A dynamic access class barring method to avoid congestion in LTE-A networks with massive M2M traffic

Israel Leyva-Mayorga, Luis Tello-Oquendo, Vicent Pla, Jorge Martinez-Bauset, and Vicente Casares-Giner

ITACA, Universitat Politècnica de València,
Camino de Vera s/n. 46022, Valencia, Spain.
email: {isleyma,luiteloq,vpla,jmartinez,vcasares}@upv.es

Abstract. In the near future, a massive number of machine-to-machine (M2M) communication devices will provide with ubiquitous information and services. Nevertheless, the actual infrastructure of relaying networks may not be capable of handling such a large number of interconnected user equipments (UEs). This is the case of LTE-A networks, where the random access channel suffers from congestion whenever a bulk of UEs transmit in a highly synchronized manner. As such, the use of congestion control methods, such as access class barring (ACB), is needed. But ACB reduces congestion in exchange of a higher access delay. Hence, maintaining an active ACB induces unnecessary delay to the access of UEs during periods of low congestion. In this paper, we present a novel approach for the dynamic modification of ACB parameters that enhances the UE accesses during periods of high congestion and avoids excessive delay during periods of low congestion.

Keywords: Access class barring; dynamic congestion-control; LTE-A networks; machine-to-machine (M2M) communications.

1 Introduction

Modern society is in the need for ubiquitous device connectivity, where small devices, known as user equipments (UEs), exchange data autonomously to provide continuous access to information and services. Machine-to-machine (M2M) communication stands for the autonomous exchange of data between UEs and is a fundamental component of the Internet of Things (IoT) [2]. But, M2M applications pose important engineering challenges regarding the signaling capabilities of relaying networks. For instance, LTE-A networks present the best option for UE interconnection as its infrastructure has already been largely deployed [7]. Nevertheless, it has been observed that the signaling capabilities of cellular base stations (evolved NodeBs, eNBs, in LTE-A) can be exceeded when a bulk of UEs transmit in a highly synchronized manner (this is a typical behavior in M2M applications) as the random access channel (RACH) of LTE-A was designed to

handle human-to-human communications, where a few UEs (when compared to M2M communications) attempt to access the eNB [5].

The UEs access the eNB by means of the random access procedure (RAP), which comprises a four-message handshake; i.e., preamble transmission (only permitted during random access opportunities, RAOs), random access response (RAR), connection request and contention resolution messages. The RAP will be described in Section 3. The performance of the LTE-A system is measured in terms of several performance indicators, which determine whether the access is being conducted efficiently. These include, among others, access success probability, access delay and the number of access attempts. The degradation of these performance indicators occurs whenever the network suffers from congestion, which is by no means desirable. To enhance the performance of the RACH, several methods have been proposed in the literature.

Access class barring (ACB) is a congestion-control method that redistributes the UE accesses through time. For this, a portion of the UEs delay the transmission of the first message of the RAP according to the barring rate and the mean barring time broadcast by the eNB. As a matter of fact, when ACB is implemented and correctly configured, the congestion in the RACH can be reduced in exchange of a longer wait to access the eNB. But maintaining a constant barring rate and barring time during periods of low congestion leads to a notorious and unnecessary increase in access delay. Building on this, several dynamic ACB methods that periodically modify its parameters have been proposed [12, 9]. But the dynamic selection of ACB parameters and, more importantly, the mechanism to activate/disable these methods is not straightforward.

In this paper, we propose and evaluate a dynamic ACB method (DACB) that relies on the periodic calculation of a load coefficient (LC). The purpose of our DACB method is to prevent congestion in the RACH. By doing so, the access success probability of UEs is maximized during sporadic periods of high congestion, while the access of UEs during periods of low congestion is not significantly affected.

The rest of the paper is organized as follows. The review of the literature of massive M2M communications through LTE-A and congestion-control methods is presented in Section 2. Then, we describe the random access procedure, along with the traditional ACB and our dynamic ACB method (DACB) in Section 3. The basic configuration of the random access channel (RACH) and the selected traffic models are described in Section 4. Results derived from our performance analysis are presented in Section 5. The article concludes with the discussion of results and future work.

