder theorem based Hint protocol is proposed for reducing RA response message size. Later, the work [11], [12] are enhanced by supporting various size data transmission in [14]. Inspired by the Hint protocols, the authors of [15] tailored the Hint protocols for RA procedures. It works in two modes: (1) collision-free paging and (2) collision-avoidance paging. The approach in this work is similar to the collision-free mode in [15]. But here we enhance it by adding a balancing mechanism to control the success probabilities of paged and non-paged devices and setting an access barring mechanism to guarantee the success probability of non-paged UEs.

## III. Preliminaries

In this section, we briefly introduce RA procedure and paging procedure.

### A. Random Access Procedure

The RA procedure is triggered by a UE when it needs to establish a connection with an eNB (e.g., for association with a network, for synchronization with an eNB after a long idle period, after radio-link failure, or after changing a serving eNB). 3GPP defines the following RA procedure, as shown in Fig. 1.

1) Preamble Transmission (Msg1): The UE selects one of the 64 physical RA channel (PRACH) signatures and transmits in that preamble.
2) Random Access Response (Msg2): The RA response (RAR) is sent by the eNB on the physical downlink shared channel (PDSCH) with each successful UE's ID, its RA radio network temporary identifier (RA-RNTI), and allocated the time-frequency slot. If multiple UEs had collided by selecting the same signature in the same preamble, they would both receive the same RAR.
3) Radio Resource Control (RRC) Connection Request (Msg3): This message is the first scheduled uplink transmission on the physical uplink shared channel (PUSCH). It makes use of hybrid automatic repeat request (HARQ) and conveys the actual RA message.
4) Contention Resolution (Msg4): The UE's behavior upon receiving this message has three possibilities. If it correctly detects its own ID, it sends back a positive ACK. If it decodes another UE's ID, it sends nothing back. If it cannot decode the message, it sends nothing back.

### B. Paging Procedure

The main purpose of paging is to initiate services for UEs that are in idle mode. It wakes up an idle UE when there is some data for it. UEs momentarily wake up in each discontinuous receive (DRX) cycle and calculate their own paging frame (PF) and paging occasion (PO). Paging messages

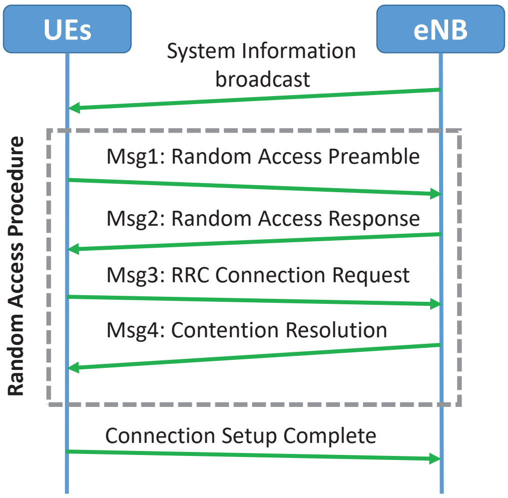



Fig. 1: Random access procedure in LTE-A.

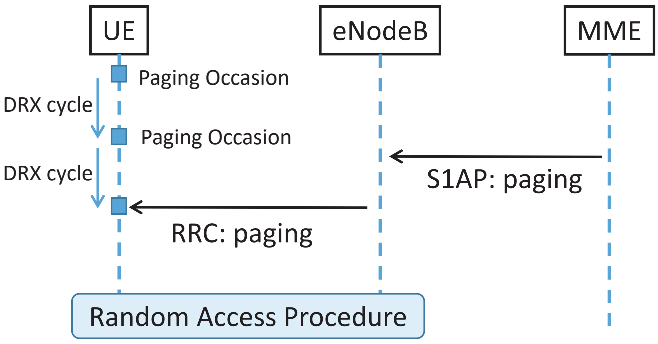



Fig. 2: Paging procedure (MME initiated)

are sent in POs, which contain the UEs' IDs (such as TMSI). If a UE does not find its ID in a paging message, it will assume that it is not paged and go back to sleep. Otherwise, it will react by triggering an RRC connection request message. Fig. 2 shows the paging procedure initiated by mobility management entity (MME).

## IV. Proposed Scheme

In this section, we first define the problem and then present Hint protocol.

### A. Problem Statement

We consider a set of UEs contending for access to an eNB at a particular time, denoted as $D = P \bigcup R$. Here, $P$ denotes the set of UEs to be paged by the eNB, and $R$ denotes the set of UEs that are not paged but need to go through RA to be attached to the eNB. The concept is shown in Fig. 3(a). These UEs are not attached to the eNB and thus need to perform the RA procedure to communicate with the eNB. Since the members of $P$ are known according to LTE paging protocol, our goal is to schedule some of them, denoted by subset $P' \subseteq P$, in scheduled preambles. The UEs in $R$ will contend in

