Low Latency Random Access with TTI Bundling in LTE/LTE-A

Kaijie Zhou*, Navid Nikaein†
*Huawei Technologies Co., Ltd. China, zhoukaijie@huawei.com
†Eurecom, France, navid.nikaein@eurecom.fr

***Abstract*—To reduce the uplink channel access latency in LTE/LTE-A, we propose a Transmission Time Interval (TTI) bundling scheme for the random access procedure. With the proposed method, a UE sends multiple preambles in consecutive subframes in order to increase the success rate of random access and hence to reduce the latency. We introduce a Semi-Markov model to accurately model and analyze the random access mechanism with TTI bundling. With this model, we formulate the access latency as a function of the number of TTI bundles and select the optimal value which minimizes the channel access latency. The proposed Semi-Markov model is validated against simulation and the performance of the TTI bundling method is also evaluated. We find that channel access latency can be significantly reduced when the preamble collision rate is not high.**

***Keywords*—LTE, Random Access, TTI bundling, Semi-Markov Model.**

I. Introduction

Low-latency protocols and access methods are becoming crucial to improve the spectral efficiency and to lower the energy consumption in end-devices, especially in view of emerging application scenarios found in machine-to-machine communication, online interactive gaming, social networking and instant messaging. However, the majority of wireless systems, including LTE/LTE-A, are designed to support a continuous flow of information, at least in terms of the time-scales needed to send several IP packets, such that the induced signaling overhead is manageable. While these systems are intended mostly for downlink-dominant and bursty traffic, emerging application scenarios are of generally different characteristics [1], namely: uplink dominant packets, periodic and event-driven packets, small and low duty cycle packets. Such applications do not necessarily need an increase in the maximal data rate though part of the 4G-5G requirement [2], they rather call for an efficient support for low-latency, potentially coordinated, channel access, and in particular for small to very-small packets. We argue that in 4G/4G+ system, this represents an opportunity to limit the modifications in the physical layer and redesign the access and transmission protocols to obtain the desired features.

Currently, to reduce the signaling overhead for the uplink scheduling, a UE can use random access channel for uplink channel access (see Ref. [3] in case of machine-to-machine communication). More specifically, to apply for uplink transmission resources from eNB, a UE sends the scheduling request (SR) message through random access, which provides a channel access latency of 14 ms in the best case (see Fig. 1). However, if the random access fails, UE has to backoff before starting another random access. This significantly increases the channel access latency, which may not be desirable for the interactive and/or realtime applications that require fast reaction time to an event.

A lot of efforts has been given to improve the performance of random access, with particular attention to machine-to-machine communication (small low duty cycle packets). In Reference [4], a self optimization method for random access is proposed, which pre-computes the random access parameters (offline approach) to guarantee the channel access latency. Different backoff schemes for random access are analyzed in [5]. Based on this method, a dynamic window assignment method is designed, and as a result, the performance of random access is greatly improved compared to the fixed window scheme. Reference [6] suggests a fast collision resolution method for random access. This method utilizes the timing advance information of the stationary UEs to identify UEs and handle collisions between UEs.

To increase the success rate of the random access, we propose a TTI bundling scheme for the transmission of preambles. In the proposed method, a UE sends multiple preambles in several subsequent TTIs (i.e. subframes), by which a random access is successful if at least one of the preambles is correctly received by an eNB without a collision. We use a semi-markov model to analyze the random access with the TTI bundling scheme. With this model the access latency is derived as a function of the number of TTI bundles, and hence the optimal TTI bundling number which minimizes the access latency is found. The idea of TTI bundling is not new and introduced in LTE Rel. 8 to improve the uplink coverage for VoIP application [7]. The method allows a UE to send multiple VoIP packets through a bundle of several subsequent TTIs before receiving the HARQ from the serving eNB, which eliminates the latency caused by the packet retransmissions and thus improves the QoS for VoIP application [8].

The reminder of this paper is organized as follows. Section II provides a background information on the random access mechanism in LTE/LTE-A and presents the basic idea of the proposed TTI bundling scheme. Random access model with TTI bundling and the assumptions are explained in Section III. The method and how the optimal number of TTI bundles is calculated are detailed in Section IV. Section V presents the model validation and simulation results. Finally, Section VI provides the concluding remarks.

II. Random access in LTE/LTE-A and the TTI bundling scheme

Fig.1 depicts the procedure for the contention based random access in LTE/LTE-A [9]. Firstly, a UE sends a randomly selected preamble when the backoff counter becomes zero. Secondly, if the preamble is correctly received by eNB, the eNB sends the random access response (RAR) information during the RAR window. Thirdly, UEs decode the RAR message. If a UE finds its preamble identifier in the RAR message, it sends a L2/L3 message, for example scheduling request SR (L2 message) or RRC connection request (L3 message). It has to be noted that multiple UEs send L2/L3 message on the same resource if they select the same preamble in the first step (preamble collision), which results in that the L2/L3 message might not be correctly received by eNB. Finally, the eNB sends the contention resolution to acknowledge the correctly received L2/L3 message. Assuming that the time used to decode preamble, RAR, and L2/L3 message is 3ms, the total latency is around 14 ms in the best case (no preamble collision and no wireless channel error). However, if the initial random access fails, a UE has to backoff for certain time before starting a new random access attempt.[1]

