IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 9, 2019, accepted April 25, 2019, date of publication May 17, 2019, date of current version May 29, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2917618

Performance Study and Enhancement of Access Barring for Massive Machine-Type Communications

JOSÉ-RAMÓN VIDAL[1], LUIS TELLO-OQUENDO[2], VICENT PLA[1], AND LUIS GUIJARRO[1]
[1]Departamento de Comunicaciones, Universitat Politècnica de València, 46022 Valencia, Spain
[2]College of Engineering, Universidad Nacional de Chimborazo, Riobamba 060108, Ecuador

Corresponding author: José-Ramón Vidal (jrvidal@upv.es)

This work was supported in part by the Ministerio de Ciencia, Innovación y Universidades (MCIU), Agencia Estatal de Investigación (AEI) y Fondo Europeo de Desarrollo Regional (FEDER), UE, under Grant PGC2018-094151-B-I00, and in part by the ITACA Institute under Grant Ayudas ITACA 2019.

ABSTRACT Machine-type communications (MTC) is an emerging technology that boosts the development of the Internet of Things by providing ubiquitous connectivity and services. Cellular networks are an excellent choice for providing such hyper-connectivity thanks to their widely deployed infrastructure, among other features. However, dealing with a large number of connection requests is a primary challenge in the cellular-based MTC. Severe congestion episodes can occur when a large number of devices try to access the network almost simultaneously. Extended access barring (EAB) is a congestion control mechanism for the MTC that has been proposed by the 3GPP. In this paper, we carry out a thorough performance analysis of the EAB and show the limitations of its current specification. To overcome these limitations, we propose the two enhanced EAB schemes: the combined use of the EAB and access class barring, and the introduction of a congestion avoidance backoff after the barring status of a UE is switched to unbarred. It is shown through extensive simulations that our proposed solutions improve the key performance indicators. A high successful access probability can be achieved even in heavily congested scenarios, the access delay is shortened, and, most importantly, the number of required preamble retransmissions is reduced, which results in significant energy savings. Furthermore, we present an accurate congestion estimation method that solely relies on the information available at the base station. We show that this method permits a realistic and effective implementation of the EAB.

INDEX TERMS Machine-to-machine communications, performance analysis, radio access networks, 5G mobile communication.

I. INTRODUCTION

The tremendous potential of machine-type communications (MTC) to offer ubiquitous connectivity among intelligent devices makes the Internet of Things (IoT) possible. Cellular networks are an excellent choice to provide such hyper-connectivity thanks to the extensively deployed infrastructure and numerous advantages such as the reduced deployment costs, security, management and QoS, among other features that 3GPP cellular technologies offer [1]–[3].

Massive machine-type communications (mMTC) is one of the main service types envisioned in 5G, along with ultrareliable and low-latency communications (URLLC) and enhanced mobile broadband (eMBB) [4], [5]. Services in mMTC are characterized by a vast number of connected devices typically transmitting a relatively low volume of delay-tolerant data.

When an MTC device (named UE herein) wants to access the network, it must first obtain some configuration parameters from the network, such as the predefined time/frequency resources where the random access attempts are allowed. Each instance of these resources, in which an access attempt can be made, is called random access opportunity (RAO) and the sequence of RAOs constitutes the random access channel (RACH). The base station (BS) broadcasts the configuration information periodically through the Master Information

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

Block (MIB) and the System Information Blocks (SIBs). Then, the UEs perform a random access procedure in which the RACH is used to signal the connection request [6], [7].

The random access procedure consists in a four-message handshake contention-based procedure. First, the UEs sends Msg1 in the next RAO using a randomly chosen preamble from a pool of available preambles. Msg1 is detected by the BS if the preamble is transmitted by only one UE in the current RAO; if not, a collision occurs. For each detected preamble, the BS sends a random access response (RAR) message, Msg2, which includes one uplink grant, from a limited number of grants available. Msg2 is used to assign time-frequency resources to the UEs for the transmission of the connection request. UEs that received an uplink grant send their connection request message, Msg3, using the resources specified by the BS. Finally, the BS responds to each Msg3 transmission with a contention resolution message, Msg4. The interested reader is referred to [6], [8]–[10] for further details.

In mMTC, a fundamental problem is the efficient management of network resources in overload situations caused when a large number of UEs simultaneously attempt to access the network. The 3GPP proposes the extended access barring (EAB) mechanism for congestion control, which selectively restricts the access attempts of the UEs that are configured for EAB. In 5G, UEs configured for EAB are considered to be configured for delay-tolerant services [11]. Furthermore, in Release 13 the 3GPP specified a new radio interface called Narrowband Internet of Things (NB-IoT) that is optimized for MTC traffic. The only congestion control mechanism available in NB-IoT is called access barring (AB) [9], [12] and, for the purposes of this paper, AB is effectively equivalent to EAB; in the remainder of this paper, EAB is used to refer to both mechanisms. Therefore, EAB is the primary congestion control mechanism for mMTC.

Each UE configured for EAB is allocated an access class (AC) in the range 0–9. When the network determines that it is appropriate to apply EAB, it barres all UEs in a given set of ACs, and broadcasts a SIB type 14 (SIB14) containing a 10 bit barring bitmap. The barring is of simple on/off type, where access to each AC is either allowed or barred. This allows the network to control the load by barring different ACs with a granularity of 10% of the device population [13]. EAB may be effective whenever the congestion occurs sparingly and during short periods of time (in the order of several seconds), which is the typical behavior of mMTC as described in [14].

When the operator determines there is congestion in a given BS, EAB is enabled and the network broadcasts necessary information to provide EAB control for UEs. This information basically consists of that contained in the SIB14 and the system information change parameter contained in the paging messages [15]. The procedure followed by UEs to acquire the SIB14 and their behavior upon its acquisition is specified in [9].

The SIB14 contains a bitmap of barred ACs. The BS broadcasts messages containing the SIB14s with a period T_{SIB14} [11]. Every time the bitmap has to be changed, the BS notifies it to the UEs through a system information change parameter contained in the paging messages [15]. The paging messages are sent at specific paging occasions (POs) within a paging cycle of duration T_P. When a UE reads a paging message with system information change set to on, it reads the next occurrence of SIB14. Each UE can use only a subset of all available POs. A UE calculates its POs from its local identifier, as specified in [15]. By doing this, the POs of the different UEs are distributed homogeneously throughout a paging cycle. As a result, after a system information change, the UEs will be evenly distributed between the SIB14 occurrences within the following paging cycle to acquire the new EAB information. The purpose of this is to reduce the number of UEs that will perform a simultaneous access attempt after the barring of their class has been lifted.

The criteria for enabling or disabling EAB are not included in the technical specifications of 3GPP, nor is the procedure for switching the barring status of each AC while EAB is enabled. In this paper, we use as a baseline EAB along with the method proposed and evaluated in [16] for enabling and disabling EAB, and for switching the barring status of each AC. The same method was used for the performance analysis presented in [17].

This method relies on two congestion coefficients, CC_{500} and CC_{1000}, that measure the level of congestion over a moving window of 500 ms and 1000 ms, respectively; further details can be found in Section VI.

When the value of CC_{1000} exceeds 0.4, EAB is enabled and all ACs except AC 0 are barred. Then, the barring of the other classes is lifted sequentially, one by one, each time CC_{500} falls below 0.4. To guarantee that all the UEs in the cell have time enough to acquire the updated bitmap, the minimum time between two consecutive updates is set to $\max(T_P, T_{\text{SIB14}})$.

It is important to note, however, that this method cannot be directly implemented. The congestion coefficients must be computed by the BS, and to do so it would require to know the number of preambles transmitted at each RAO. However, when collisions occur the BS cannot know the exact number of UEs that transmitted a preamble. One of the main contributions of this paper is the proposal of an estimation method that allows the congestion coefficients to be calculated by relying only on the information available at the BS.

In this paper, we conduct a thorough performance analysis of EAB by discrete-event simulations. We show that a limiting factor of the current specification of EAB is that, when the barring of an AC is lifted, UEs belonging to this AC initiate their access procedure in bursts of periodicity T_{SIB14}. Each burst is caused by those UEs whose POs, and consequently their readings of the paging message, occur between two consecutive broadcasts of SIB14. These traffic bursts cause

many preamble collisions during the first RAOs, deteriorating the overall network performance.

We propose and evaluate two alternative enhancements of EAB to relieve the congestion arising after the barring of an AC is released. Firstly, we assess the effect of using ACB [9] in conjunction with EAB. For this, when a barred AC is released, the UEs in this class perform an ACB check before accessing the RACH. Secondly, we propose and evaluate a simpler scheme consisting in a collision-avoidance (CA) backoff. In this scheme, when a barred AC is released, the UEs in this class wait for a random backoff period before accessing the RACH. We show that both proposals reduce the traffic bursts and enhance the key performance indicators (KPIs).

The main contributions of this paper are as follows. First, a comprehensive performance analysis of EAB is carried out by extensive and comprehensive computer simulations. Our simulation model accurately reproduces the behavior and characteristics, as specified by the 3GPP, of the elements involved in the random access procedure and EAB (UEs, BS and RACH). This analysis reveals that the current specification of EAB has significant limitations and provides a detailed insight into their causes and consequences. In the limited existing literature on this topic (see the next section for details), the available performance studies ignore some important details of the system (e.g., which is the information available at the BS) and do not assess the performance in terms of delay, or this assessment is limited to the mean delay.

Our performance evaluations are based on the three main KPIs for MTC [14]: the probability to successfully complete the random access procedure, the number of necessary preamble transmissions to the random access procedure, and statistics of the access delay of the successful accesses (in addition to the mean, several percentiles are given). These KPIs are in conformance with the 3GPP directives to assess the efficiency of the random access procedure with MTC.

Second, we propose two easily implementable modifications of EAB to overcome the identified limitations. The effectiveness of the proposed solution is illustrated by means of an exhaustive performance study. Among the limited existing research on EAB, there are very few proposals aiming to improve its performance. To our knowledge, the two of such proposals are [13], [18]. Our proposed improvements differ from these in that ours consistently improve the performance in all the KPIs, whereas the other proposals focus exclusively on the success probability and ignore the impact on the other KPIs; please, refer to the next section for further details.

Finally, we propose an estimation method that allows the practical implementation of EAB by relying only in the information available at the BS. The criteria we used for setting the barring status of each AC is the same as in the most complete performance studies of EAB so far [13], [18]. However, in these previous studies no details are given about how the BS calculates the required congestion coefficients for the implemented criteria.

The rest of the paper is organized as follows. In Section II, we conduct a review of the literature regarding the studies that evaluate the performance of EAB and some proposed algorithms for its implementation. In Section III, we analyze the performance of EAB and investigate its limitations. In Section IV, we describe and evaluate two enhanced EAB schemes which overcome the former limitations. In Section V, we evaluate quantitatively the improvement obtained by both proposals in terms of KPIs. In Section VI, we propose an evaluate a method to implement EAB realistically by estimating the CC_W from the information available at the BS. Finally, in Section VII we draw the conclusions.