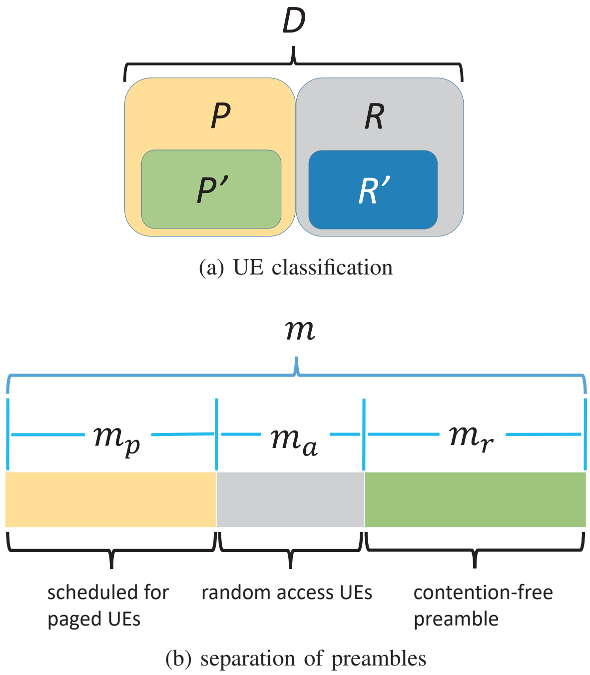



Fig. 3: UE subsets and preambles

separate preambles (not used by $P$). The purpose of our design is to increase the access success probability $F_s$:

$$F_s = \frac{S_s}{N}, \tag{1}$$

where $S_s$ is the number of successful UEs in set $D$ and $N = |D|$. By separating $F_s$ into paging access success probability $F_p$ and RA success probability $F_r$, we have

$$F_s = \frac{F_p \times N_p + F_r \times N_r}{N}, \tag{2}$$

where $|P| = N_p$ and $|R| = N_r$. We will show how to control the two probabilities $F_p$ and $F_r$ to improve the overall $F_s$.

### B. Hint Protocol

In the legacy LTE, as shown in Fig. 3(b), there are totally $m$ preambles, among which $m_r$ preambles are reserved for priority use (such as handover UEs). The rest of the $m - m_r$ preambles are to be used by those UEs in $D$ through RA. The collision probability increases rapidly as the number of contending UEs grows, leading to spectrum under-utilization. In this work, we leverage the property that an eNB could easily predict the members in $P$ and thus can schedule some safe preambles for a selected subset $P' \subseteq P$ using a Hint-based approach. This is achieved through a specially designed hashing function with UE's ID and a common hint message as inputs and separation of the preamble pools for $P$ and $R$. We shall control $P'$ and the actual contending devices in $R$ to keep a balance between paged and non-paged UEs. The detail design is as follows. We only modify UE's behaviors in Msg1 in Fig. 4.

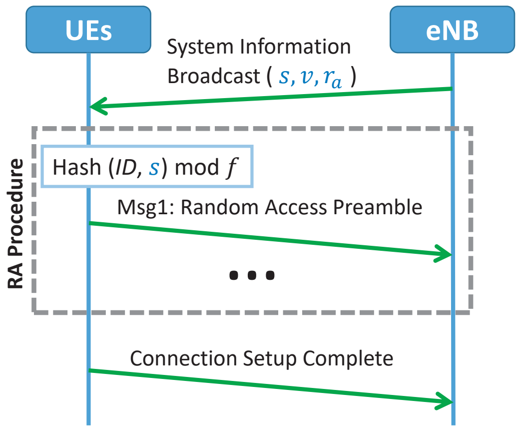



Fig. 4: Hint-based RA procedure.

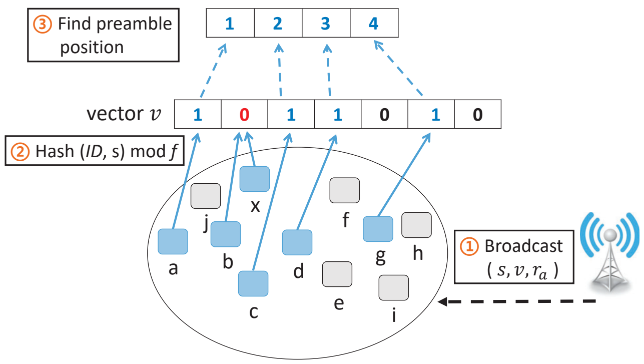



Fig. 5: Example of scheduling preambles for $P'$.

1) The eNB computes two parameters, the paging ratio $r_p$ and the access barring $r_a$ (how to compute them will be discussed in Section V. The eNB broadcasts a hint message $(s, v, r_a)$ in System Information broadcast, which is defined as follows.
   a) $s$ is a random seed chosen by the eNB.
   b) $v$ is a binary vector of length $f$, which is computed as follows. The eNB will schedule up to $r_p \times |P|$ preambles for devices in $P$. Let $P' \subseteq P$ be the subset of devices that will be scheduled. Initially, let $P' = \emptyset$ and all elements of $v$ are 0's. At the end, the corresponding elements of $P'$ in $v$ will be set to 1's. We repeat the following steps until $P'$ is finalized.
      i) Randomly pick an $x \in P$.
      ii) If $Hash(s, x)\ mod\ f \neq Hash(s, x')\ mod\ f$ for all other $x' \in P$ but $x' \neq x$, we add $x$ into $P'$ and set $v[Hash(s, x)\ mod\ f] = 1$. (This means that $x$ will be scheduled with a dedicated preamble.) Otherwise, we set $x$ as "examined".
      iii) If $|P'| = r_p \times |P|$ or all devices in $P$ have been "examined", exit the loop. Otherwise, go to step i).
2) For each UE $x$ receiving the above hint message $(s, v, r_a)$, it counts the number of 1's in $v$, denoted as $m_p$. The first $m_p$ preambles are for devices in $P'$. The rightmost $m_r$ preambles are reserved ones. The middle $m_a = m - m_p - m_r$ preambles are for devices in $R$. Such partitioning is shown in Fig. 3(b). There are two cases.
   a) If $x \in P$, it checks if $v[Hash(s, x)\ mod\ f] = 1$. If so, it counts the number of 1's in $v$ in front of its '1', denoted as $c$. Then it transmits in the $(c+1)th$ preamble in the $m_p$ preambles reserved for $P'$. Otherwise, $x \notin P'$ and it will not contend in this round.
   b) If $x \in R$, it picks a random number $r \in [0, 1]$. If $r \leq r_a$, then this UE will consider itself in set $R'$. It contends by selecting any of the $m_a$ middle preambles for $R$ to transmit its preamble.