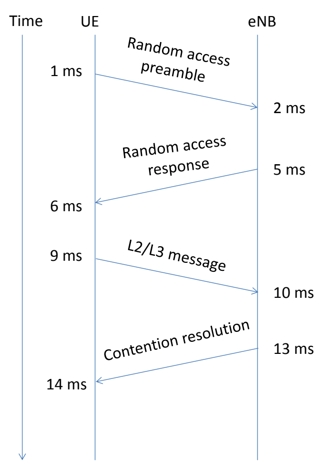

Fig. 1. Contention based random access in LTE

To improve the success rate in random access, we propose a TTI bundling scheme as shown in Fig.2. With the proposed scheme, a UE sends several randomly selected preambles in consecutive subframes to perform multiple random access attempts, which is referred to as TTI bundling for random access. Here we consider the multiple random access in consecutive subframes as a random access round. It is obvious that if one of these preambles in a random access round is correctly received by eNB and without collision, the random access round is successful, which eliminates the time that a UE has to wait for to start another random access when the initial random access fails. It seems that increasing the number of bundling TTIs yields higher successful probability for a random access round and thus reduces latency. However, this is not always true as the preamble collision rate increases with the number of bundling TTIs. This is because each UE has to trigger more transmissions when bundling larger number of TTIs, which in turn could reduce the success rate of a random access round and accordingly increases the latency. Therefore, the optimal selection of TTI bundling number in non-trivial.

[1]The maximum length of the backoff time is signaled by the eNB and can vary from 0 to 960 ms.

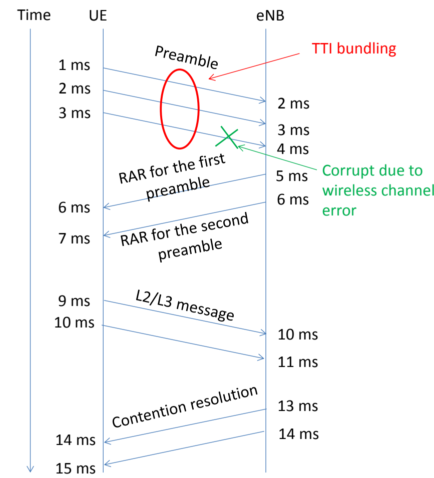

Fig. 2. Contention based random access with TTI bundling

III. Random Access With TTI Bundling Model and Assumptions

To find the optimal number of TTI bundles, a mathematical model is needed to analyze the random access mechanism. The existing random access models are based on a multichannel slotted ALOHA [10]- [11], where the delay and throughput of the random access procedure can be derived. However, they cannot be used to analyze the benefits of the TTI bundling scheme as they do not consider the backoff procedure after an unsuccessful transmission and the waiting state for a random access response.

We apply the Semi-Markov process to model the random access in LTE/LTE-A and to analyze both the regular random access as well as the random access with the TTI bundling. In the proposed model, the following assumptions are made.

Collision: We assume that each packet collides with a constant and independent probability. The assumption is feasible when the backoff window and number of UE are large [12].

Packet Transmission: Regardless of the packet size, all the packets in a UE's buffer can be sent by one uplink transmission. This assumption is reasonable as a UE can provide a buffer status report for the eNB scheduler through the L2/L3 message (see Fig. 1). Note that during the random access, it is possible that new packets are generated. With this assumption, these new packets are delivered with precedent packets. Therefore, when a UE re-starts at the initial state, there is no packet in its buffer. Moreover, due to the memoryless characteristic, the probability that a packet arrives in one subframe is not changed.

Traffic Model: The packet arrival is Poisson distributed. This is a simplifying assumption, for analytical purposes, but can be relaxed later (not considered in this paper).

Random Access opportunity: The random access channel is available in every subframe, which is related to random access resource configuration index 14 specifed by 3GPP [13].

Fig. 3 shows the proposed Semi-Markov process model for random access with TTI bundling, where there are three types of state: idle, backoff, and random access.