II. RELATED WORK

The performance of EAB has been analyzed in a few studies such as [13], [17]–[20]. However, the study of EAB has received very little attention so far in comparison with pre-existing ACB. Besides, some literature uses the name EAB to refer to the application of ACB in MTC scenarios [21]–[24]. Thus, the actual number of studies on EAB is even lower than it may seem.

Some of these studies are restricted to assessing the performance of EAB; and although they usually confirm the existing limitations, no solutions are proposed to improve the performance of EAB [17], [19], [20]. Besides, most of these studies tended to focus on the theoretical performance of EAB, whereas less attention was paid to its implementation. For example, the implementation of EAB proposed in [16], [17] is based on two congestion coefficients which cannot be directly calculated by the BS, and in [19], [20] no details are given regarding the procedure that is followed to activate and deactivate EAB control, or to set ACs barring status.

One of the earliest performance studies of EAB is that in [13]. The results there show that, even with the longest paging cycle, the performance of EAB can be rather poor in highly loaded scenarios. As a solution, the authors propose a selective paging solution in which users are notified of a system information change over several paging cycles. A significant impact on the access delay can be expected with this solution. However, the reported results only included the access success probability and no statistics about access delay are provided.

In [17], a complete analytical model for the evaluation of EAB is presented. The model allows obtaining the access success probability, the collision probability and the mean access delay in an efficient and rather accurate manner. The results show that to obtain an acceptable performance in highly loaded scenarios, the ratio between the duration of the paging cycle and the SIB14 period (T_P/T_{SIB14}) must be above a certain threshold. However, other than selecting sufficiently long paging cycles, no solutions are proposed to improve the performance of EAB.

The authors of [18] propose an improvement to EAB that is implemented in the UE side. The proposed method aims to enable UEs to detect the congestion arising after the barring of an AC is lifted. In that situation, after a number of failed access attempts, a UE may decide (with certain probability) to defer new attempts until its next paging occasion. This method is similar to ours in the sense that randomness is introduced in the UE side to alleviate the congestion that occurs after an AC is unbarred. However, whereas our approach aims at avoiding collisions, this method is basically reactive and requires a minimum level of collisions to work properly. This would go against one of the design principles of EAB, which focuses on reducing collisions (to save battery) rather than reducing access delay [20].

Non-orthogonal multiple access (NOMA) is expected to provide important spectral efficiency gains in 5G. Although NOMA is more commonly applied to downlink transmissions, it has also been shown to be effective for uplink transmissions [25], [26].

In [26], [27], the authors propose a new uplink multiple access scheme for mMTC based on NOMA. This scheme will eliminate the need of the initial random access procedure, thus reducing significantly not only the signaling congestion in the RACH but also the access delay.

While this approach is undoubtedly promising for cellular mMTC, it requires to completely overturn the current specifications of the radio interface. Moreover, there are still a number of practical challenges that must be addressed before this technology can be used in real systems, such as channel estimation and synchronization among UEs. All these aspects are beyond the scope of this paper, which aims instead at enhancing a congestion control mechanism (EAB) in the framework of the current random access procedure specified by the 3GPP.

In [25], [28], [29], a power-domain NOMA is applied to slotted multichannel ALOHA to increase the throughput. The throughput gain is achieved by subdividing each channel into power levels. The receiver can then decode several transmission in the same channel, if their received power levels are sufficiently separated.

It is known that the random access procedure employed in the RACH is based on slotted multichannel ALOHA, in which the the different channels represent the preambles used for Msg1 [30]. Accordingly, the application of the NOMA scheme proposed in [25] to the random access procedure could be seen, in a sense, as equivalent to increasing the number of preambles. This increase, however, is limited since the number of power levels must be kept to a low value due to practical reasons. Therefore, while this scheme will ease congestion episodes, they will not be completely eliminated and effective congestion control mechanisms will still be required. Furthermore, a known drawback of power-domain NOMA schemes is the increase of the transmission power. Consequently, such schemes may not be appropriate for those cases in which low energy consumption is a critical requirement.

III. EAB PERFORMANCE ANALYSIS

For the EAB mechanism evaluation, we measure three network KPIs, namely: the probability to successfully complete the random access procedure, P_s; the mean number of preamble transmissions needed by the UEs to successfully complete the random access procedure, K; and statistics of the access delay of the successful accesses, D. These KPIs are in conformance with the 3GPP directives to assess the efficiency of the random access procedure with MTC [14].

The access requests of MTC UEs follow a *Beta*(3, 4) distribution over a period of 10 s, according to traffic model 2 specified by the 3GPP in [14]. This traffic model can be seen as an extreme scenario in which a vast number of MTC UE arrivals (ranging from 5 000 to 30 000) occur in a highly synchronized manner (e.g., after an alarm that activates them).

We have assumed a single cell environment in which the model above captures the access requests in a single cell. However, we believe that the results obtained could be generalized to a multi-cell environment, because the effect of handoffs in the traffic model would not be relevant. Note that the only handoffs that would affect the access rate are those in which a UE leaves the cell during the access procedure, interrupting the procedure, and these would be very rare, considering that most M2M devices are static and the access procedure is of short duration.

TABLE 1. KPIs obtained for a heavy-loaded scenario ($N_M = 30\,000$), and $T_{\text{SIB14}} = 320$ ms.

		no control	EAB		
	$\frac{T_P}{T_{\text{SIB14}}}$	-	2	4	8
Success probability	P_s	0.31	0.38	0.81	0.99
Number of preamble transmissions, K	$\mathrm{E}[K]$	3.4	3.45	5.04	3.50
Access delay, D [ms]	$\mathrm{E}[D]$	70	1 315	5 972	9 647

To obtain the KPIs, we developed a discrete-event simulator that fully reproduces the behavior of UEs, BS, and RACH during the random access procedure. It consists in a set of Java classes implementing the monitor, events, traffic generators, measurement probes, and processes that reproduce the behavior of UEs and BS. The simulator has been validated through the results obtained in a reference configuration (Table 1). We assume a typical PRACH configuration, prach-ConfigIndex 6 [7], where the subframe length is 1 ms and the periodicity of RAOs is 5 ms. $R = 54$ out of the 64 available preambles are used for contention-based random access and the maximum number of preamble transmissions of each UE, preambleTransMax, is set to 10. Additional system configuration parameters can be found in [8, Table III]. These parameters are those specified by the 3GPP in [14] to conduct the study on the RACH capacity for MTC devices, and maintained in [16] for the evaluation of EAB.

The results presented in this section have been obtained using the congestion coefficients defined in [16] and

assuming that the BS has the necessary information for its exact calculation. In Section VI, we show that our proposed realistic calculation of these congestion coefficients has a very limited impact on the results.

In Table 1, we present the KPIs obtained in a heavy-loaded scenario of $N_M = 30\,000$ MTC UE arrivals. We consider two scenarios, without congestion control (i.e., with no EAB) and with EAB, for different configuration values of T_P. The results displayed in Table 1 have been obtained by averaging the results of 100 independent simulation runs. The results for EAB were obtained for $T_{\text{SIB14}} = 320$ ms and $T_P \in \{640, 1280, 2560\}$ ms. Previously, we verified that, as in [16], [17], the combinations of parameters in which $T_P < T_{\text{SIB14}}$ result in very poor performance in terms of P_s. The results shown in Table 1 are in line with those presented in [13], [17], [19], [20]. We can see that, with no congestion control, performance in terms of P_s and K is very poor. With EAB, as the T_P duration increases P_s increases. This is because the greater the T_P, the greater the number of updates per T_P, which results in a lower intensity of the traffic burst after each SIB14 update. However, the cost of increasing T_P is that ACs are unbarred at a slower rate (one AC per T_P), thus increasing the access delay.

Fig. 1 shows in detail how the EAB mechanism handles congestion episodes. This figure plots the temporal distribution of preamble transmissions during a congestion episode of $N_M = 30\,000$ MTC UEs, for $T_{\text{SIB14}} = 320$ ms and $T_P \in \{640, 1280, 2560\}$ ms. It can be seen that, when congestion builds up, the CC_{1000} calculated at the BS grows, and when it exceeds the barring threshold (0.4), EAB is enabled and all ACs except AC 0 are barred, which occurs at $t \approx 2$ s. From this moment, the traffic declines as the UEs read the SIB14 message following their respective *PO*s. When the CC_{500} is under the unbarring threshold (0.4), the BS starts to release the barred ACs one at a time, at intervals not shorter than $\max\{T_P, T_{\text{SIB14}}\} = T_P$. Every time that a barred AC is released, a number of UEs access the RACH in bursts of periodicity $T_{\text{SIB14}} = 320$ ms. Each burst corresponds to those UEs whose *PO*s are between two consecutive broadcastings of the SIB14 message; these UEs access the RACH concurrently because they update their SIB14 simultaneously. For each AC released, a total of T_P/T_{SIB14} traffic bursts are produced. When all the traffic from a released AC ends, the congestion vanishes and, when the CC_{500} is under the unbarring threshold (0.4), the BS releases a new barred AC. The full episode ends when all the barred ACs (from 1 to 9) are released.