In the following, we give an example. Let $P = \{a, b, c, d, g, x\}$ are the UEs to be paged. Suppose that $ID_a = 259$, $ID_b = 279$, $ID_c = 221$, $ID_d = 222$, $ID_g = 184$, and $ID_x = 41$. Let the optimal $r_p = 0.66$, so $0.66 \times |P| = 4$ UEs will be allocated distinct preambles. Suppose that the hash function $Hash(ID, s) = ID\ mod\ s$, where the random seed $s = 139$ is found after some trying. The hash results are 120, 1, 82, 83, 45, and 41, respectively. Letting $f = 40$, the final results are 0, 1, 2, 3, 5, and 1, respectively. Because the eNB should schedule 4 UEs' preambles, $b$ and $x$, (which have collision) will not be scheduled. Then, the eNB would set 1's in the positions of 0, 2, 3, and 5 in vector $v$. Afterwards, the eNB broadcasts $(s = 139, v, r_a)$. For $c$, its hashing result is 2. The bit in $v$ is '1'. Since this is the second '1' in $v$, $c$ will choose the second preamble to transmit. For $x$, its hashing result is 1. The bit in $v$ is '0'. Thus, $x$ would not contend in this round. This is the same for $b$. The example is shown in Fig. 5.

## V. Mathematical Analysis

In this section, we will show how to find the optimal paging ratio $r_p$ and access barring ratio $r_a$. We first define some important parameters. In the conventional RA procedure, since all UEs contend on the same pool of preambles, the success probability of a UE is:

$$F_c = \left(1 - \frac{1}{m - m_r}\right)^{N_r + N_p - 1}. \tag{3}$$

The expected number of successful UEs is $E_c = F_c \times (N_p + N_r)$. Given $r_p$ and $r_a$ in our method, the success probability of a non-paged UE is a function:

$$F_r(r_p, r_a) = \left(1 - \frac{1}{m - m_r - r_p N_p}\right)^{r_a N_r - 1}. \tag{4}$$

The expected number of successful non-paged UEs is $E_r(r_p, r_a) = F_r(r_p, r_a) \times r_a N_r$. Since only a ratio of $r_p$ UEs are scheduled, the success probability of paged UEs is simply written as

$$F_p(r_p) = r_p. \tag{5}$$

TABLE I: List of Notations

| Notation | Definition |
|---|---|
| $D$ | A set of UEs |
| $P$ | The set of UEs to be paged |
| $P'$ | The set of scheduled UEs controlled by $r_p$ |
| $R$ | The set of random UEs |
| $R'$ | The set of UEs controlled by access barring $r_a$ |
| $F_s$ | The RA success probability of Hint protocol |
| $F_c$ | The RA success probability of legacy LTE-A |
| $F_p$ | The RA success probability of paged UEs |
| $F_r$ | The RA success probability of random UEs |
| $N$ | The number of total UEs |
| $N_p$ | The number of paged UEs |
| $N_r$ | The number of random access UEs |
| $E_c$ | The number of successful UEs of legacy LTE-A |
| $E_r$ | The number of successful non-paged UEs |
| $m$ | The number of total preambles |
| $m_p$ | The number of scheduled preambles for paged UEs |
| $m_r$ | The number of reserved preambles |
| $m_a$ | The number of preambles for random access UEs |
| $r_p$ | Paging ratio |
| $r_a$ | Access barring factor |
| $S_s$ | The number of successful UEs |
| $s$ | Random seed |
| $v$ | A binary vector of length $f$ |
| $f$ | The size of contention-based preambles defined in SIB2 |

Our analysis includes two steps. First, we set $r_a = 1$ (i.e., without access barring) and try to find a good $r_p$. Second, with a fixed $r_p$, we try to find the best $r_a$. With $r_a = 1$, we simply write $F_r(r_p, r_a)$ as $F_r(r_p)$. By differentiating $F_r(r_p)$, we find that $F_r(r_p)$ is monotonically decreasing with respect to $r_p$. On the other hand, $F_c$ is a constant with respect to $r_p$.

We draw the conceptual curves of $F_r(r_p)$ and $F_p(r_p)$ in Fig. 6. If we set $F_r(r_p) = F_p(r_p)(= r_p)$, the intersection point $B$ would represent the fair point where both paged and non-paged UEs have the same success probability $B_y$. In Fig. 6, we also draw the conceptual line of $F_c$. Let the intersection of $F_r(r_p)$ and $F_c$ be C. The following theorem proves that $B_y > C_y$, which means that our scheme always outperforms the conventional scheme by setting $F_r(r_p) = r_p$.