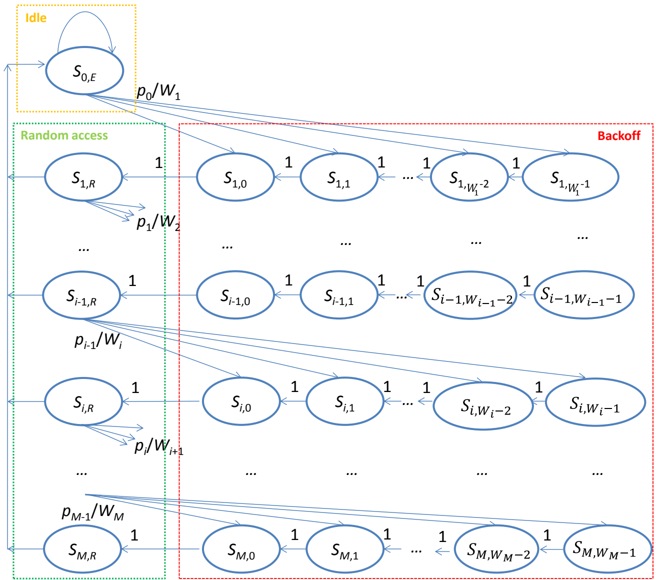

Fig. 3. Semi-Markov process model for random access with TTI bundling

- **Idle state** $S_{0,E}$ means there is no packet in the UE's buffer.
- **Backoff state** $S_{i,j}, i \in [1, M], j \in [0, W_i - 1]$, means that the UE in the ith backoff stage and the backoff counter is j, where M is the transmission limit and $W_i - 1$ is the maximum backoff counter size for ith backoff stage.
- **Random access transmission state** $S_{i,R}$, $i \in [1, M]$, means that a UE is performing multiple random access attempts, i.e., sending preambles or L2/L3 messages, and waiting for the response from eNB, such as RAR or contention resolution message.

A UE transfers between states as follows.

- When a UE is at state $S_{0,E}$, if a packet arrives in one subframe then UE selects a random backoff counter j over $[0, W_1 - 1]$ and transfers to state $S_{1,j}$ to start the first backoff. Otherwise it remains at state $S_{0,E}$. When a UE is at state $S_{i,j}, i \in [1, M], j \in [0, W_i - 1]$, it transfers to $S_{i,j-1}$ after 1ms.
- When a UE is at state $S_{i,R}$, $i \in [1, M-1]$, it transfers to state $S_{0,E}$ if a contention resolution indicating a successful random access is received. In contrast, it transfers to state $S_{i+1,j}$ (j is randomly selected over $[0, W_{i+1} - 1]$) to start another random access round if the $n, n \in [1, N]$, random accesses in one random access round all fail, where n is the number of bundling TTIs and N is the limit of TTI bundling number.
- When a UE is at state $S_{M,R}$, whether the random access attempts in this round are successful or not, it transfers to state $S_{0,E}$ when the random access round ends.

IV. Optimal TTI bundling for random access

Following the model presented in the Fig. 3, we denote the probability that a packet arrives during one subframe (1ms) as p_0, the state transition probability from $S_{0,E}$ to $S_{1,j}$, $j \in [0, W_1 - 1]$, is p_0/W_1. Similarly, we denote p_i, $i \in [1, M - 1]$, as the unsuccessful probability for the ith random access round, therefore the state transition probability from $S_{i,R}$, $i \in [1, M - 1]$ to $S_{i+1,j}, j \in [0, W_{i+1} - 1]$ is p_i/W_{i+1}.

Denoting $\pi_{i,j}$ as the stationary probability for state $S_{i,j}$ (the probability that UE remains at a given state), it can be calculated as:

$$\begin{cases} \pi_{1,W_1-1} = \pi_{0,E}\dfrac{p_0}{W_1} \\ \pi_{1,j} = \pi_{0,E}\dfrac{p_0}{W_1} + \pi_{1,j+1}, \quad j \in [0, W_1 - 2]. \\ \pi_{i,W_i-1} = \pi_{i-1,R}\dfrac{p_{i-1}}{W_i}, \quad i \in [2, M] \\ \pi_{i,j} = \pi_{i-1,R}\dfrac{p_{i-1}}{W_i} + \pi_{i,j+1}, \quad i \in [2, M], j \in [0, W_i - 2]. \end{cases} \tag{1}$$

With the first and second equations in equation system (1), we have:

$$\pi_{1,j} = (W_1 - j)\frac{p_0}{W_1}\pi_{0,E}, \quad j \in [0, W_1 - 1]. \tag{2}$$

$$\pi_{1,R} = p_0\pi_{0,E}. \tag{3}$$

By the use of the third and fourth equations in equation system (1), we get

$$\pi_{i,j} = (W_i - j)\frac{p_{i-1}}{W_i}\pi_{i-1,R}, \quad i \in [2, M], j \in [0, W_i - 1]. \tag{4}$$

$$\pi_{i,R} = p_{i-1}\pi_{i-1,R}, \quad i \in [2, M]. \tag{5}$$

The sum of the all state's stationary probabilities is 1, which yields:

$$\begin{aligned} 1 &= \pi_{0,E} + \sum_{i=1}^{M}\pi_{i,R} + \sum_{i=1}^{M}\sum_{j=0}^{W_i-1}\pi_{i,j} \\ &= \pi_{0,E} + \sum_{i=1}^{M}\prod_{j=0}^{i-1}p_j\pi_{0,E} + \sum_{i=1}^{M}\sum_{j=0}^{W_i-1}\frac{W_i - j}{W_i}\pi_{i,R} \\ &= \pi_{0,E} + \sum_{i=1}^{M}\prod_{j=0}^{i-1}p_j\pi_{0,E} + \sum_{i=1}^{M}\frac{W_i + 1}{2}\prod_{j=0}^{i-1}p_j\pi_{0,E}. \end{aligned} \tag{6}$$

Therefore, we have

$$\pi_{0,E} = \frac{1}{1 + \sum_{i=1}^{M}\prod_{j=0}^{i-1}p_j + \sum_{i=1}^{M}\frac{W_i+1}{2}\prod_{j=0}^{i-1}p_j}. \tag{7}$$

Now let us calculate the state transition probabilities. Assuming the packet arrives following Poisson distribution with arrival rate λ, the probability that a packet arrives in one subframe is $p_0 = 1 - e^{-\lambda}$.