The dynamics of EAB are most clearly observed in the case of a large paging cycle ($T_P = 2560$ ms). In Fig. 1c, we can see that the barred ACs begin to be released starting on $t \approx 5$ s. In this case, barred ACs are released at intervals slightly greater than 2560 ms (T_P). For each AC released, a group of eight (T_P/T_{SIB14}) traffic bursts at intervals of 320 ms (T_{SIB14}) is generated. The time gap between two consecutive groups of bursts is about 1 s, which is the time it takes for CC_{500} to go below the unbarring threshold. In Fig. 1c, nine groups

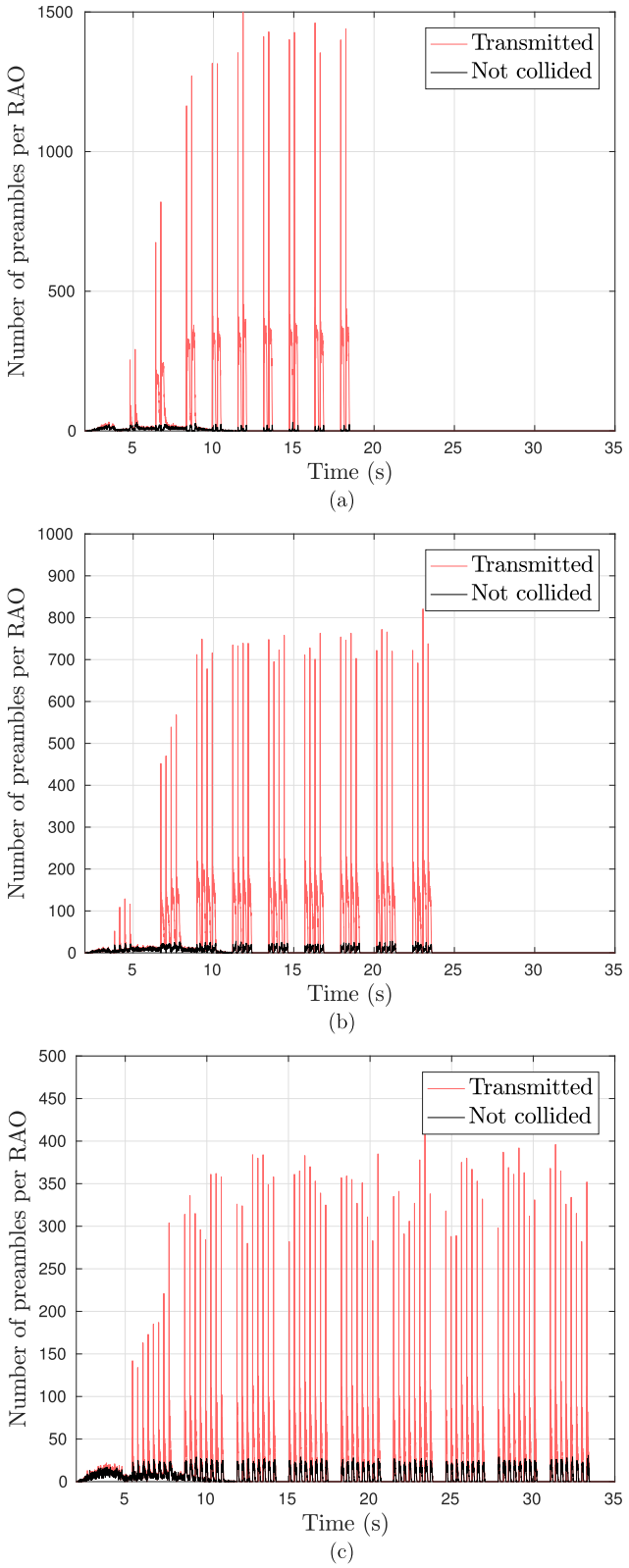

FIGURE 1. Average number of preamble transmissions for $N_M = 30\,000$, and $T_{\text{SIB14}} = 320$ ms. (a) $T_P = 640$ ms. (b) $T_P = 1280$ ms. (c) $T_P = 2560$ ms.

of bursts, corresponding to the nine barred ACs, are clearly observed. The first barred ACs are released when new UEs are still arriving, which gives rise to a more complex dynamic

at the beginning of the episode (recall that the benchmark traffic model generates a congestion episode of 10 s, and this group lasts approximately between 5 and 8 s). Therefore, when the first AC was released a significant proportion of its UEs had not arrived yet. The consequence of this is that, as seen in the figure, the bursts of the first group are of lower intensity than those of the following groups. This also explains why the intensity of the bursts in the first group exhibits an increasing pattern.

For shorter paging cycles (Figs. 1a and 1b), the specifics of the first group of bursts described above are more pronounced and extend to subsequent groups (up to the second group in Fig. 1b and up to the fourth in Fig. 1a). This is mainly because a shorter duration of the paging cycle causes that the UEs of barred ACs notice their status before. Therefore, once EAB is activated, the traffic is reduced more quickly, the congestion relieves faster, and the barred ACs start to be released before.

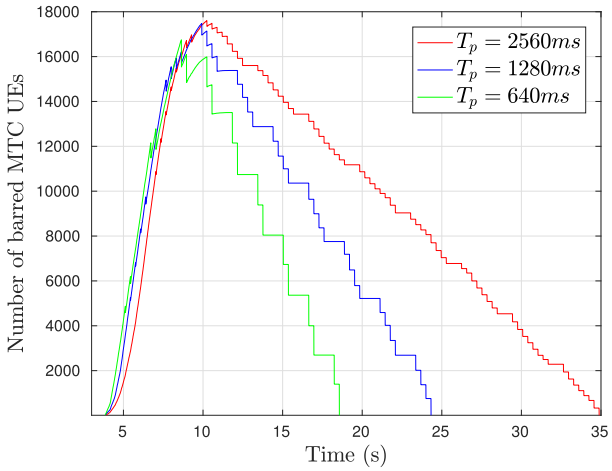

FIGURE 2. Number of barred UEs during a congestion episode with EAB, $N_M = 30\,000$, and $(T_P, T_{\text{SIB14}}) = (\{640, 1280, 2560\}, 320)$ ms.

Fig. 2 shows the evolution in time of the number of barred UEs in the same scenario, and provides an additional insight into what has been explained above. When congestion starts to build up, the number of barred UEs increases, until it reaches a maximum when the congestion episode ends (at $t \approx 10$s). It can be seen that the curve in its growing part is not completely smooth. The peaks in this part of the curve are due to those barred ACs that are released before the congestion ends, that is, while new UEs are still arriving. This happens especially for short paging cycles, as explained before. When the congestion ends, the number of barred UEs decreases stepwise every T_{SIB14} in groups of T_P/T_{SIB14} steps, which correspond to the groups of traffic bursts described above.

Note that a when the value of T_P is increased, the downward steps in Fig. 2 become smaller, which means less intense traffic bursts. However, an increase of T_P also provokes a significant increase in the time required to release all the UEs, which means greater access delay. The intensity of the traffic bursts is inversely proportional to the number of steps in a group, which is T_P/T_{SIB14}. This because the total number of UEs belonging to a barred AC are divided into T_P/T_{SIB14} traffic burst when its AC is released. This can be confirmed in Fig. 1: for $T_P/T_{\text{SIB14}} = 2$ (Fig. 1a), the bursts reach approximately 1500 preambles per RAO; for $T_P/T_{\text{SIB14}} = 4$ (Fig. 1b), they reach approximately 750 preambles per RAO; and for $T_P/T_{\text{SIB14}} = 8$ (Fig. 1c), they reach approximately 375 preambles per RAO. These values correspond to the number of UEs per AC ($N_M/10 = 3000$ UEs) divided by the number of bursts per group (T_P/T_{SIB14}).

From this analysis, we conclude that the main limitation in EAB performance is due to the fact that, when a barred AC is released, groups of UEs initiate their access procedure in a synchronized mode. The UEs that are synchronized are those that access the paging information in the same *PO*. This synchronization causes traffic bursts that degrade the network performance. For small values of T_P (640 or 1280 ms), these traffic bursts are far above the RACH capacity, which results in a low successful access probability and a high number of preamble transmissions. By increasing T_P, the traffic from each released AC is distributed between a greater number of bursts, thus reducing their intensity. As can be seen in Table 1, this improves the successful access probability, but it hardly changes the number of required transmissions to successfully complete the access procedure and increases the access delay.

IV. EAB PERFORMANCE IMPROVEMENT

In this section, we evaluate two alternatives to improve the EAB scheme which overcomes the limitations mentioned above, thus improving the network performance. Firstly, we study the combination of EAB and access class barring (ACB), and show that ACB can reduce the intensity of the traffic bursts that occur every time a barred AC is released. Secondly, we propose and evaluate a more simple and effective way to obtain the same result, by means of including a congestion avoidance (CA) backoff in the UE access procedure.

A. EAB WITH ACB

ACB [9], [11] is a congestion control scheme designed to redistribute the access requests of UEs through time to alleviate congestion episodes. When ACB is implemented, the BS broadcasts periodically a message containing a barring rate $P_{\text{ACB}} \in \{0.05, 0.1, \ldots, 0.3, 0.4, \ldots, 0.7, 0.75, 0.8, \ldots, 0.95\}$, and a mean barring time $T_{\text{ACB}} \in \{4, 8, 16, \ldots, 512\,\text{s}\}$. Before starting the random access procedure, the UE performs a test to find out if it can initiate the access procedure immediately or it has to wait. For this, it generates a random number uniformly distributed between 0 and 1, $g = \mathcal{U}[0, 1)$: if $g \leq P_{\text{ACB}}$, the UE transmits its preamble; otherwise, the UE waits for a random time calculated as

$$T_{\text{barring}} = [0.7 + 0.6\,\mathcal{U}[0, 1)]\; T_{\text{ACB}}, \tag{1}$$

and then it repeats the test.

In the normal operation of ACB, all the new arriving UEs must perform the ACB test with the last values of $\{P_{\text{ACB}}, T_{\text{ACB}}\}$ broadcast by the BS. However, here we want to test the simultaneous operation of ACB and EAB to determine the effect of ACB on the performance of EAB. For this reason, we have tested a scheme in which ACB operates only when EAB is enabled. This means that UEs will pass the ACB test only if they have been previously barred; if not, they initiate the access procedure immediately.

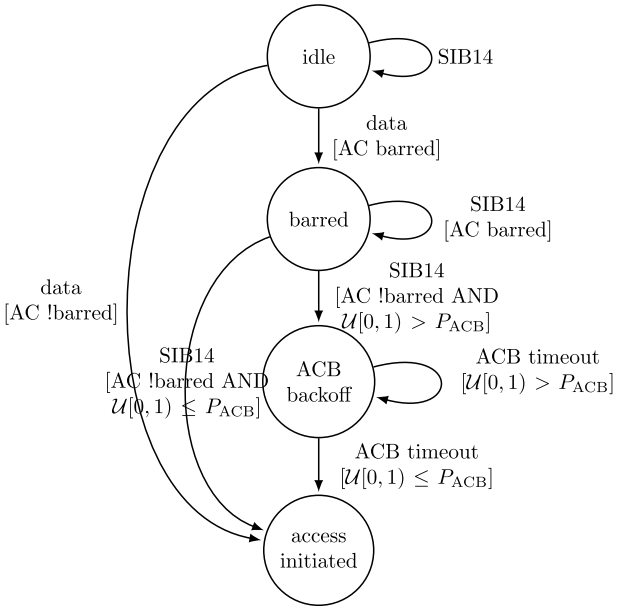

FIGURE 3. State diagram for a UE with EAB + ACB congestion control.

The behavior of the UE with EAB+ACB congestion control is illustrated in the state diagram in Fig. 3. While the UE is in idle state, it periodically wakes up at its *PO* and, if the system information change in the paging message is set, it updates its AC barring status by reading the next message containing the SIB14. When, in idle state, a request of sending data is received, the UE initiates the access or goes to barred state, depending on its AC barring status. While in barred state, it keeps updating its AC barring status, and when it is not barred, it initiates the access or goes to ACB backoff state, depending on the result of the ACB test. In ACB backoff state, every time that the ACB backoff ends, the UE repeats the ACB test and, depending on the result, it initiates the access or remains in ACB backoff state.