**Theorem 1.** *By setting $F_r(r_p) = F_p(r_p)$, it is guaranteed that $F_r(r_p) > F_c$.*

*Proof.* We would prove that point $B$ is above point $C$ in Fig. 6, i.e., $B_y > C_y$. This is equivalent to proving that $D_y > C_y$. Since $C_x = D_x = D_y$, this is equivalent to finding the $r_p$ such that $F_r(r_p) = F_c$. According to Eqs. (3)(4), we have

$$\left(1 - \frac{1}{m - m_r}\right)^{N_r + N_p - 1} = \left(1 - \frac{1}{m - m_r - r_p N_p}\right)^{r_a N_r - 1}.$$

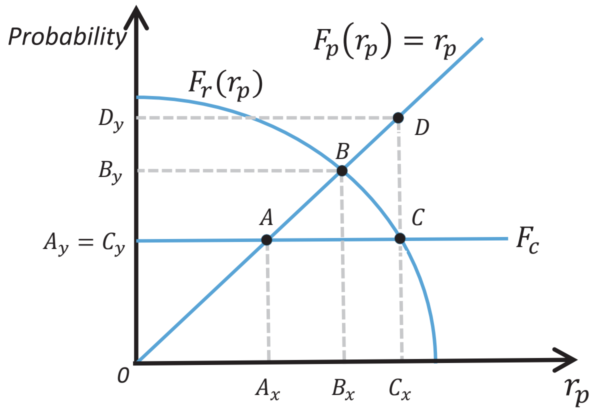



Fig. 6: Conceptual trend of $F_r(r_p)$, $F_p(r_p)$, and $F_c$.

It is not hard to derive that

$$C_x = D_x = D_y = \frac{m_p - \frac{m_p^k}{m_p^k - (m_p - 1)^k}}{N_p},$$

where $k = \frac{N_p + N_r - 1}{r_a N_r - 1}$ and $m_p = m - m_r$. On the other hand, $C_y = F_c$, the gap between points $D$ and $C$ on the vertical axis is

$$\begin{aligned} g(N_p) &= D_y - C_y \\ &= \frac{m_p - \frac{m_p^k}{m_p^k - (m_p - 1)^k}}{N_p} - (1 - \frac{1}{m_p})^{N_r + N_p - 1}. \end{aligned}$$

Proving $g(N_p) > 0$ is equivalent to proving $g(N_p)N_p > 0$ because $N_p > 0$. Letting $t = \frac{N_p}{N_r - 1}$, we have

$$\begin{aligned} g_2(N_p) = & g(N_p)N_p = m_p - \frac{1}{1 - (1 - 1/m_p)^{t+1}} \\ & - N_p(1 - \frac{1}{m_p})^{(\frac{1}{t}+1)N_p}. \end{aligned}$$

Taking the derivation of $N_p$,

$$\begin{aligned} g_2'(N_p) = & -\frac{1}{t}(\frac{1}{m_p}(m_p - 1))^{\frac{N_p}{t}(t+1)} \\ & \times (N_p(t+1)\ln(\frac{1}{m_p}(m_p - 1)) + t). \end{aligned}$$

When $\alpha_p = -\frac{t}{(t+1)\ln(1 - \frac{1}{m_p})}$, $g_2'(\alpha_p) = 0$. The global minimum is

$$\begin{aligned} g_2(\alpha_p) = & m_p - \frac{1}{1 - (1 - \frac{1}{m_p})^{t+1}} \\ & + \frac{1}{(1 + \frac{1}{t})\ln(1 - \frac{1}{m_p})}\left(1 - \frac{1}{m_p}\right)^{-\frac{1}{\ln(1 - \frac{1}{m_p})}} \\ = & m_p - \frac{1}{1 - (1 - \frac{1}{m_p})^{t+1}} + \frac{1}{e(1 + \frac{1}{t})\ln(1 - \frac{1}{m_p})}. \end{aligned}$$

Next, we would prove that $g_2(\alpha_p) > 0$. Because

$\frac{1}{\ln(1-\frac{1}{m_p})} > -m_p$, we have

$$g_2(\alpha_p) > m_p - \frac{1}{1-(1-\frac{1}{m_p})^{t+1}} - \frac{tm_p}{(1+t)\mathrm{e}}.$$

Because $\frac{t}{t+1} < 1$, we have

$$g_2(\alpha_p) > m_p - \frac{1}{1-(1-\frac{1}{m_p})^{t+1}} - \frac{m_p}{\mathrm{e}} = g_3(\alpha_p).$$

When $t \geq 1$, $g_3(\alpha_p)$ increases monotonically. When $t = 1$,

$$\begin{aligned} g_3(\alpha_p) &= m_p - \frac{1}{1-(1-\frac{1}{m_p})^2} - \frac{m_p}{\mathrm{e}} \\ &= \frac{m_p((\mathrm{e}-2)m_p + 1 - \mathrm{e})}{(2m_p - 1)\mathrm{e}}. \end{aligned}$$

When $m_p \geq 3$, $(\mathrm{e}-2)m_p + 1 - \mathrm{e} > 0$. Therefore, $g_3(\alpha_p) > 0$ when $t \geq 1$. When $0 \leq t \leq 1$, because $(1-\frac{1}{m_p})^t < (1-\frac{t}{m_p})$,

$$\begin{aligned} g_2(\alpha_p) &> m_p - \frac{1}{1-(1-\frac{1}{m_p})(1-\frac{t}{m_p})} - \frac{tm_p}{(t+1)\mathrm{e}} \\ &= \frac{m_p}{(m_p\frac{t+1}{t} - 1)\mathrm{e}}\Big[(m_p-1)\mathrm{e} - \mathrm{e} + \frac{t}{t+1}\Big]. \end{aligned}$$