One random access round is unsuccessful if all the random access attempts in this round are unsuccessful, therefore $p_i = p_{F,i}^n$ where $p_{F,i}$ is the unsuccessful probability for one random access in the ith random access round.

An unsuccessful random access is caused by erroneous transmission for the preamble or the unsuccessful delivery of the L2/L3 message. More specifically, the unsuccessful delivery of the L2/L3 message is also caused by two subcases: (1) collision of the preamble, which leads to the failure for the L2/L3 message delivery, and (2) the L2/L3 is corrupted due to wireless channel error. In the later case, the preamble is correctly received by the eNB without any collisions, however the L2/L3 cannot be successfully decoded by the eNB due to the wireless channel error. With the above analysis, we have

$$p_{F,i} = p_{E,i} + (1 - p_{E,i})p_c + (1 - p_c)(1 - p_{E,i})p_{ES}^{N_{HARQ}}. \quad (8)$$

In the above equation p_c is the collision rate for a preamble; $p_{E,i}$ is the error probability caused by wireless channel for a preamble in the ith random access round; p_{ES} is the error rate to send the L2/L3 message which contains SR and N_{HARQ} is the maximum number of HARQ transmissions. Since the L2/L3 message containing SR is of very small size, therefore its error rate p_{ES} is very small (less than 0.1) and hence $p_{ES}^{N_{HARQ}} \approx 0$ considering that N_{HARQ} is usually larger than 2. With this result, we have $p_{F,i} \approx p_c + p_{E,i} - p_c p_{E,i}$.

From the perspective of one UE, collision happens when there are other UEs selecting the same preamble, therefore

$$p_c = \sum_{i=1}^{N_u - 1} \binom{N_u - 1}{i} \tau^i (1 - \tau)^{N_u - 1 - i} (1 - (1 - \frac{1}{N_p})^i). \quad (9)$$

In the above equation N_u is the total amount of UE; τ is the probability that a UE sends a preamble in one subframe; N_p is the number of available preambles for random access.

Now let us calculate the state holding time for this Semi-Markov process model. It is obvious that the state holding time for $S_{0,E}$ and $S_{i,j}, i \in [1, M], j \in [0, W - 1]$ is 1ms.

For the UE at $S_{i,R}, i \in [1, M]$, the calculation for state holding time is less obvious. We denote the duration that starts at the end of a preamble transmission and ends at the time instant when receiving the RAR message for that preamble as T_{RAR} and the time used to decode the RAR message as T_D. Therefore, the SR message is sent $T_{RAR} + T_D$ ms after the preamble's transmission if the RAR message is received (no wireless error for the transmitted preamble). As stated above, when the UE is at $S_{i,R}, i \in [1, M]$, the state transition happens when one random access is successful or all the random access in one random access round fail. Hence, we calculate the state holding time for three cases:

1) **The jth, $j \in [1, n]$, random access in the ith, $i \in [1, M]$, random access round is successful.**

   The probability for the first case $p_{i,j}^S$ is

   $$\prod_{k=1}^{i-1} p_k p_{F,i}^{j-1}(1 - p_{F,i}), i > 1 \quad (10)$$

   or

   $$p_{F,1}^{j-1}(1 - p_{F,1}), i = 1. \quad (11)$$

   When a random access succeeds, the UE transfers to the initial state $S_{0,E}$ after decoding the contention resolution message. Denoting T_{CR} as the average duration which starts at time instant when a UE sends the SR message and ends at the time instant when a UE decodes the contention resolution message, the state holding time for the first case is $T_{i,j}^S = j + T_{RAR} + T_D + T_{CR}$

2) **None of the random access in the ith, $i \in [1, M]$, random access round is successful, and the UE receives the RAR message from eNB for the last transmitted preamble.**

   In this case, a UE sends the L2/L3 message. However, as the random access is unsuccessful, it cannot receive the contention resolution message. This UE will transfer to the initial state $S_{0,E}$ when the contention resolution timer expires.
   Therefore, the state holding time for the second case is $T^R = n + T_{RAR} + T_D + T_{timer}$ where T_{timer} is the duration for contention resolution timer. The probability for this second case when $i > 1$ is

   $$p_i^R = \prod_{k=1}^{i-1} p_k p_{F,i}^{n-1} P_{RAR,i}. \quad (12)$$

   When $i = 1$, the probability for the second case is

   $$p_1^R = p_{F,1}^{n-1} P_{RAR,1}. \quad (13)$$

   where $P_{RAR,i}, i \in [1, M]$, is the probability that the collision happens for a random access in the ith random access round and UE receives the RAR message. The $P_{RAR,i}$ is calculated by equation (14). In equation (14) $r_{i,j+1}$ is the detection rate for the preamble in the ith random access round when $j + 1$ UEs (one UE plus j contending UEs) send the same preamble. If $r_{i,j+1} \approx 1$ for $j \geq 1$, i.e., a preamble can mostly be detected when it is sent by multiple UEs, then $P_{RAR,i} \approx p_c$.