To determine the values of $\{P_{\text{ACB}}, T_{\text{ACB}}\}$ to be tested, we have considered that the objective is to distribute the transmissions of a traffic burst over the interval between the start of the current burst and the start of the next one. Ideally, this would be achieved if all the UEs waited a time uniformly distributed between 0 and T_{SIB14} ($\mathcal{U}[0, T_{\text{SIB14}})$). This is not feasible with ACB because: firstly, P_{ACB} cannot be 0, and therefore a proportion of the incoming UEs will not wait any backoff; and secondly, the ACB backoff time given by (1) is not uniformly distributed. To approximate the ideal behavior, we have chosen combinations of $\{P_{\text{ACB}}, T_{\text{ACB}}\}$ that lead to a mean total ACB waiting time of $T_{\text{SIB14}}/2$ (the mean of $\mathcal{U}[0, T_{\text{SIB14}})$). The total ACB waiting time is the sum of N ACB backoff times, where N is the number of times a UE will be barred until its access is finally allowed. These backoff times are independent and follow the same distribution, which is given by (1). The distribution of the number of ACB backoffs prior to the access attempt is $\text{P}(N = n) = (1-P_{\text{ACB}})^n P_{\text{ACB}}$, for $n \geq 0$, and consequently $\text{E}[N] = (1-P_{\text{ACB}})/P_{\text{ACB}}$. Now, the expected total ACB waiting time can be obtained as $\text{E}[N]T_{\text{ACB}}$. Therefore, for a given P_{ACB}, in order to obtain a mean total ACB waiting time equal to $T_{\text{SIB14}}/2$, the required value of T_{ACB} is given by

$$T_{\text{ACB}} = \frac{P_{\text{ACB}}}{1-P_{\text{ACB}}} \frac{T_{\text{SIB14}}}{2}. \tag{2}$$

For most of the values of P_{ACB}, the value obtained from (2) is much smaller than the minimum value of T_{ACB} specified in [9], which is 4 s. For $P_{\text{ACB}} = 0.95$, which is the maximum value of P_{ACB} below 1 (note that $P_{\text{ACB}} = 1$ means that ACB is disabled), and $T_{\text{SIB14}} = 320$ ms, the required mean barring time given by (2) is $T_{\text{ACB}} = 3\,040$ ms. For $P_{\text{ACB}} = 0.5$, $T_{\text{ACB}} = 160$ ms, and for $P_{\text{ACB}} = 0.1$, $T_{\text{ACB}} = 17.78$ ms.

TABLE 2. KPIs obtained for EAB+ACB for a heavy-loaded scenario ($N_M = 30\,000$), $T_{\text{ACB}} = 4$ s, and $T_{\text{SIB14}} = 320$ ms.

	$\frac{T_P}{T_{\text{SIB14}}}$	2	4	8
$P_{\text{ACB}} = 0.1$	P_s	0.99	0.99	0.99
	E[K]	1.68	1.63	1.80
	E[D] (ms)	27 631	29 900	32 549
$P_{\text{ACB}} = 0.5$	P_s	0.84	0.99	0.99
	E[K]	3.03	2.78	2.28
	E[D] (ms)	7 386	8 544	12 410
$P_{\text{ACB}} = 0.95$	P_s	0.41	0.84	0.99
	E[K]	3.34	4.85	3.36
	E[D] (ms)	2 095	6 308	9 844

Table 2 shows the KPIs obtained with $T_{\text{ACB}} = 4$ s (the minimum value in the specifications) and $P_{\text{ACB}} \in \{0.1, 0.5, 0.95\}$, in a heavy-loaded scenario of $N_M = 30\,000$ MTC UE arrivals with EAB+ACB, $T_{\text{SIB14}} = 320$ ms, and $T_P \in \{640, 1280, 2560\}$ ms. These values are the average of 100 independent simulation runs. As can be seen, for medium or low values of P_{ACB}, acceptable values of P_s and K are obtained but at the cost of excessively increasing the delay (note that $T_{\text{ACB}} = 4$ s is much higher than the values given by (2)). On the other hand, for high values of P_{ACB}, the results obtained do not differ significantly from those obtained for EAB only (Table 1), as expected, because with values of P_{ACB} close to 1, the ACB effect is negligible. From these results, we conclude that the desired effect of EAB+ACB cannot be obtained with standard values of T_{ACB}.

In order to evaluate the potential capacity of EAB+ACB of reducing traffic bursts, and its effects in the KPIs, hereafter we test EAB+ACB with $P_{\text{ACB}} \in \{0.5, 0.1\}$, and the corresponding values of T_{ACB} given by (2), that is, $T_{\text{ACB}} = 160$ ms and $T_{\text{ACB}} = 17.78$ ms, respectively)

Fig. 4 shows the temporal distribution of preamble transmissions during a congestion episode of $N_M = 30\,000$ MTC UEs, for $T_{\text{SIB14}} = 320$ ms and $T_P \in \{640, 1280, 2560\}$ ms, with EAB+ACB using the parameters $P_{\text{ACB}} = 0.5$ and $T_{\text{ACB}} = 160$ ms. As expected, the intensity of the traffic bursts is reduced by half, because only half of the UEs pass the ACB test at the first attempt. The other half of UEs are delayed by the ACB backoff. Depending on the duration of the paging cycle, this reduction may be not enough to achieve a high successful access probability. In fact, for the values of T_P tested, the intensity of the traffic bursts is still above the RACH capacity.

In Fig. 5, the results for $P_{\text{ACB}} = 0.1$ and $T_{\text{ACB}} = 17.78$ ms in the same scenario are shown. Now, the intensity of the traffic bursts is further reduced. As the numerical results shown in Section V confirm, this behavior results in a significant performance improvement, in terms of both successful access probability and number of preamble transmissions.

B. EAB WITH CA BACKOFF

As seen in the previous section, with the EAB+ACB scheme, the objective of reducing the intensity of traffic bursts that deteriorate network performance can be achieved. However, this requires values in the parameters of ACB not contemplated in the specification. In this section, we describe a simpler and more efficient way to eliminate the traffic bursts without the intervention of ACB. Our solution is implemented in the UE-side and consists in performing a collision-avoidance (CA) backoff before the access attempt. The behavior of the UE with EAB and CA backoff congestion control is illustrated in the state diagram in Fig. 6. The CA backoff is performed only by those UEs leaving the barred state. When a UE in barred state updates its SIB14 and finds out that its ACs has been unbarred, it waits for a random time uniformly distributed between 0 and T_{SIB14} before starting the access procedure.

Thanks to the CA backoff, UEs that update their SIB14 simultaneously do not access the RACH concurrently, but they distribute their access attempts throughout a T_{SIB14} period. These UEs are those which had their POs in the previous T_{SIB14} period. In the same way, UEs with their POs in the current T_{SIB14} period will distribute their access attempts along the next T_{SIB14} period. By doing so, the bursts of arrivals described above are eliminated, which allows resolving the congestion in a shorter time.

Fig. 7 illustrates the temporal distribution of the preamble transmissions during a congestion episode of $N_M = 30\,000$ MTC UEs, for $T_{\text{SIB14}} = 320$ ms and $T_P \in \{640, 1280, 2560\}$ ms, with EAB+CA. As seen, now all the traffic coming from the release of a barred AC is evenly distributed throughout a paging cycle. For short paging cycles (as in the case of $T_P = 640$ ms), this may be not enough to reach a high probability of successful access, but for larger paging cycles ($T_P \in \{1280, 2560\}$ ms), it is observed that the

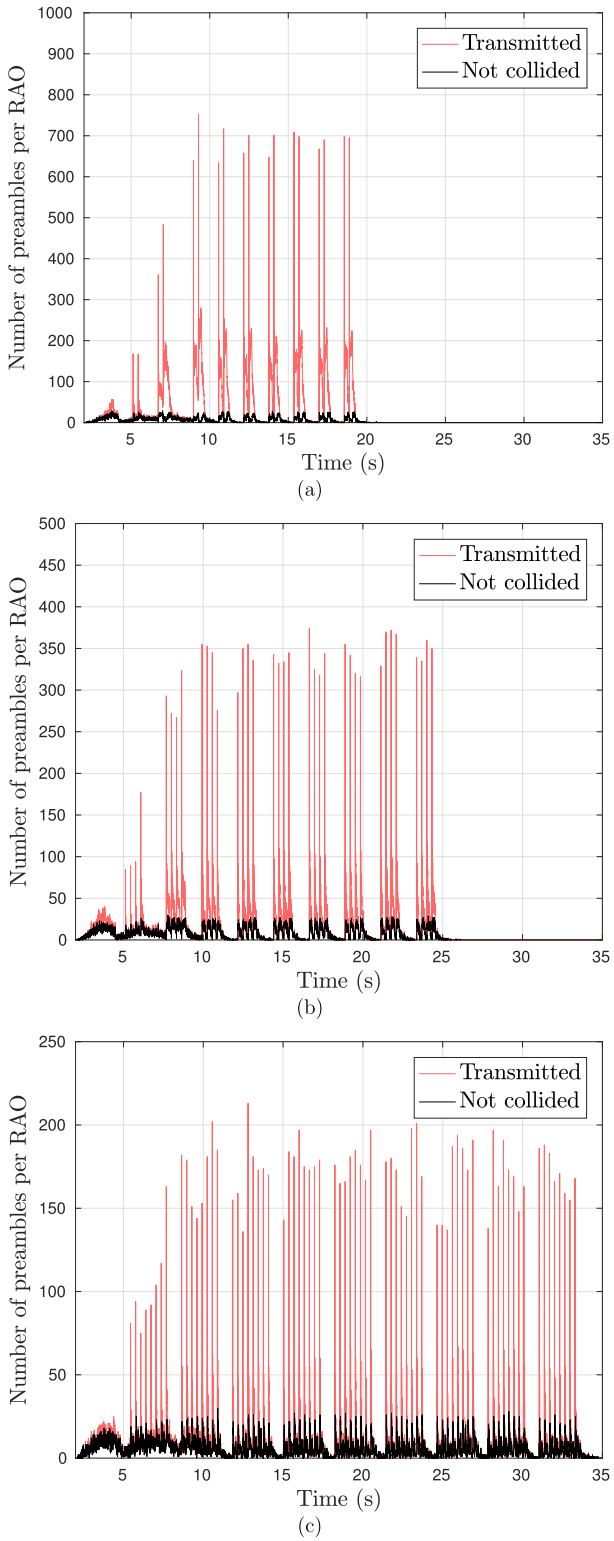

FIGURE 4. Average number of preamble transmissions with EAB+ACB, $N_M = 30\,000$, $T_{\text{SIB14}} = 320$ ms, $P_{ACB} = 0.5$, and $T_{ACB} = 160$ ms. (a) $T_P = 640$ ms. (b) $T_P = 1280$ ms. (c) $T_P = 2560$ ms.