Because $\frac{m_p}{(m_p\frac{t+1}{t}-1)\mathrm{e}} > 0$ and $(m_p - 1)\mathrm{e} - \mathrm{e} + \frac{t}{t+1} > 0$, $g_2(\alpha_p) > 0$ when $m_p \geq 2$. Therefore, $g_2(\alpha_p) > 0$ when $t \geq 0$, which proves that $D_y > C_y$. □

The above theorem suggests a fair success probability for both paged and non-paged UEs that is higher than that of the conventional method. As shown in Fig. 6, $F_r(r_p)$ is higher than $F_c$ when $r_p \in [0, C_x)$. On the other hand, $F_p(r_p)$ is higher than $F_c$ when $r_p \in (A_x, 1]$. Therefore, $(A_x, C_x)$ is the range for $r_p$ such that our method works better. However, in Theorem 1, the access barring $r_a$ is not applied. In the following theorem, we show that after setting $r_p$, we can further adjust $r_a$ to improve $F_r(r_p)$.

**Theorem 2.** *Let $\hat{r}_p$ be the value of $r_p$ such that $F_r(\hat{r}_p, 1) = F_p(\hat{r}_p)$. $F_r(\hat{r}_p, r_a)$ is maximized when $r_a = \frac{-1}{N_r \ln Q}$, where $Q = (1 - \frac{1}{m - m_r - \hat{r}_p N_p})$.*

*Proof.* With a fixed $\hat{r}_p$, we rewrite

$$\begin{aligned} E_r(\hat{r}_p, r_a) &= F_r(\hat{r}_p, r_a) \times r_a N_r \\ &= \Big(1 - \frac{1}{m - m_r - \hat{r}_p N_p}\Big)^{r_a N_r - 1} \times r_a N_r \\ &= Q^{r_a N_r - 1} \times r_a N_r. \end{aligned}$$

Derivating $E_r(\hat{r}_p, r_a)$ by $r_a$,

$$E_r'(\hat{r}_p, r_a) = N_r Q^{r_a N_r - 1}(1 + r_a N_r \ln Q). \quad (6)$$

By solving the differential equation $E_r'(\hat{r}_p, r_a) = 0$, we have

$$r_a = \frac{-1}{N_r \ln Q}.$$

Next, we would prove that $E_r'(\hat{r}_p, r_a) < 0$ when $r_a > \frac{-1}{N_r \ln Q}$, and that $E_r'(\hat{r}_p, r_a) > 0$ when $r_a < \frac{-1}{N_r \ln Q}$. Since $N_r$ is a positive integer and $Q \in (0, 1)$, it is easy to see that $N_r Q^{r_a N_r - 1} > 0$. According to Eq. (6), whether $E_r'(\hat{r}_p, r_a)$ is positive or negative is determined by $(1 + r_a N_r \ln Q)$.

When $r_a < \frac{-1}{N_r \ln Q}$, we have

$$\begin{aligned} r_a N_r &< \frac{-1}{\ln Q}, \\ r_a N_r \ln Q &> -1, \\ (1 + r_a N_r \ln Q) &> 0, \end{aligned}$$

so $E_r'(\hat{r}_p, r_a) > 0$.

Similarly, $E_r'(\hat{r}_p, r_a) < 0$ when $r_a > \frac{-1}{N_r \ln Q}$. Therefore, $E_r(\hat{r}_p, r_a)$ is maximized when $r_a = \frac{-1}{N_r \ln Q}$. □

In the above proof, it is essential to find the point $\hat{r}_p$. However, we are unable to find a closed form for $\hat{r}_p$. We propose an estimation as follows.

(1) Find $A_x$: This is equivalent to solving $F_c = r_p$. Hence,

$$A_x = (1 - \frac{1}{m_p})^{N_r + N_p - 1}. \quad (7)$$

(2) Find $C_x$ (or $D_x$): This is equivalent to solving $F_c = F_r(r_p)$. We can derive that

$$C_x = \frac{m_p - \frac{m_p^k}{m_p^k - (m_p - 1)^k}}{N_p}, \quad (8)$$

where $k = \frac{N_p}{N_r - 1} + 1$.

(3) Conduct a binary search in $(A_x, C_x)$ to find an approximation of $B_x$.