2 Related work

Several studies have concluded that the current RA procedure (RAP) of LTE-A is not capable of handling massive M2M communications [1, 8, 10, 13]. As such, several methods have been proposed to enhance the performance of the LTE-A RACH [5, 6]. Among these, access class barring (ACB) methods are part of the

LTE-A specification [3]. The purpose of these methods is to spread the UE accesses through time by restricting the access of certain classes of UEs. ACB is oftentimes modeled as a static method, in which the configuration parameters remain the same throughout its operation, and most studies agree with its efficiency during sporadic periods of congestion [6, 5]. Nevertheless, to the best of our knowledge, no study has evaluated the impact on access delay of a static implementation of ACB during periods of low congestion, where an unnecessary access delay may be induced.

The dynamic modification of ACB parameters has been proposed previously as an efficient solution for enhancing the access success probability with a minimum increase in access delay [11, 9]. This is the case of [12], where the authors propose a dynamic barring method in which the ideal barring factor is calculated at each RAO. To calculate the ideal barring factor, the authors estimate the number of contending UEs based on the probability distribution of successful and failed accesses. But this information may not be known by the eNB. This is recognized in [9], where the authors propose the use of a state transition diagram for the dynamic activation of a barring method. Here, the state of the system depends on the average number of successful preamble transmissions and the dynamic barring method is activated when the system reaches the state of severe congestion. However, the authors do not consider that the maximum number of uplink grants may be lower than the number of successful preambles, which is a relevant limitation of the RACH. In this paper, we propose a simple dynamic ACB method (DACB) that considers the limitations of the RACH.

3   Random access procedure

The UEs access the eNB by means of the RAP, which comprises a four-message handshake. First, the timing configuration of the random access channel (RACH) must be obtained by the UEs through the System Information Blocks, which are periodically broadcast by the eNB. Then:

Msg1: The UEs perform the random selection and transmission of one out of the R available preambles towards the eNB during one of the available random access opportunities (RAOs). These preambles are orthogonal sequences that lack any type of identification field. A collision occurs if the same preamble is transmitted by multiple UEs in the same RAO, i.e., the eNB is unable to detect the transmitted preamble [1].

Msg2: The eNB sends up to N_{UL} uplink grants to the UEs per RAO (up to one for each detected preamble) by means of random access response (RAR) messages. Uplink grants assign time-frequency resources to the UEs for the transmission of *Msg3*. After the transmission of *Msg1*, the UEs wait for a random access response window, W_{RAR}, to receive the uplink grant. If the uplink grant is not received within the W_{RAR}, the UE shall increase its preamble transmission counter, k, ramp up its power, perform backoff and go back to the transmission of *Msg1* if the maximum number of preamble transmissions, *preambleTransMax*, has not been reached. Otherwise, the RAP is terminated.

Msg3: The UEs that received an uplink grant send its connection request message (*Msg3*) during the time-frequency resources specified by the eNB. As such, no collisions can occur during the transmission of this message.

Msg4: Finally, the eNB responds to each *Msg3* transmission with a contention resolution message, *Msg4*. This is the end of the RAP and now the UEs proceed with the transmission of their data packets towards the eNB.

3.1 Access class barring (ACB)

Access class barring (ACB) is a congestion control method that aims to redistribute the first preamble transmission of UEs through time to reduce the number of access requests per RAO. In ACB, every UE that belongs to the normal classes (0 to 9) shall acquire the mean barring time, T_{ACB}, and barring rate, P_{ACB}, from the eNB through the System In-

Fig. 1. Four-message handshake in the contention-based random access procedure (RAP) of LTE-A.

formation Block Type 2 (SIB2). SIB2 is broadcast periodically by the eNB. Then, the UEs perform ACB checks at the beginning of the RAP; i.e., only before its first preamble transmission. At each ACB check, the UEs transmit its preamble with probability P_{ACB}. Otherwise, the UE waits for a random time,

$$T_b = (0.7 + 0.6 \times U(0,1)) \times T_{\mathrm{ACB}} \text{ seconds.} \tag{1}$$

In highly congested scenarios, the static implementation of ACB has proven to be a valuable congestion control method given an adequate selection of P_{ACB} and T_{ACB} In the practice, the implementation of a static ACB induces an unnecessary delay in the UE accesses during periods of low or no congestion.

3.2 Dynamic ACB (DACB)

In this section we present a dynamic access class barring method (ACB) and the load coefficient (LC) that is used to calculate the barring rate at the ith RAO, $P_{\mathrm{ACB}}[i]$. It is worth noting that in this study we select a constant $T_{\mathrm{ACB}} = 1$ s as its dynamic selection is not straightforward. Herein we assume that the eNB updates LC at each RAO and sends this information to the UEs as in [12].