3) **None of the random access in the ith, $i \in [1, M]$, random access round is successful, and the UE does not receive RAR for the last random access.**

   In this case, after sending the last preamble the minimum time that a UE will stay at state $S_{i,R}$ is T_W, i.e., the minimum state holding time for state $S_{i,R}$ is T_W, where T_W is the duration which starts at the time instant when a UE sends a preamble and ends at the last subframe of RAR window. Besides that, if this UE has received the RAR for a random access in this round (not the last random access in this round), it cannot transfer to the initial state $S_{0,E}$ until the contention resolution timer ends, which may extend the state holding time. Denoting the time instant when a UE sends the jth, $j \in [1, n]$, preamble as t_j and assuming the RAR is received by UE for this preamble, its contention resolution timer ends at $t_j + T_{RAR} + T_D + T_{timer}$. Therefore, for a preamble transmission which can extend the UE's state holding time at state $S_{i,R}, i \in [1, M]$, its transmission time instant t_j should satisfy the following the condition

   $$t_j + T_{RAR} + T_D + T_{timer} > t_n + T_W \quad (15)$$

   where t_n is the time instant when a UE sends the last preamble.
   Since $t_n = t_j + (n - j)$ where n is the index of the last preamble sent in a random access round and j is the index of the jth preamble sent in a random access round, the above calculation is rewritten as

   $$j + T_{RAR} + T_D + T_{timer} > n + T_W. \quad (16)$$

   We can find the minimum preamble index j satisfying the above formula, which is denoted as z.
   Therefore, when a UE can transfer from state $S_{i,R}$ to another state is determined by the status whether RAR is received for the kth preamble, $k \in [z, n-1]$. Specifically, a UE transfers from state $S_{i,R}$ to another state when the kth preamble's contention resolution timer ends if the following

$$P_{RAR,i} = \sum_{n=1}^{N_u-1} \binom{N_u-1}{n} \tau^n (1-\tau)^{N_u-1-n} \sum_{j=1}^{n} \binom{n}{j} (\frac{1}{N_p})^j (1-\frac{1}{N_p})^{n-j} r_{i,j+1}. \quad (14)$$

conditions hold: (a) RAR is received for this preamble; (b) the random accesses are unsuccessful for the preambles sent before it; (c) no RAR messages are received for the preambles sent after it.

Accordingly, the state holding time is

$$T_{i,k}^N = k + T_{RAR} + T_D + T_{timer} \quad (17)$$

and its probability $p_{i,k}^N$ is

$$\prod_{j=1}^{i-1} p_j p_{F,i}^{k-1} P_{RAR,i} P_{NRAR,i}^{n-k}, \;\; i > 1 \quad (18)$$

or

$$p_{F,1}^{k-1} P_{RAR,1} P_{NRAR,1}^{n-k}, \;\; i = 1 \quad (19)$$

where $P_{NRAR,i}, i \in [1, M]$, is the probability that no RAR is received in one random access of the ith random access round. The RAR is not sent to a UE if the transmitted preamble sent by one UE (or multiple UEs) is not correctly detected by eNB, therefore we have equation (20).

It is also possible that no RAR is received for all the random accesses whose index are larger than z. Then the state holding time is $T_{i,n}^N = n + T_W$ and its probability $p_{i,n}^N$ is

$$\prod_{j=1}^{i-1} p_j p_{F,i}^{z-1} P_{NRAR,i}^{n-z+1}, \;\; i > 1 \quad (21)$$

or

$$p_{F,i}^{z-1} P_{NRAR,i}^{n-z+1}, \;\; i = 1 \quad (22)$$

If $r_{i,j+1} \approx 1$ for $j \geq 1$, i.e., a preamble can mostly be detected when it is sent by multiple UEs, we have $P_{NRAR,i} \approx (1-p_c)p_{E,i}$.