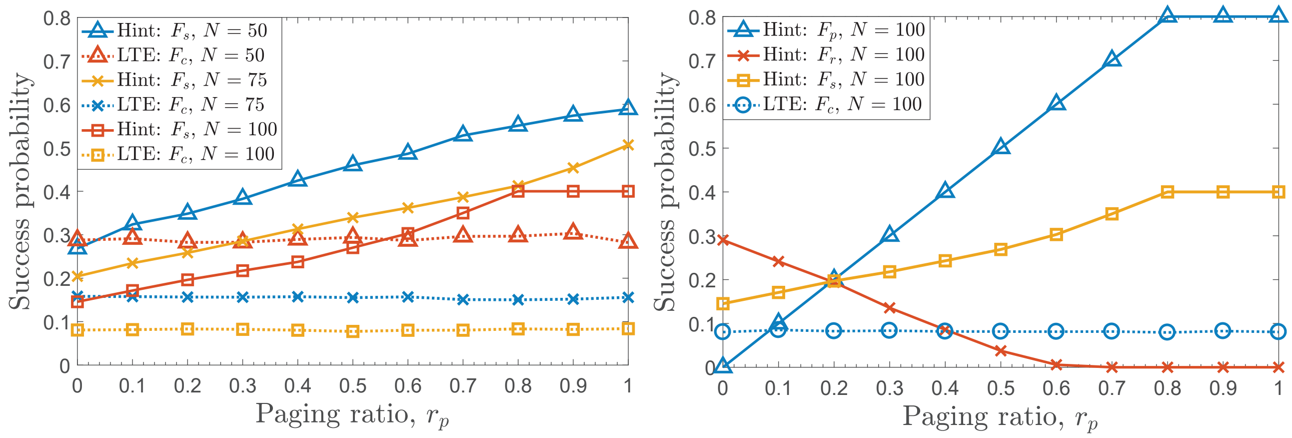
## VI. Performance Evaluation

In this section, we show through extensive simulations to compare the performance of the proposed Hint-based RA protocol with the legacy LTE-A. The comparison is based on two performance metrics: *success probability* and *paging latency*. The *success probability* is $F_s$, $F_c$, $F_r$, and $F_p$ as defined in Eqs. (1)(3)(4)(5), while the *paging latency* is the average number of paging iterations required to successfully page all the members in $P$. We evaluate our performance from different aspects. In each simulation, we report the average results of 1000 simulations.

### A. Successful Probability

*1) Impact of paging ratio $r_p$:* Figs. 7(a) and 7(b) show the impact of paging ratio $r_p$ on the success probability of Hint protocol and legacy LTE-A. The percentage of the paged UEs is set to 0.5, i.e., $N_p/N = 0.5$. The range of paging ratio is from 0.1 to 1. The results reveal the following findings.

- The correctness of the analytical models is cross-validated by extensive simulations, where the both results match very well.
- The paging ratio $r_p$ has significant impact on the success probability of Hint protocol, yet has no help on legacy LTE-A. Increasing $r_p$ improves the performance of Hint protocol.




(a) Success probability comparison between Hint protocol and legacy LTE.

(b) Success probability in terms of $F_p$, $F_r$, $F_s$, and $F_c$.

Fig. 7: Impact of paging ratio on success probability, where $m_p = 40$ and $N_p/N = 0.5$.

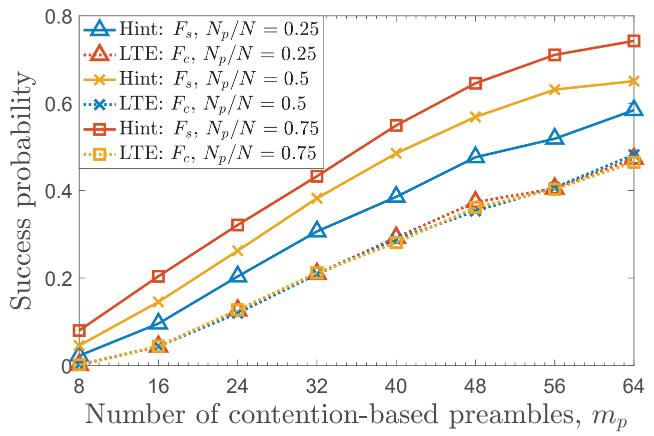



Fig. 8: Impact of the number of contention-based preambles on success probability, $N = 100$.

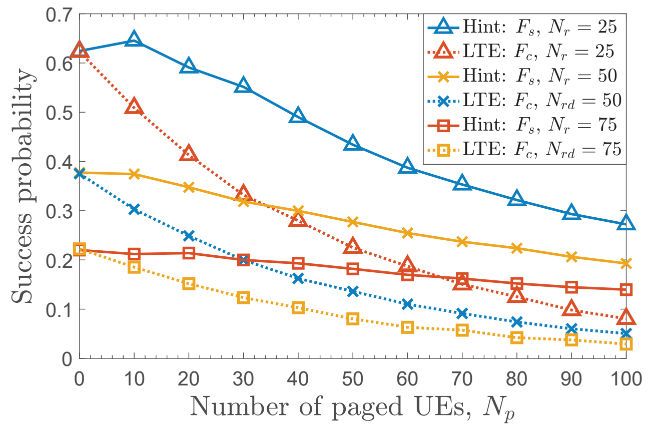



Fig. 9: Impact of the number of paged UEs on success probability, $m_p = 50$.

- The performance of Hint protocol is better than legacy LTE-A. And a larger $r_p$ leads to a bigger gap between them, as shown in Fig. 7(a). Indeed, increasing $r_p$ leads to higher success probability. However, the success probability of RA UEs $F_r$ is reduced as $r_p$ grows (see Fig. 7(b)). To guarantee fairness, i.e., $F_r = F_p = F_s$, the intersection of the curves $F_r$ and $F_p$ can be found in Fig. 7(b). When $r_p = 0.2$, the successful probabilities of Hint protocol $F_r = F_p = F_s$ are better than that of legacy LTE-A, which is coincident with Theorem 1.