At the end of the ith RAO, the eNB is only aware of the number of successfully decoded preambles, $N_{sp}[i]$, and the number of uplink grants that will be sent within the next W_{RAR}, $N_s[i]$. Other information such as the number of UEs that transmit its kth preamble at the ith RAO, $N[i,k]$, the total number
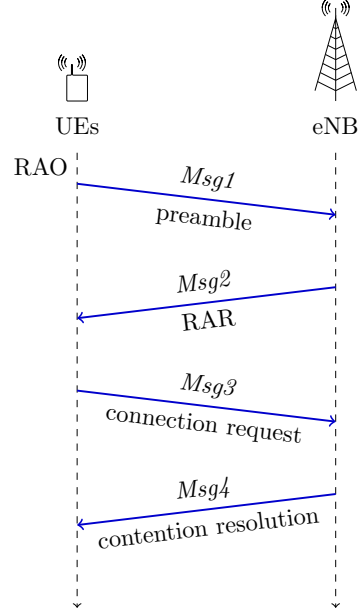
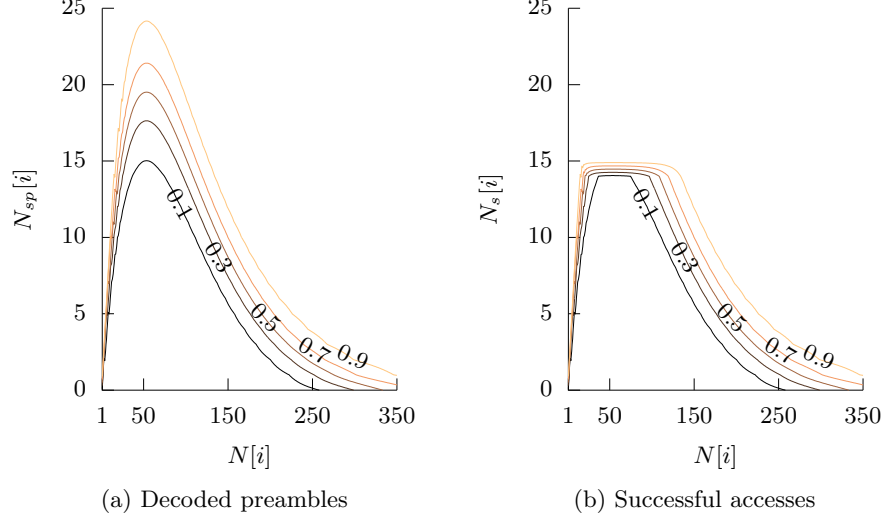(a) Decoded preambles   (b) Successful accesses

Fig. 2. Cumulative distribution function (CDF) of the decoded preambles, $N_{sp}[i]$, and the successful access, $N_s[i]$, given $N[i] \in \{1, 2, \ldots, 350\}$.

of transmitted preambles per RAO and the number of colliding UEs is certainly not known by the eNB. Building on this, the eNB can only measure the traffic load in terms of $N_{sp}[i]$ and $N_s[i]$. Fig. 2a and Fig. 2b show the cumulative distribution function (CDF) of $N_{sp}[i]$ and $N_s[i]$ for $N[i] \in \{1, 2, \ldots, 350\}$ in a typical RACH configuration (further explained in Section 4), where the number of available preambles is $R = 54$ and the number of uplink grants per RAO is $N_{UL} = 15$.

Instead of attempting to control the congestion once it has occurred, we adopt a preventive approach towards the use of ACB. For this, we propose the use of a load coefficient, given as

$$\ell[i] = \frac{\max\{N_s[i] - 1, 0\}}{N_{UL} - 1},\tag{2}$$

to control the barring rate at each RAO as

$$P_{\mathrm{ACB}}[i] = 1 - \ell[i].\tag{3}$$

The block diagram for the implementation of the DACB is shown in Fig. 3.2.