With the above results, the average holding time $T_{i,R}, i \in [1, M]$, for state $S_{i,R}$ is

$$T_{i,R} = \frac{\sum_{j=1}^n p_{i,j}^S T_{i,j}^S + p_i^R T^R + \sum_{j=z}^n p_{i,j}^N T_{i,j}^N}{\sum_{j=1}^n p_{i,j}^S + p_i^R + \sum_{j=z}^n p_{i,j}^N}. \quad (23)$$

When a UE is at state $S_{i,R}, i \in [1, M]$, the average duration which is used for sending preambles is

$$T_{i,TX} = \frac{\sum_{j=1}^n p_{i,j}^S j + p_i^R n + \sum_{j=z}^n p_{i,j}^N n}{\sum_{j=1}^n p_{i,j}^S + p_i^R + \sum_{j=z}^n p_{i,j}^N}. \quad (24)$$

Therefore, the proportion of time that a UE is sending a preamble, i.e., the probability that a UE sends a preamble in one subframe, is

$$\tau = \sum_{i=1}^M \frac{\pi_{i,R} T_{i,TX}}{T} \quad (25)$$

where

$$T = \pi_{0,E} + \sum_{i=1}^M \sum_{j=1}^{W-1} \pi_{i,j} + \sum_{i=1}^M \pi_{i,R} T_{i,R}. \quad (26)$$

is the average holding time for all the states.

It can be seen that equations (25) and (9) comprise a equation system with two unknowns p_c and τ, which can be solved by the use of numerical method.

Provided that the $i \in [1, M]$ random access round is unsuccessful, then the duration that UE stays at state $S_{i,R}$ is the latency introduced by this random access round. Denoting the latency in this case as d_i', it is calculated by

$$d_i' = \frac{p_i^R T^R + \sum_{j=z}^n p_{i,j}^N T_{i,j}^N}{p_i^R + \sum_{j=z}^n p_{i,j}^N}. \quad (27)$$

If a random access is successful at the first random access round, no latency is caused by the subsequent random access rounds. Therefore, we have $T_1' = 0$. However, if a random access is successful at the ith random access round, the latency caused by the precedent unsuccessful random access is

$$T_i' = \sum_{j=1}^{i-1} d_j', \;\; i \in [2, M]. \quad (28)$$

Now let us calculate the access latency for random access which is defined as the duration that starts at the time instant when a UE wants to trigger a random access and ends at the time when that UE receives a contention resolution message indicating the random access is successful. Assuming the random access is succeed in the jth transmission of the ith random access round, the access latency includes (1) the duration of $\sum_{j=1}^i \frac{W_j}{2}$ which is used for backoff, (2) the duration of $T_{i,j}^S$ which is the time used for the successful random access in the current round, and (3) the latency T_i' which is used for the precedent unsuccessful random access rounds.

With above analysis, the average channel access latency caused by random access is calculated by

$$d = \sum_{j=1}^n \frac{p_{F,1}^{(j-1)}(1-p_{F,1})}{1-\prod_{i=1}^M p_i} (\frac{W_1}{2} + T_{1,j}^S) + \sum_{i=2}^M \sum_{j=1}^n \frac{\prod_{k=1}^{i-1} p_k p_{F,i}^{(j-1)}(1-p_{F,i})}{1-\prod_{i=1}^M p_i} (\sum_{k=1}^i \frac{W_k}{2} + T_{i,j}^S + T_i'). \quad (29)$$

The optimal TTI bundling number which minimizes the access latency is

$$\begin{aligned} &\arg\min_n \quad d \\ &\text{subject to} \quad n < N \end{aligned} \quad (30)$$

where N is the limit of TTI bundling number. The L2/L3 message is sent $T_{RAR} + T_D$ ms after the first preamble's transmission if the preamble is correctly received by eNB. As a UE cannot send a L2/L3 message as well as a preamble at the same time, the maximum bundling TTI number N should be no larger than $T_{RAR} + T_D$. Since we do not have a closed form of d in the term of n, therefore the above optimization problem can only be solved by exhaustive search.

$$P_{NRAR,i} = \sum_{n=0}^{N_u-1} \binom{N_u-1}{n} \tau^n (1-\tau)^{N_u-1-n} \sum_{j=0}^{n} \binom{n}{j} (\frac{1}{N_p})^j (1-\frac{1}{N_p})^{n-j} (1-r_{i,j+1}). \quad (20)$$

For some power constrained devices, the optimal TTI bundling number obtained from the above equation might be adjusted to achieve a trade-off between the power consumption and the latency constraint.

V. Results

A. Simulation Parameters

The simulation parameters are shown in Table I. Here the total number of available preambles is assumed to be 20. While the total number of available preambles in LTE is 64, this assumption allows us to study the behaviour of the proposed method under a high collision rate regime. In addition, this assumption is also related to the scenario where the available preambles are divided into multiple (non-)overlapping sets to control the collision across multiple applications, e.g. one for human type communication and the other for the machine type communication [14]. Two packet arrival rates are considered: $\lambda = 1/100$ and $1/50$ packet/ms, which correspond to the interactive and/or realtime application scenarios (e.g. realtime machine-to-machine communication and online interactive gaming). In case of no collision, the preamble detection rate is assume to be $1 - \frac{1}{e^i}$ similar to [3], where $i \in [1, M]$ indicates the ith preamble transmission. When a preamble are sent by multiple UEs, it can always be correctly decoded, $r_{i,j+1} \approx 1$ for $j \geq 1$. This assumption is quite typical in LTE. If a preamble is sent by two UEs, two peaks appear at the eNB side. The probability that neither peak can be decoded by eNB is relatively low.