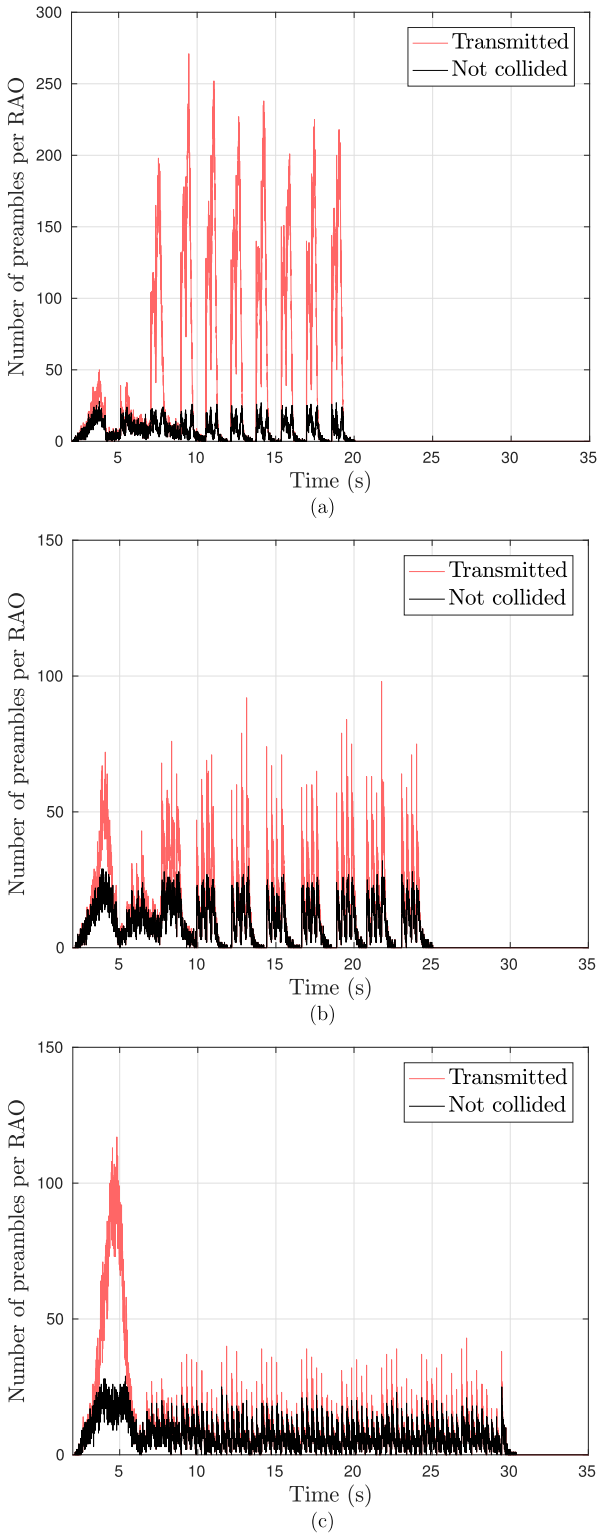

FIGURE 5. Average number of preamble transmissions with EAB+ACB, $N_M = 30\,000$, $T_{SIB14} = 320$ ms, $P_{ACB} = 0.1$, and $T_{ACB} = 160$ ms. (a) $T_P = 640$ ms. (b) $T_P = 1280$ ms. (c) $T_P = 2560$ ms.

resulting traffic does not substantially exceed the capacity of the channel, which results in an improvement of both access probability and number of preamble transmissions, as we show in the following section.

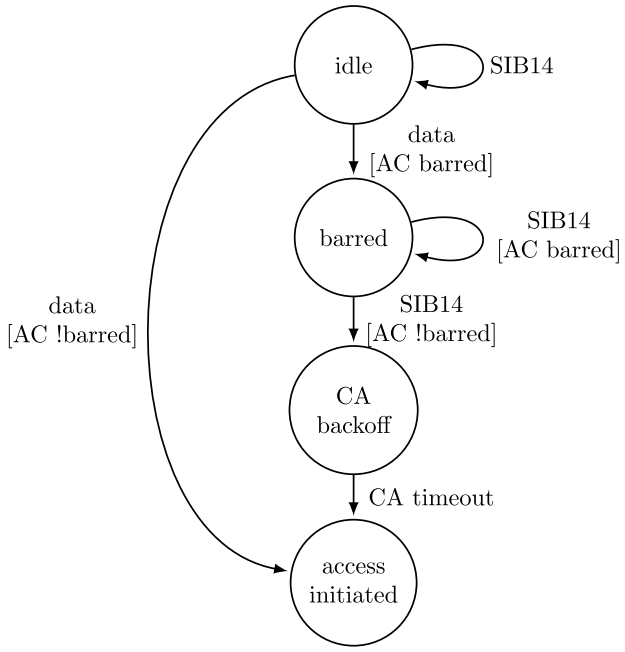

FIGURE 6. State diagram for a UE with EAB+CA congestion control.

V. NUMERICAL RESULTS

In this section, we present the quantitative results of the performance evaluation of the two improvement schemes described above, and compare them with each other and with the performance of the baseline EAB.

Fig. 8 depicts the successful access probability, P_s, for the baseline EAB, EAB+ACB with $P_{\text{ACB}} = 0.5$, EAB+ACB with $P_{\text{ACB}} = 0.1$, and EAB+CA. The value of P_s has been obtained for several traffic intensities, ranging from $N_M = 6\,000$ UEs (light congestion) to $N_M = 30\,000$ (heavy congestion). The shaded areas around the curves represent the confidence intervals for a confidence level of 99 %. All the results have been obtained by averaging the results of 100 independent simulation runs.

As can be seen, for a short paging cycle ($T_P = 640$ ms, Fig. 8a) and moderate or heavy congestion ($N_M > 20\,000$ UEs), none of the tested schemes obtains a sufficiently high successful access probability (above 0.9). Nevertheless, EAB+ACB and EAB+CA represent an important improvement with respect to EAB. They obtain a very high P_s for moderate congestion up to $N_M = 20\,000$ UEs, and, for a heavy congestion of $N_M = 30\,000$ UEs, P_s goes from ≈ 0.4 in plain EAB to ≈ 0.7 in the enhanced schemes. However, this probability is still low because, for a severe congestion, the paging cycle duration is not enough to accommodate all the traffic generated when an AC is released, even if it has been distributed evenly throughout the paging cycle (see Fig. 7a).

For a medium paging cycle ($T_P = 1280$ ms, Fig. 8b), plain EAB can successfully cope with a moderate congestion up to $N_M \approx 25\,000$ UEs, while EAB+ACB and EAB+CA obtain an excellent P_s even with heavy congestion. With respect to the effect of the value of P_{ACB} used in EAB+ACB, with

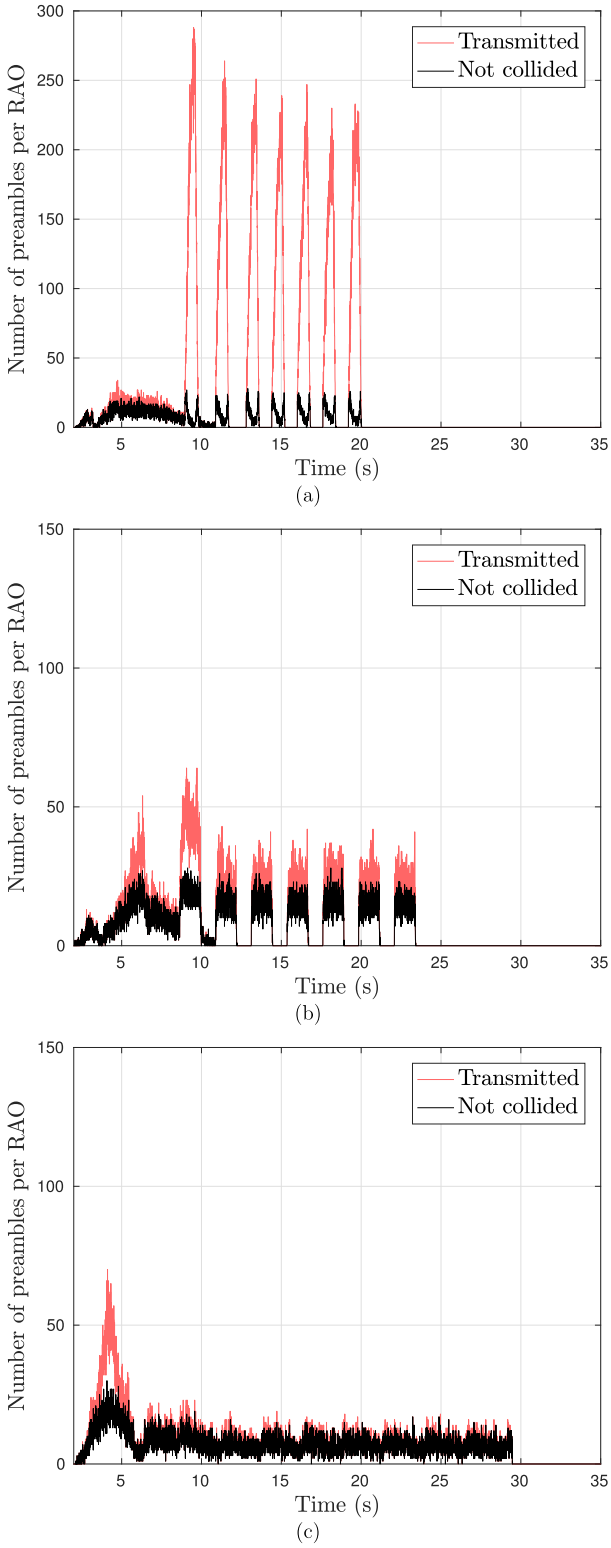

FIGURE 7. Average number of preamble transmissions with EAB+CA, $N_M = 30\,000$, and $T_{SIB14} = 320$ ms. (a) $T_P = 640$ ms. (b) $T_P = 1280$ ms. (c) $T_P = 2560$ ms.