*2) Impact of the number of preambles $m_p$:* In Fig. 8, we fix the number of all contention-based UEs $N$ to 100. The paging ratio $r_p$ is set to the optimal ratio, which is found by the binary search method addressed in Section V. It is easy to see that Hint protocol has much higher successful probabilities than legacy LTE-A no matter the setting on $N_p/N$. We observe that the gap between Hint protocol and legacy LTE-A increases when $N_p/N$ grows. The results show that Hint protocol has better success probability when there are more paged UEs. We also see that $m_p$ has significant impact on success probability. Increasing $m_p$ leads to success probability grows. But the three lines of legacy LTE-A schemes almost overlap entirely, meaning that the percentages of paged UEs has no impact on $F_c$, which are coincident with our discussion.

*3) Impact of the number of paged UEs $N_p$:* Fig. 9 depicts the success probability by varying $N_p$ from 0 to 100. We fix the number of contention-based preambles to 50. There are three scales for the number of random UEs $N_r$. For each $N_r$, the corresponding paging ratio is set to be optimal value as we derived in Theorem 2. As $N_p$ grows, both successful probabilities descend. However, the success probability of Hint protocol decreases slower than that of legacy LTE-A does. The results show that Hint protocol performs much better than legacy LTE-A scheme and has better performance when $N_p$ is large.

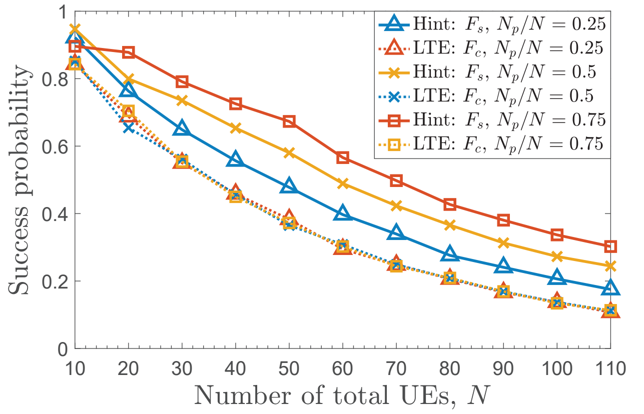



Fig. 10: Impact of total UEs on success probability, $m_p = 50$.

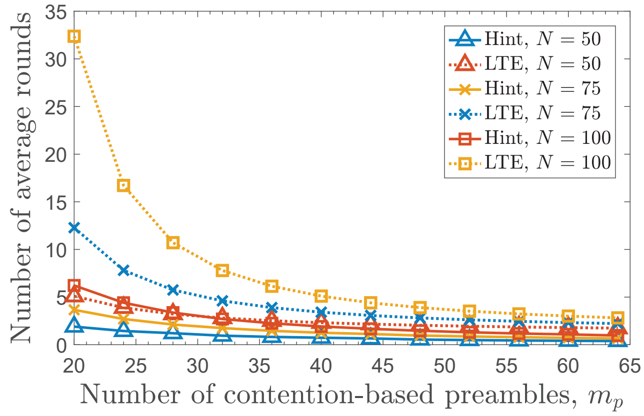



Fig. 12: Impact of contention-based preambles on the number of paging rounds, $N_p/N = 0.5$.

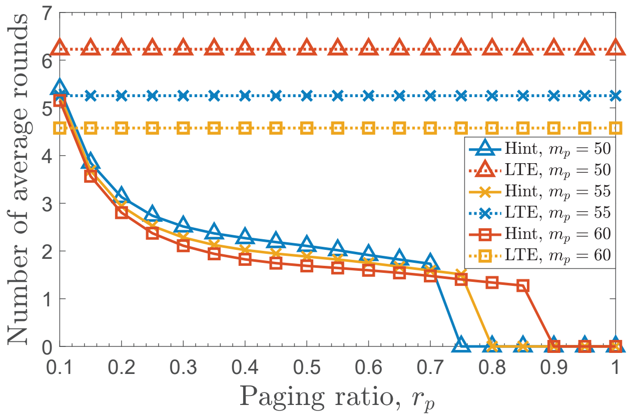



Fig. 11: Impact of paging ratio on the number of paging rounds, $N_p = N_r = 70$.

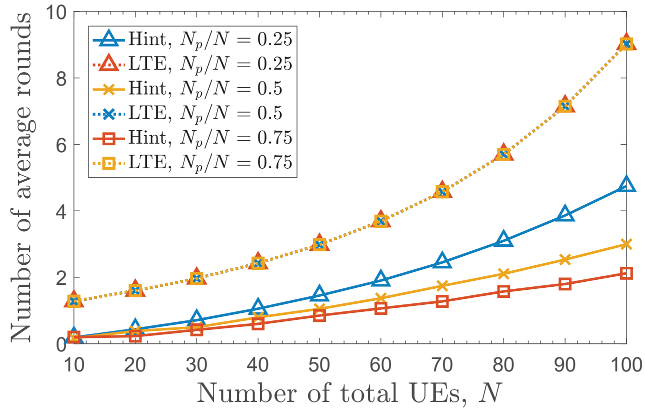



Fig. 13: Impact of total UEs on the number of paging rounds, $m_p = 30$.