TABLE I. Simulation parameters

Parameter	Value	Description
T_{RAR}	5 ms	Time elapsed between the preamble transmission and the reception of RAR message
T_D	3 ms	Time used to decode a RAR message
T_{CR}	8 ms	Time elapsed between the transmission of an SR and the reception of a contention resolution message
T_W	15 ms	Time elapsed between the preamble transmission and the last subframe of RAR window
T_{timer}	24 ms	Duration of contention resolution timer
M	5	Transmission limit for random access
N_p	20	Number of available preambles for random access
N	8	Max TTI bundling number
W_i, $i \in [1, 5]$	30	Backoff Window Size

B. Model Validation

To validate the proposed method, we compare the simulation results with the analytical results obtained using equation (29) in Fig. 4. To validate the applicability of the modelling approach for both regular random access and random access with TTI bundling, simulations are performed with 1 and 2 TTI bundles (i.e. $n = 1$ and $n = 2$). As shown in Fig. 4, the analytical results match the simulation results validating the modelling approach.

C. Optimal TTI bundles and The Induced Latency

Fig. 5 depicts the optimal TTI bundling number under different numbers of UE and packet arrival rates. It can be seen that the optimal TTI bundling number non-increases as the number of UE and the packet arrival rates increase. The reason for this phenomenon is that, the preamble collision rate grows with the number of UEs and the packet arrival rate. Therefore, when the number of UEs and/or packet arrival rate become large, a UE should bundle smaller (or same) number of TTI to reduce the collision rate. Moreover, we also find that the TTI bundling number scales with the packet arrival rate in that higher packet arrival rate limits the number of bundles. This is reasonable since the packet collision rate increases with packet arrival rate. Therefore, smaller TTI bundling should be used to control the collision.

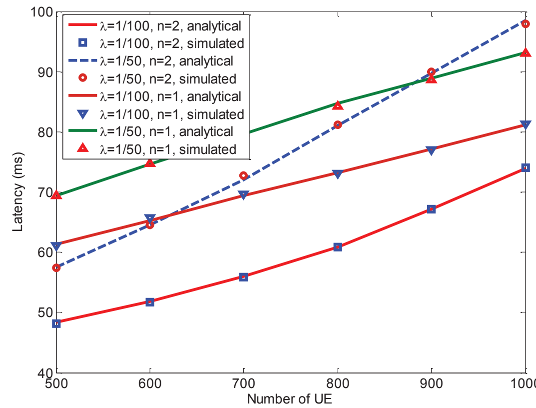

Fig. 4. Comparison of simulation and analytical results

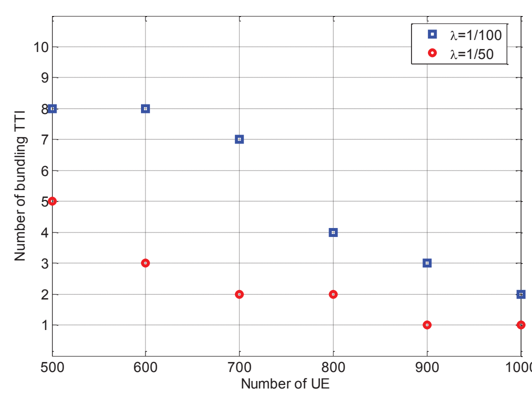

Fig. 5. Optimal number of bundling TTI

Fig. 6 compares the latency obtained using the results shown in Fig. 5 to the latency without the TTI bundling scheme. It can be inferred that the latency is greatly reduced until the average number of simultaneous channel access is less than the number of available preambles, e.g. with $\lambda = 1/50$ and 1000 users, there will be 20 random channel access attempts per subframe. Thus, the achievable latency gain is also dependent on the number of available preambles. From the figures, it can be seen that there are two regimes. In the first regime, the preamble collision rate is not very high, and thus bundling multiple TTIs greatly increases the successful rate of the random access procedure. For example when $\lambda = 1/100$, $n = 1$, and $N_u = 500$,the collision rate is 0.22 and the first random access round successful rate is 0.50 and the resulted latency is 61ms. When the TTI bundling number n increases to 8, though the collision rate increases to 0.36, the first random access round successful approximately equals 1 which reduces the latency to 33ms as shown in Fig.6. In the second regime, the preamble collision rate is high, thus bundling multiple TTIs further increases the preamble collision rate. As a result, the random access success rate could not be improved. For example when $\lambda = 1/50$, $n = 1$ and $N_u = 800$, the collision rate is 0.48; the first random access round successful rate is 0.33 and the latency is 85ms. When the TTI bundling number n increases to 2, the preamble collision rate jumps to 0.65 and the successful rate for the first random access round only increases to 0.45 which slightly reduces the latency to 81ms (see Fig.6). Therefore, to reduce the random access channel access latency when the preamble collision rate is high, the TTI bundling is applicable only when more preambles and/or more PRACH resources can be (dynamically) allocated to lower the collision rate.