Please observe that the access of UEs will not be affected if $N_s[i] \in \{0, 1\}$ since $\ell[i] = 0$. This is a typical scenario of low congestion in LTE-A, where few preamble transmissions per second occur. Therefore, the UEs would not experience unnecessary access delay when the traffic load is low. $\ell[i]$ will increase with the number of preamble transmissions per RAO, $N[i]$, until a maximum of one when $N_s[i] = N_{UL}$. Given the maximum load coefficient, $\ell[i] = 1$, is obtained,
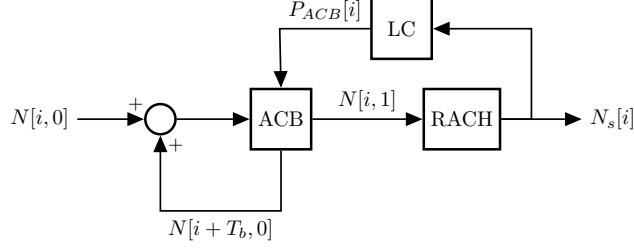
Fig. 3. Block diagram for the implementation of a dynamic ACB (DACB).

Table 1. RACH configuration.

Parameter	Setting
RAO periodicity	$T_{\mathrm{RAO}} = 5$ ms
Number of available preambles	$R = 54$
RAR window length	$W_{\mathrm{RAR}} = 5$ subframes
Maximum number of uplink grants per subframe	$N_{\mathrm{RAR}} = 3$
Maximum number of uplink grants per RAO	$N_{UL} = W_{\mathrm{RAR}} \times N_{\mathrm{RAR}} = 15$
Maximum number of preamble transmissions	$preambleTransMax = 10$
Preamble detection probability for the kth preamble transmission	$P_d = 1 - 1/e^k$
Backoff Indicator	$B = 20$ ms

$P_{\mathrm{ACB}}[i] = 0$, hence the first preamble transmissions will not be permitted until $\ell[i]$ drops. From Fig. 2b, it can be seen that our DACB will react accordingly given $N[i] \leq 130$. Therefore, the main objective of our DACB method is to hinder the growth of $N[i]$ beyond 130.

4 RACH configuration and selected traffic model

Throughout this study we assume the basic configuration of the RACH as suggested in [1]. Table 1 shows the selected RACH configuration.

In this configuration (PRACH Configuration Index 6), the duration of a subframe is 1 ms and RAOs occur every $T_{RAO} = 5$ ms. The number of available preambles, R, determines the probability distribution of successful preambles, $N_{sp}[i]$, in the ith RAO for a given number of preamble transmissions, $N[i]$ (see Fig. 2a). $N_{UL} = W_{\mathrm{RAR}} \cdot N_{\mathrm{RAR}}$ determines the maximum number of uplink grants that the eNB can transmit within the RAR window (for each RAO). Note that only the UEs that receive an uplink grant may proceed with the transmission of *Msg3* and *Msg4*; therefore, the maximum number of successful UE access per RAO for any given $N[i]$ is $\max\{N_s[i]\} = N_{UL}$ as observed in Fig. 2b. The preambles transmitted by single UEs are detected by the eNB with probability P_d. The backoff time of failed UEs is calculated randomly as

$$T_{\mathrm{B}} = U(0, B), \tag{4}$$

Table 2. M2M traffic models for RACH Evaluation [1]

Characteristics	Traffic model 1	Traffic model 2
Number of M2M UEs (N)	1000, 3000,..., 180000	30000
Arrival distribution over T	Uniform	$Beta(3,4)$
Distribution period, T	12000 RAOs	2000 RAOs

where B is the *Backoff Indicator* broadcast by the eNB.

The 3GPP has proposed two traffic models for RACH evaluation in [1]. Traffic model 1 corresponds to a typical scenario, where the UE arrivals are uniformly distributed through time within $T = 12000$ RAOs (60 seconds). While the 3GPP defines the maximum number of M2M UEs in the traffic model 1 as 30000, we use this model to evaluate the effect of ACB on the access delay of UEs during periods of low congestion, when up to 180000 M2M UEs access the eNB. On the other hand, Traffic model 2 corresponds to a highly congested scenario, where $N = 30000$ M2M UEs arrivals follow a $Beta(3,4)$ distribution within $T = 2000$ RAOs (10 seconds). We use this model to evaluate the effect of ACB on the access success probability, P_s, and the access delay of UEs that access the eNB during periods of high congestion. Table 2 shows the traffic models and the number of M2M UEs considered during this study.

5   Performance analysis

In this section we evaluate the performance of the RACH and compare the efficiency of a static ACB and our DACB method. For this, we perform simulations of the RAP until the cumulative results obtained in the jth simulation differ from those obtained in the $j-1$th simulation by less than 1%. First, we evaluate the performance of the RACH given that the UE access the eNB according to traffic model 2, depicted in Table 2. Fig. 5 shows the average number of UE arrivals, $N[i,0]$[1] and the average number of successful accesses per RAO, $N_s[i]$, for three cases in the given scenario. In the first, no ACB is implemented, and it can be clearly seen that the performance of the RACH is degraded when $N[i,0]$ exceeds $N_{UL} = 15$. Consequently, the access success probability if no ACB is implemented is $P_s = 31.305\%$, which is by no means desirable.