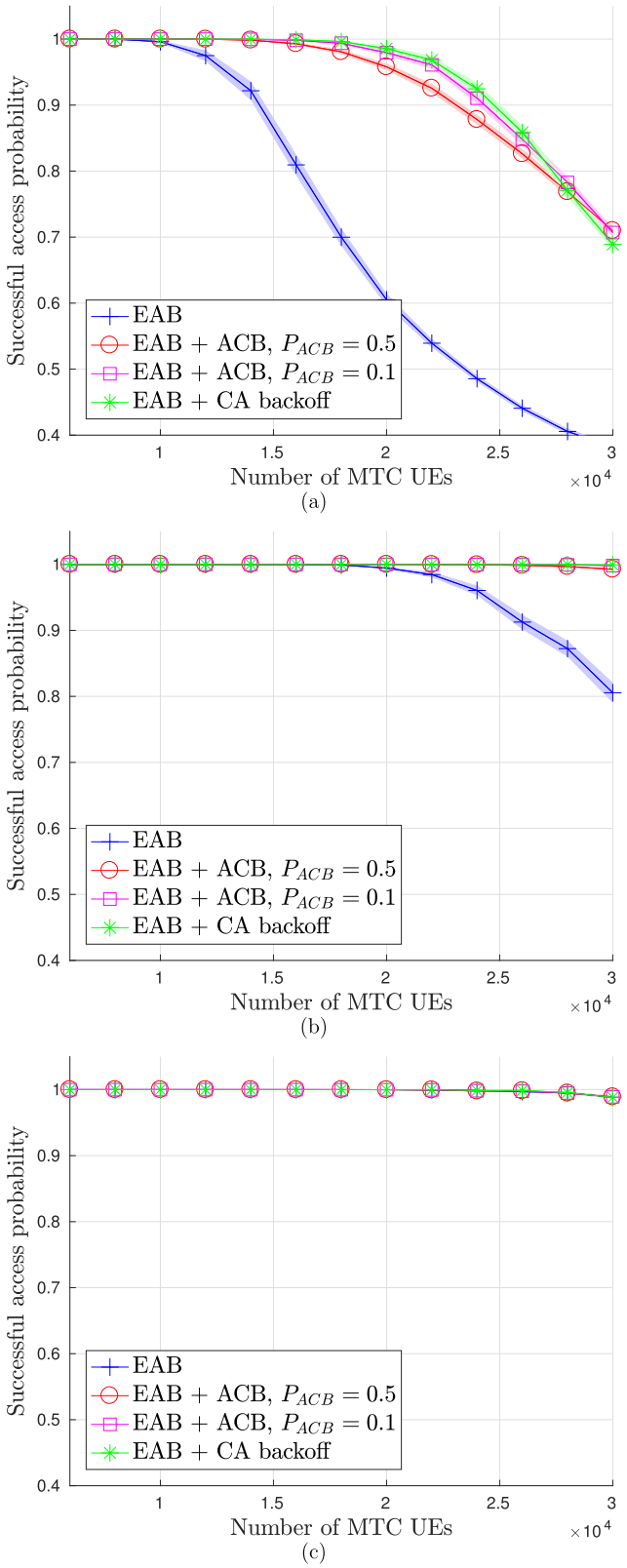

FIGURE 8. Successful access probability. (a) $T_P = 640$ ms. (b) $T_P = 1280$ ms. (c) $T_P = 2560$ ms.

$P_{\text{ACB}} = 0.5$ a slight performance deterioration is appreciated as the load increases, whereas with $P_{\text{ACB}} = 0.1$ the behavior is indistinguishable from that of EAB+CA.

For a long paging cycle ($T_P = 2560$ ms, Fig. 8c), all the tested schemes obtain a high successful access probability even for heavily congested scenarios. In this case, the

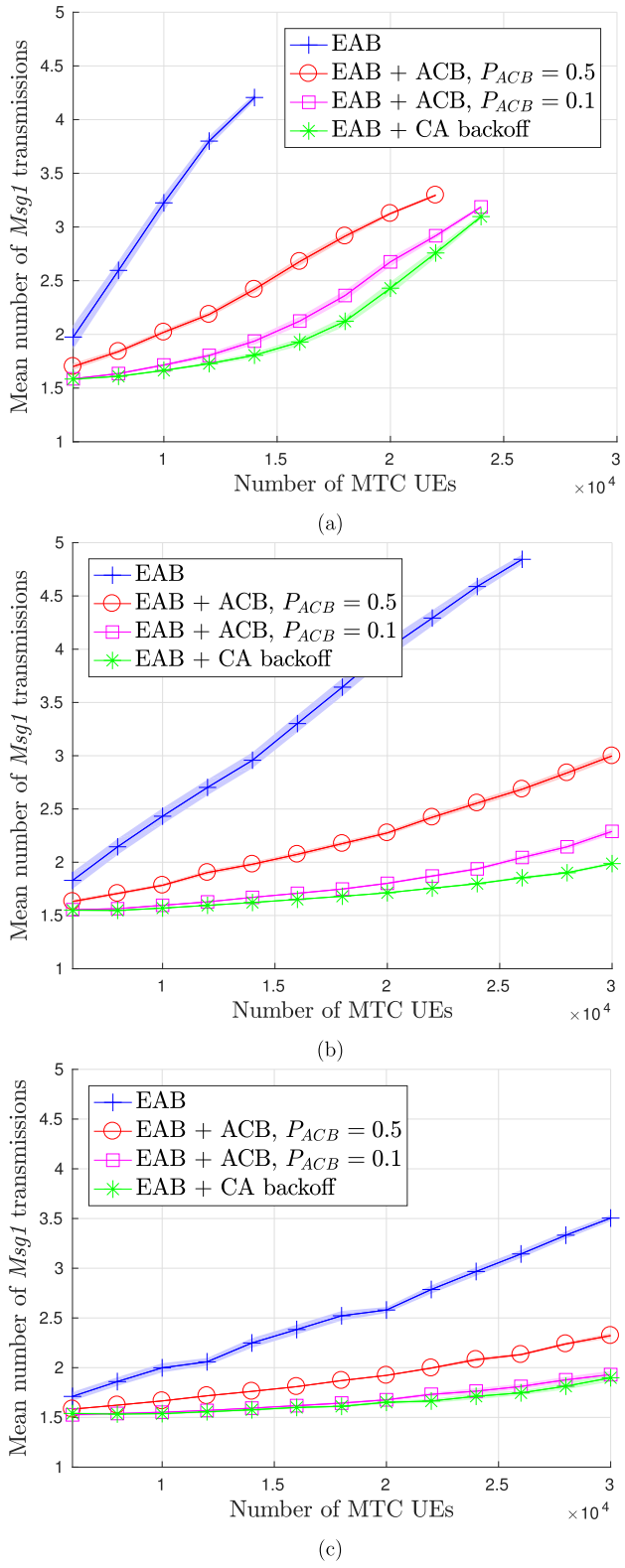

FIGURE 9. Average number of preamble transmissions. (a) $T_P = 640$ ms. (b) $T_P = 1280$ ms. (c) $T_P = 2560$ ms.

performance comparison will be made in terms of other KPIs (number of preamble transmissions and delay).

Fig. 9 shows the mean number of preamble transmissions (E[K]) needed to successfully complete the random access procedure, corresponding to the previous experiment. Note that, for the calculation of K (and also for the delay), only the successful accesses are counted. For this reason, the comparison of two cases in terms of E[K] only makes sense if in both cases the number of successful accesses is similar. As seen in Fig. 8, the values of P_S differ greatly in some cases. Thus, in Fig. 9 we have shown only those results in which the successful access probability is sufficiently high, that is, those cases in which the majority of accesses have been accounted for the calculation of E[K]. For this, we have chosen a threshold of $P_s > 0.9$. For those cases in which $P_s < 0.9$, the comparison in terms of K would not be fair and moreover the performance would be too low for being of practical interest. In all cases, EAB+ACB and EAB+CA reduce to approximately half the mean number of preamble transmissions, which results in power savings. Moreover, here it is confirmed that EAB+ACB performs better with $P_{\text{ACB}} = 0.1$ than with $P_{\text{ACB}} = 0.5$, and in the first case the result is very close (although slightly worse) to that of EAB+CA. The reduction in the number of preamble transmissions is especially relevant in delay-tolerant scenarios where UEs are mainly power-constrained [31].

Fig. 10 shows the mean access delay corresponding to the same experiment. As before, and for the same reasons, only those results corresponding to a $P_s > 0.9$ have been included. As can be seen in this figure, the confidence intervals obtained for the delay is wider than those obtained for the other KPIs, with the same number of simulation runs and the same confidence level, which means that the delay caused by EAB has a high variability, especially for a long paging cycle. These results show that EAB+ACB with $P_{\text{ACB}} = 0.1$ and EAB+CA reduce the mean access delay by approximately one second in most cases. However, this reduction is not particularly significant, given the magnitude of the total delay. What is important here is that, with the improved schemes it is possible to obtain better KPIs than with plain EAB, but using a shorter paging cycle, which does result in a significant reduction in the delay. Thus, for example, with CA backoff and $T_P = 1280$ ms we obtain the same successful access probability ($P_s > 0.99$) as with EAB and $T_P = 2560$ ms, but in the first case we obtain E[K] ≈ 2 transmissions and E[D] ≈ 6.5 s, while with plain EAB we obtain E[K] ≈ 3.5 transmissions and E[D] ≈ 10 s.

In Table 3, we present the numerical values of KPI statistics for the four tested schemes in a heavy-loaded traffic scenario ($N_M = 30\,000$), for $T_{\text{SIB14}} = 320$ ms, and $T_P \in \{640, 1280, 2560\}$ ms. These values are the average of 100 independent simulation runs. The delay statistics, (mean value and 50th, 75th, 90th, 95th, and 99th percentiles) are included only in those cases in which $P_s > 0.9$. The values of the percentiles of the delay confirm what the mean values in Fig. 10 suggested.