*4) Impact of the number of total contention UEs $N$:* Fig. 10 illustrates the success probability by varying $N$ from 10 to 110. Three different settings are set for the percentage of paged UEs. In general, the curves descend as $N$ grows. The reason is that more UEs lead to more contention collisions leading to lower success probability. However, we can see that Hint protocol shows higher successful probabilities than legacy LTE-A. In addition, we observe that the three curves of legacy scheme almost overlap. Since the random access UEs and the paged UEs randomly choose the same range of preambles, the proportion of paged UEs has no impact on the success probability of the legacy scheme. While in Hint protocol, the larger proportion of paged UEs means that we can pre-assign more preambles to paged UEs. The results reveal that Hint protocol is scalable and especially suitable for large-scale MTC.

### B. Paging Latency

*1) Impact of paging ratio $r_p$:* Recall that paging ratio $r_p$ is used to control the number of UEs (i.e., $|P|r_p$) to obtain preambles successfully. As shown in Fig. 11, the results show that Hint protocol has lower paging latency than legacy LTE-A scheme. Specifically, legacy scheme is shown as three horizontal lines in different settings. The reason is that in legacy scheme paging ratio is set to the same value (i.e., $r_p = 1$). Regarding Hint protocol, the paging latency decreases as $r_p$ grows. A large $r_p$ means more preambles reserved for paged UEs so that more paged UEs will be successful, which is coincident with our analysis.

*2) Impact of the number of contention-based preambles $m_p$:* In Fig. 12, $N_p/N$ is set as 0.5. We use the optimal paging ratio $r_p$ and access barring $r_a$ derived in our model. It is easy to see that Hint protocol outperforms the legacy scheme no matter the value of $m_p$. We also observe that even $N$ is a large value, e.g., $N = 100$, Hint protocol still has very low paging latency, meaning that it is very suitable for large scale MTC.

*3) Impact of the number of total UEs $N$:* In Fig. 13, we fix $m_p$ to 30. Again, the paging ratio we set here is the optimal value derived based on our model. The results show that Hint protocol outperforms the legacy scheme and the gap between the two schemes increases as $N$ grows. We also observe that the curves of Hint protocol grow very slow when $N_p/N = 0.75$, meaning that Hint protocol is suitable for large scale MTC and especially good at the case with more paged UEs.

## VII. Conclusions

In this paper, we have proposed a Hint protocol for improving the random access procedure in LTE-A. Compared with the legacy LTE-A scheme, Hint protocol reduces the collisions in random access procedure, providing higher success probability and lower paging latency. We also prove the existence of a fairness point $\hat{r}_p$ that provides a success probability for both paged and non-paged UEs which outperforms the conventional scheme, which allows all UEs to compute randomly. With this fairness point, we are able to further calculate an optimal access barring ratio for random access UEs. Demonstrated through extensive simulations, the performance of Hint protocol is much better than the legacy LTE-A scheme in terms of RA success probability and paging latency.

## References


[1] Ericsson, "Mobility report," Jun. 2018.

[2] 3GPP TR 45.820, *Cellular system support for ultra-low complexity and low throughput Internet of Things (CIoT) (Release 13)*, Nov. 2015.

[3] K. D. Lee, S. Kim, and B. Yi, "Throughput comparison of random access methods for M2M service over LTE networks," in *Proc. IEEE GLOBECOM Workshops 2011*, pp. 373–377.

[4] D. T. Wiriaatmadja and K. W. Choi, "Hybrid random access and data transmission protocol for machine-to-machine communications in cellular networks," *IEEE Trans. Wirel. Commun.*, vol. 14, no. 1, pp. 33–46, 2015.

[5] Y.-Y. Chu, R. Harwahyu, R.-G. Cheng, and C.-H. Wei, "Study of generalized resource allocation scheme for multichannel slotted ALOHA systems," in *Proc. IEEE PIMRC*, 2015, pp. 1702–1706.

[6] 3GPP TS 36.321, *Medium Access Control (MAC) protocol specification (Release 14)*, 3GPP Std., Sep. 2017.

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

[8] O. Galinina, A. Turlikov, S. Andreev, and Y. Koucheryavy, "Stabilizing multi-channel slotted ALOHA for machine-type communications," in *Proc. IEEE Int'l Symp. Information Theory (ISIT)*, 2013, pp. 2119–2123.

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

[10] C.-W. Chang, Y.-H. Lin, Y. Ren, and J.-C. Chen, "Congestion control for machine-type communications in LTE-A networks," in *Proc. IEEE GLOBECOM*, 2016.

[11] Y. Ren, R.-J. Wu, and Y.-C. Tseng, "The Hint protocol: Using a broadcast method to enable ID-free data transmission for dense IoT devices," in *Proc. IEEE PerCom Workshop (PerIoT)*, 2017.

[12] Y. Ren, R.-J. Wu, T.-W. Huang, and Y.-C. Tseng, "Give me a Hint: An ID-free small data transmission protocol for dense IoT devices," in *Proc. IEEE Wireless Days (WD)*, 2017.

[13] T.-W. Huang, Y. Ren, K. C.-j. Lin, and Y.-C. Tseng, "r-Hint: A message-efficient random access response for mMTC in 5G networks," in *Proc. IEEE PIMRC*, 2017.

[14] T.-Y. Chan, Y. Ren, Y.-C. Tseng, and J.-C. Chen, "eHint: An efficient protocol for uploading small-size IoT data," in *Proc. IEEE WCNC*, 2017.

[15] C.-A. Hsu, Y. Ren, K. C.-J. Lin, and Y.-C. Tseng, "Hey! I have something for you: paging cycle based random access for LTE-A," in *Proc. IEEE ICC*, 2018.