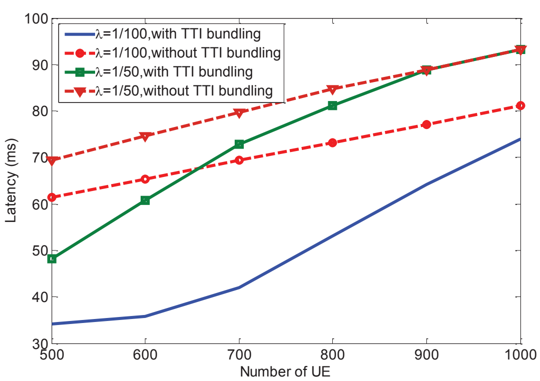

Fig. 6. Latency comparison with and without TTI bundling

Fig.7 shows the channel access latency under different TTI bundling number when $\lambda = 1/100$ and $1/50$, and $N_u = 500$. It can be observed that the latency can increase if the number of bundles is not optimally selected (see the curve for $\lambda = 1/50$).

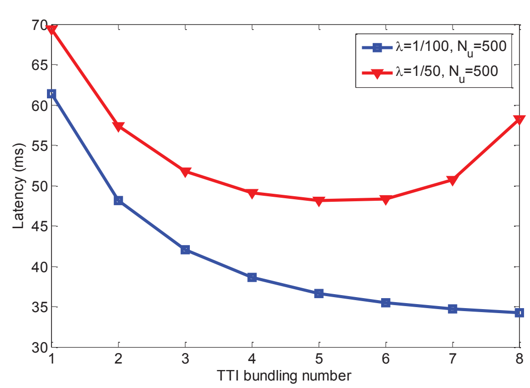

Fig. 7. Latency under different number of TTI bundles

VI. Conclusion

In this paper, we propose a TTI bundling method to reduce the uplink channel access latency in LTE/LTE-A. With TTI bundling, a UE sends one or multiple preamble(s) in one random access round to increase the random access success rate. To find the optimal TTI bundling number which minimizes the channel access latency, we apply a Semi-Markov model to formulate the access latency as a function of TTI bundling number. The proposed model is validated through simulations and performance of the proposed TTI bundling scheme is also evaluated. We find that the TTI bundling scheme can significantly reduce the uplink latency when the preamble collision rate due to simultaneous channel access is not high while the latency improvement vanishes as the preamble collision rate increases.

Acknowledgement

The research leading to these results has received funding from the European Research Council under the European Community Seventh Framework Programme (FP7/2014-2017) grant agreement 612050 for FLEX project.

References

[1] N. Nikaein, M. Laner, K. Zhou, P. Svoboda, D. Drajic, M. Popovic and S. Krco, "Simple Traffic Modeling Framework for Machine Type Communication," 2013.

[2] Ericsson, "5G Radio Access: Research and Vision," White paper, Ericsson, 2013.

[3] 3GPP TR 37.868, "RAN Improvements for Machine-type Communications," V. 11.0.0., Oct., 2011.

[4] M. Amirijoo, F. Gunnarsson, F. Andren, "Random Access Channel Self-Optimization in 3GPP LTE ," *IEEE Transactions on Vehicular Technology*, No.99, pp.1, Feb. 2014.

[5] Jun-Bae Seo, Leung, V.C.M, "Back to Results Design and Analysis of Backoff Algorithms for Random Access Channels in UMTS-LTE and IEEE 802.16 Systems," *IEEE Transactions on Vehicular Technology*, vol. 60, No. 8, pp. 3975 - 3989, Oct. 2011.

[6] K.S. Ko, M.J.Kim, *etal*., "A Novel Random Access for Fixed-Location Machine-to-Machine Communications in OFDMA Based Systems," *IEEE Commun. Lett.*, vol.16,No.9, pp.1428-1431, Sept. 2012.

[7] 3GPP TDoc R2-072630, HARQ operation in case of UL Power Limitation, Ericsson, June 2007.

[8] R. Susitaival, M.Meyer, "LTE coverage improvement by TTI bundling," *Proc. IEEE VTC Spring*, April, 2009, pp.1-5.

[9] 3GPP TS 36.321, Evolved Universal Terrestrial Radio Access (E-UTRA); Medium Access Control (MAC)," V.12.0.0, Jan., 2014.

[10] E.Osipov, L.Riliskis, A. Eldstal-Damlin, M.Burakov, M.Nordberg, Min Wang, "An improved model of LTE random access channel", *Proc. IEEE VTC Spring*, April, 2014, pp.1-5.

[11] Michael Burakov, "An LTE Random Access Channel Model for Wireless Sensor Network Applications", MSc Thesis, Lules university of technology, 2012.

[12] Giuseppe Bianchi, "Performance Analysis of the IEEE 802.11 Distributed Coordination Function", *IEEE Journa on Selected Areas in Communications*, Vol. 18, No. 3, pp. 535-547.

[13] 3GPP TS 36.211, "Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Physical Channels and Modulation," V 12.0.0, Dec., 2013.

[14] 3GPP TS 36.300, "‘Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN)," V. 10.4.0, Jun. 2011.