VI. CALCULATION OF THE CONGESTION COEFFICIENTS

The results presented thus far have been obtained using the congestion coefficients defined in [16]. According to this

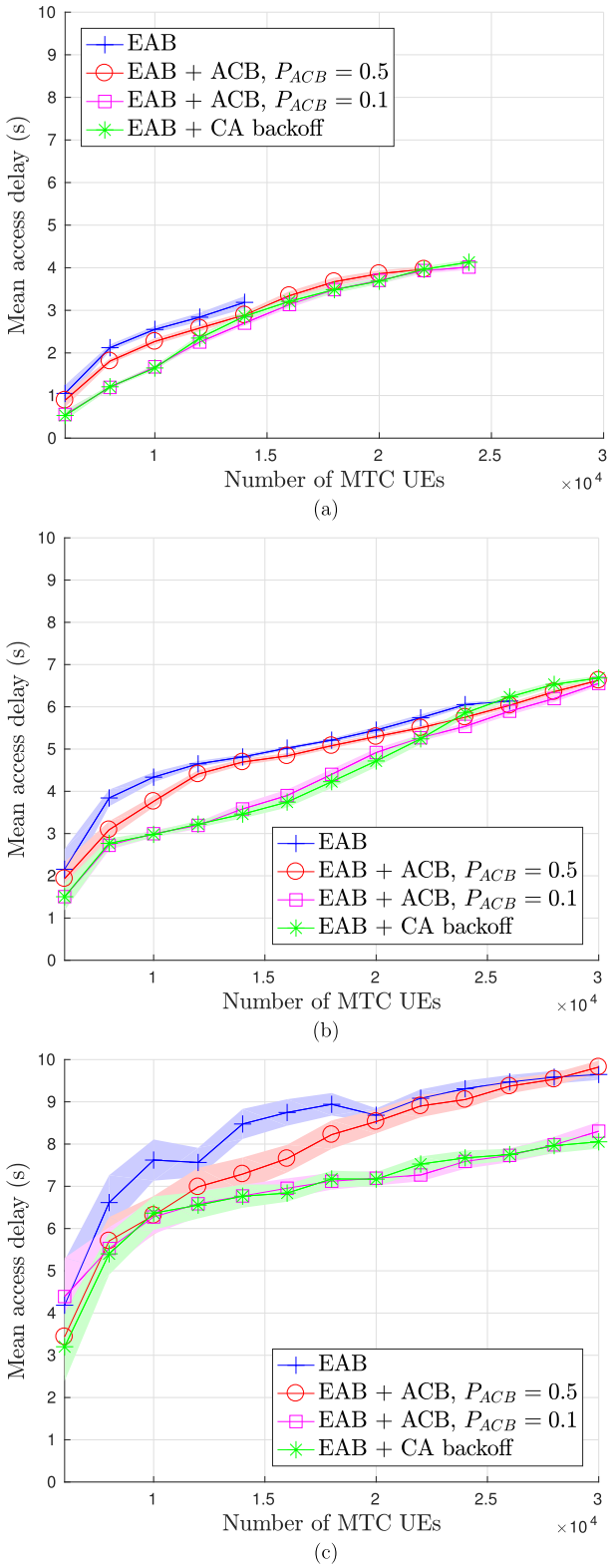

FIGURE 10. Average access delay. (a) $T_P = 640$ ms. (b) $T_P = 1280$ ms. (c) $T_P = 2560$ ms.

definition, the congestion coefficient for a moving time-window of W ms (CC_W) is

$$\mathrm{CC}_W = 1 - \frac{\mathrm{nRAR}_W}{\mathrm{nPT}_W}, \tag{3}$$

TABLE 3. KPIs obtained for a heavy-loaded scenario ($N_M = 30\,000$), and $T_{\mathrm{SIB14}} = 320$ ms.

T_P			EAB	EAB+ACB ($P_{\mathrm{ACB}} = 0.5$)	EAB+ACB ($P_{\mathrm{ACB}} = 0.1$)	EAB+CA
640 ms		P_s	0.38	0.71	0.71	0.69
1280 ms		P_s	0.81	0.99	0.99	0.99
		E[K]	-	3.00	2.29	1.99
	delay stats. [ms]	E[D]	-	6 627	6 554	6 682
		D_{99}	-	18 918	18 778	19 169
		D_{95}	-	17 072	16 892	17 354
		D_{90}	-	15 591	15 447	15 816
		D_{75}	-	11 873	11 773	12 013
		D_{50}	-	5 726	5 635	5 691
2560 ms		P_s	0.99	0.99	0.99	0.99
		E[K]	3.50	2.32	1.93	1.90
	delay stats. [ms]	E[D]	9 647	9 822	8 305	8 053
		D_{99}	27 956	28 027	24 267	23 459
		D_{95}	25 781	25 854	21 948	21 530
		D_{90}	23 895	23 958	20 084	19 731
		D_{75}	17 732	17 897	15 123	14 785
		D_{50}	7 663	8 047	6 898	6 590

where nRAR_W is the number of RARs sent during W ms and nPT_W is the number of preamble transmissions during W ms. To calculate CC_Ws defined in this way, the BS would need to know the number of preamble transmissions during the time window. However, this number is unknown at the BS because those preambles transmitted by more than one UE are not decoded. Therefore, to evaluate the performance of EAB in a realistic environment, the value of nPT_W in (3) should be estimated from the number of preambles used (by at least one UEs) at each RAO. In order to demonstrate that the results shown in the previous sections are valid also in a realistic implementation of EAB, in this section, we propose and evaluate a way to calculate the CC_Ws from the information available in the BS, namely, from the number of preambles used and the number of RARs sent at every RAO.

Let $Y_j(i) \in \{0, 1\}$ be the random variable that denotes the transmission of preamble j at RAO(i) given that the total number of transmissions at RAO(i) is $n_t(i)$. Then, $Y_j(i) = 0$ when preamble j has not been transmitted by any UE at RAO(i), and $Y_j(i) = 1$ otherwise. Therefore, the probability distribution of $Y_j(i)$ is

$$\mathrm{P}(Y_j(i) = 0) = \left(1 - \frac{1}{R}\right)^{n_t(i)}, \tag{4}$$

$$\mathrm{P}(Y_j(j) = 1) = 1 - \left(1 - \frac{1}{R}\right)^{n_t(i)}, \tag{5}$$

where R is the number of available preambles. Thus,

$$\mathrm{E}[Y_j(i)] = 1 - \left(1 - \frac{1}{R}\right)^{n_t(i)}. \tag{6}$$

The number of used preambles at RAO(i), $n_u(i)$, is

$$n_u(i) = \sum_{j=0}^{R} Y_j(i), \tag{7}$$

and its expected value is

$$\mathrm{E}[n_u(i)] = R\left[1 - \left(1 - \frac{1}{R}\right)^{n_t(i)}\right]. \tag{8}$$

Since $n_u(i)$ is known at the BS, and assuming that $\mathrm{E}[n_u(i)]$ changes slowly, it can be estimated from a short-term time average of $n_u(i)$. Let $\hat{n}_u(i)$ be an estimate of $\mathrm{E}[n_u(i)]$ at RAO(i) obtained by an exponential smoothing of $n_u(i)$,

$$\hat{n}_u(i) = \alpha\, \hat{n}_u(i-1) + (1-\alpha)\, n_u(i), \tag{9}$$

with $\alpha < 1$. Then, from (8), the estimated value of the number of transmitted preambles used to calculate the congestion coefficients can be obtained as

$$\hat{n}_t(i) = \frac{\log\left(1 - \frac{\hat{n}_u(i)}{R}\right)}{\log\left(1 - \frac{1}{R}\right)}. \tag{10}$$

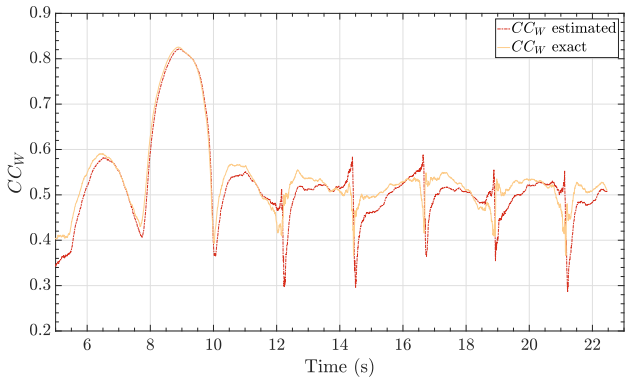

FIGURE 11. $\mathbf{CC_W}$ **for** $\boldsymbol{W = 1000}$ **ms exact and estimated during a congestion episode with** $\boldsymbol{N_M = 30\,000}$ **and** $\boldsymbol{(T_P, T_{SIB14}) = (2560, 320)}$ **ms.**

We have repeated the previous simulations but this time estimating the CC_W as described above. Fig. 11 shows an example of the evolution of CC_{1000} during a congestion episode with $N_M = 30\,000$ MTC UEs arrivals and $(T_P, T_{\mathrm{SIB14}}) = (2560, 320)$ ms. In the figure, the CC_{1000} obtained from the estimated number of transmissions is compared with the obtained from the exact value of transmissions. As can be seen, the values of CC_{1000} obtained by estimation are very close to those obtained using the real number of preambles transmitted. The error is only significant for brief intervals that occur with periodicity T_{SIB14}, each time an AC is unbarred.

Fig. 12 shows a comparison of the KPIs obtained with exact congestion coefficients with those obtained with estimated ones in a heavily congested scenario ($N_M = 30\,000$), for the different EAB schemes and paging cycle durations tested in the previous sections. As seen, using the estimated CC_Ws, the effect of the proposed EAB improvements on performance is approximately the same as the one we obtained using the exact congestion coefficients. In fact, all the conclusions stated in the previous section on the performance of the different EAB schemes tested remain valid in a scenario of realistic implementation in which the congestion coefficients are calculated by means of an estimator.

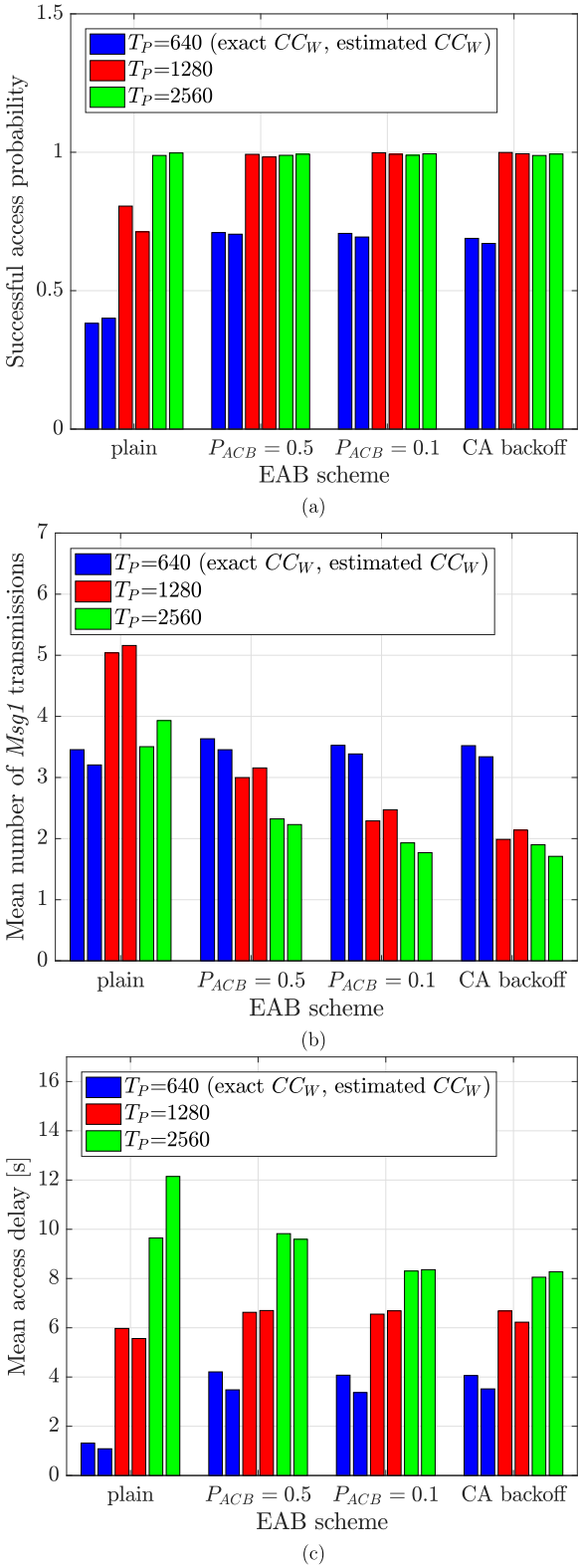

FIGURE 12. **Comparison of the KPIs with exact** $\mathbf{CC_W}$ **and estimated** $\mathbf{CC_W}$**, for** $\boldsymbol{N_M = 30\,000}$**, and** $\boldsymbol{(T_P, T_{SIB14}) = (\{640, 1280, 2560\}, 320)}$ **ms. (a) Successful access probability. (b) Mean number of Msg1 transmissions. (c) Mean access delay.**

VII. CONCLUSION

We have shown that the current extended access barring (EAB) suffers from a limiting factor which is that, every time a barred access class is released, UEs belonging to that class access the channel in bursts. We have evaluated two enhanced EAB schemes to remove this burstiness: a first one based on the joint use of EAB and access class barring, and a second one based on the execution of a congestion avoidance backoff after the barring status of a UE is switched to unbarred. We showed that these schemes improve the key performance indicators and that the scheme based on a congestion avoidance backoff, besides being simpler, is the most effective.

The congestion avoidance backoff scheme offer two important benefits: firstly, it provides very high successful access probability, even in heavily congested scenarios, with shorter paging cycles than those required by plain EAB, what results in a very important reduction of the access delay; and secondly, it reduces drastically the number of preamble transmissions required to access the channel. In particular, the important reduction in the number of preamble transmissions makes this solution a well-suited option for power-constrained MTC devices.

Finally, we presented an accurate estimation method to calculate the congestion coefficients using only the information available at the base station, as would be required in a realistic setting. We showed that the application of our method permits a realistic and effective implementation of all the studied EAB schemes, and in particular of the congestion avoidance backoff scheme.

REFERENCES

[1] Qualcomm. (Jul. 2016). *Paving the Path to Narrowband 5G with LTE Internet of Things (IoT)*. [Online]. Available: http://www.5gamericas.org/files/1914/6862/6516/Whitepaper_-_Paving_the_path_to_Narrowband_5G_with_LTE_Internet_of_Things_-_v1.0.pdf
[2] *Architecture Enhancements to Facilitate Commun.with Packet Data Networks and Applications*, document TS 23.682, 3GPP, Sep. 2017.
[3] Ericsson. (Nov. 2017). *Ericsson Mobility Report*. [Online]. Available: https://www.ericsson.com/en/mobility-report/reports/november-2017
[4] C. Bockelmann *et al.*, "Towards massive connectivity support for scalable mMTC communications in 5G networks," *IEEE Access*, vol. 6, pp. 28969–28992, 2018.
[5] P. Marsch, O. Bulakci, O. Queseth, and M. Boldi, *5G System Design: Architectural and Functional Considerations and Long Term Research*. Hoboken, NJ, USA: Wiley, 2018.
[6] *Medium Access Control (MAC) Protocol Specification*, document TS 36.321, 3GPP, Sep. 2017.
[7] *Physical Channels Modulation*, document TS 36.211, 3GPP, Sep. 2017.
[8] L. Tello-Oquendo *et al.*, "Performance analysis and optimal access class barring parameter configuration in LTE-A networks with massive M2M traffic," *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 3505–3520, Apr. 2018.
[9] *Radio Resource Control (RRC), Protocol Specification*, document TS 36.331, 3GPP, Sep. 2017.
[10] *Feasibility Study for Further Advancements for E-UTRA*, document TR 36.912, 3GPP, Mar. 2017.
[11] *Service Accessibility*, document TS 22.011, 3GPP, Sep. 2017.
[12] J. Schlienz and D. Raddino, "Narrowband Internet of Things. Whitepaper," Rohde & Schwarz, Munich, Germany, Tech. Rep. NarrowBand_IoT—1MA266_0e, 2016. [Online]. Available: https://www.rohde-schwarz.com/us/applications/narrowband-internet-of-things-white-paper_230854-314242.html
[13] A. Larmo and R. Susitaival, "RAN overload control for machine type communications in LTE," in *Proc. IEEE Globecom Workshops*, Dec. 2012, pp. 1626–1631.
[14] *Study RAN Improvements for Machine Type Communications*, document TR 37.868, 3GPP, Sep. 2011.
[15] *User Equipment (UE) Procedures in Idle Mode, V14.4.0*, document TS 36.304, 3GPP, Sep. 2017.
[16] *Further Performance Evaluation of EAB Information Update Mechanisms*, document 77 R2-120270, 3GPP, Feb. 2012.
[17] R.-G. Cheng, J. Chen, D.-W. Chen, and C.-H. Wei, "Modeling and analysis of an extended access barring algorithm for machine-type communications in LTE-A networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 2956–2968, Jun. 2015.
[18] Z. Zhang, H. Chao, W. Wang, and X. Li, "Performance analysis and UE-side improvement of extended access barring for machine type communications in LTE," in *Proc. IEEE Veh. Technol. Conf. (VTC Spring)*, May 2014, pp. 1–5.
[19] U. Phuyal, A. T. Koc, M.-H. Fong, and R. Vannithamby, "Controlling access overload and signaling congestion in M2M networks," in *Proc. IEEE ASILOMAR*, Nov. 2012, pp. 591–595.
[20] W. T. Toor and H. Jin, "Comparative study of access class barring and extended access barring for machine type communications," in *Proc. IEEE ICTC*, Oct. 2017, pp. 604–609.
[21] H. Kim, S.-S. Lee, and S. Lee, "Dynamic extended access barring for improved M2M communication in LTE-A networks," in *Proc. IEEE SMC*, Oct. 2017, pp. 2742–2747.
[22] C. M. Chou, C. Y. Huang, and C.-Y. Chiu, "Loading prediction and barring controls for machine type communication," in *Proc. IEEE ICC*, Jun. 2013, pp. 5168–5172.
[23] C.-W. Chang, J.-C. Chen, C. Chen, and R.-H. Jan, "Scattering random-access intensity in LTE machine-to-machine (M2M) communications," in *Proc. IEEE Globecom*, Dec. 2013, pp. 4858–4863.
[24] C.-Y. Oh, D. Hwang, and T.-J. Lee, "Joint access control and resource allocation for concurrent and massive access of M2M devices," *IEEE Trans. Wireless Commun.*, vol. 14, no. 8, pp. 4182–4192, Aug. 2015.
[25] J. Choi, "NOMA-based random access with multichannel ALOHA," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2736–2743, Dec. 2017.
[26] M. Shirvanimoghaddam *et al.*, "Massive non-orthogonal multiple access for cellular IoT: Potentials and limitations," *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 55–61, Sep. 2017.
[27] M. Shirvanimoghaddam, M. Condoluci, M. Dohler, and S. J. Johnson, "On the fundamental limits of random non-orthogonal multiple access in cellular massive IoT," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 10, pp. 2238–2252, Oct. 2017.
[28] J.-B. Seo, B. C. Jung, and H. Jin, "Nonorthogonal random access for 5G mobile communication systems," *IEEE Trans. Veh. Technol*, vol. 67, no. 8, pp. 7867–7871, Aug. 2018.
[29] J.-B. Seo, B. C. Jung, and H. Jin, "Performance analysis of NOMA random access," *IEEE Commun. Lett.*, vol. 22, no. 11, pp. 2242–2245, Nov. 2018.
[30] C. H. Wei, G. Bianchi, and R. G. Cheng, "Modeling and analysis of random access channels with bursty arrivals in OFDMA wireless networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 4, pp. 1940–1953, Apr. 2015.
[31] J. Choi, "Channel-aware multichannel random access for energy-limited sensors and MTC devices," *IEEE Access*, vol. 6, pp. 59929–59939, 2018.

JOSÉ-RAMÓN VIDAL received the Ph.D. degree in telecommunication engineering from the Universitat Politècnica de València (UPV), Valencia, Spain. He is currently an Associate Professor with the Department of Communications, Higher Technical School of Telecommunication Engineering, UPV. His current research interest includes the application of game theory to resource allocation in cognitive radio networks and to economic modeling of telecommunication service provision. In these areas, he has authored or coauthored several papers in refereed journals and conference proceedings.

LUIS TELLO-OQUENDO received the B.E. degree (Hons.) in electronic and computer engineering from the Escuela Superior Politécnica de Chimborazo (ESPOCH), Ecuador, in 2010, and the M.Sc. degree in telecommunication technologies, systems, and networks and the Ph.D. degree in telecommunications from the Universitat Politècnica de València (UPV), Spain, in 2013 and 2018, respectively. From 2013 to 2018, he was a Graduate Research Assistant with the Broadband Internetworking Research Group, UPV. From 2016 to 2017, he was a Research Scholar with the Broadband Wireless Networking Laboratory, Georgia Institute of Technology, Atlanta, GA, USA. He is currently a Research Faculty Member with the College of Engineering, Universidad Nacional de Chimborazo. His research interests include machine-type communications, wireless software-defined networks, 5G and beyond cellular systems, the Internet of Things, and machine learning. He is a member of ACM. He received the Best Academic Record Award from the Escuela Técnica Superior de Ingenieros de Telecomunicación, UPV, in 2013, and the IEEE ComSoc Award for attending the IEEE ComSoc Summer School at The University of New Mexico, Albuquerque, NM, USA, in 2017.

VICENT PLA received the B.E. and M.E. degrees in telecommunication engineering and the Ph.D. degree from the Universitat Politècnica de València (UPV), Valencia, Spain, in 1997 and 2005, respectively, and the B.Sc. degree in mathematics from the Universidad Nacional de Educación a Distancia, Madrid, Spain, in 2015. In 1999, he joined the Department of Communications, UPV, where he is currently a Professor. His research interest includes modeling and performance analysis of communication networks. During the past few years, most of his research activity has focused on traffic and resource management in wireless networks. In these years, he has authored or coauthored numerous papers in refereed journals and conference proceedings and has been an active participant in several research projects.

LUIS GUIJARRO received the M.Eng. and Ph.D. degrees in telecommunications from the Universitat Politècnica de València (UPV), Valencia, Spain. He is currently an Associate Professor in telecommunications policy with the Department of Communications, UPV. He has authored or coauthored the book *Electronic Communications Policy of the European Union* (UPV, 2010). He researched in traffic management in ATM networks and in e-Government. His current research interest includes economic modeling of telecommunication service provision. He has contributed in the areas of peer-to-peer interconnection, cognitive radio networks, search engine neutrality, wireless sensor networks, and 5G.