In the second, a static ACB is implemented. We have selected $P_{\mathrm{ACB}} = 0.5$ and $T_{\mathrm{ACB}} = 4$ s for the static ACB because it leads to the lowest access delay while maintaining an adequate $P_s \geq 95\%$, concretely $P_s = 97.44\%$. In the third, our DACB is implemented and even a higher $P_s = 99.93\%$ is achieved. Such high access success probabilities are achieved because both ACB methods successfully reduce the number of first preamble transmissions, $N[i,1]$, below N_{UL}. Also, in Fig. 5 we can observe a noticeable difference in the shape of $N_s[i]$ between both ACB methods. For the static ACB, $N_s[i]$ rises rapidly and almost reaches

---

[1] In case no ACB is implemented, the average number of UE accesses, $N[i,0]$, is equal to the average number of first preamble transmissions, $N[i,1]$, in the ith RAO.
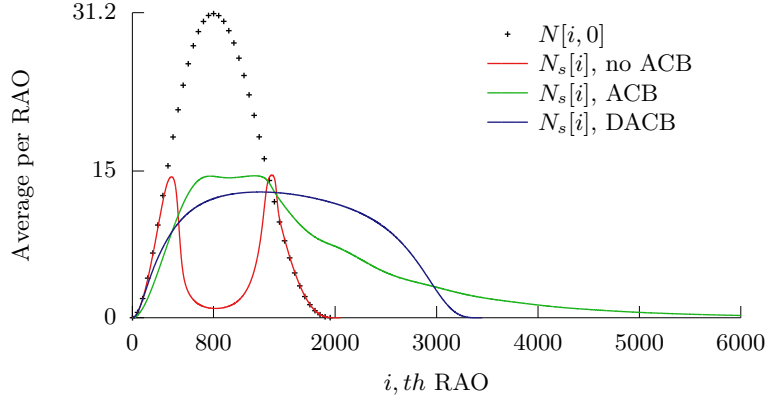
Fig. 4. Average number of UE accesses, $N[i, 0]$, and the average number successful UE accesses, $\overline{N_s[i]}$, given the implementation of no ACB, a static ($P_{\mathrm{ACB}} = 0.5$ and $T_{\mathrm{ACB}} = 4$ s) and the dynamic ACB (DACB), traffic model 2.

$N_{UL} = 15$. Then, after $N[i, 0]$ drops below N_{UL}, $N_s[i]$ drops slowly. For the DACB, $N_s[i]$ rises rapidly but stops at around 12.9 successful accesses per RAO. Then, after 2000 RAOs, $N_s[i]$ drops rapidly.

The difference in the shape of $N_s[i]$ between both methods results in a noticeable disparity in the probability distribution of access delay. To closely observe this performance indicator, we have obtained the cumulative distribution function (CDF) of access delay based on the timing values of Table 16.2.1-1 in [4]. In Fig. 5, it can be seen that the initial growth of the CDF of the static backoff is more rapid than that of our DACB method. However, this growth stops just below 0.5 and the CDF of the DACB is higher after 2 s have elapsed. As a result, high percentiles of access delay are much lower for the DACB than that of the static ACB. Hence, our DACB method presents a lower access delay than the best possible combination of P_{ACB} and T_{ACB} for the static ACB in an extremely congested scenario (traffic model 2).

Finally, we are interested in evaluating the effect of both ACB methods in access delay during periods of low congestion, described by traffic model 1 (see Table 2). Specifically, we assess the access delay in these scenarios in terms of the 95th percentile, t_{95} in seconds; i.e., 95% of the UEs experience an access delay lower than or equal to t_{95}. As Fig. 6 shows, the static ACB causes the exact same t_{95} regardless of the number of UEs that access the eNB per RAO. In other words, if a static ACB is implemented, 5% of the total UE accesses will be delayed, at least, 15.6 seconds in any scenario, which is by no means desirable. On the other hand, the use of our DACB method sharply reduces t_{95}. In fact, the lowest $t_{95} = 0.057$ s is achieved when $N[i] = 1$ and increases with $N[i]$, but is always lower than the t_{95} achieved by the static ACB.
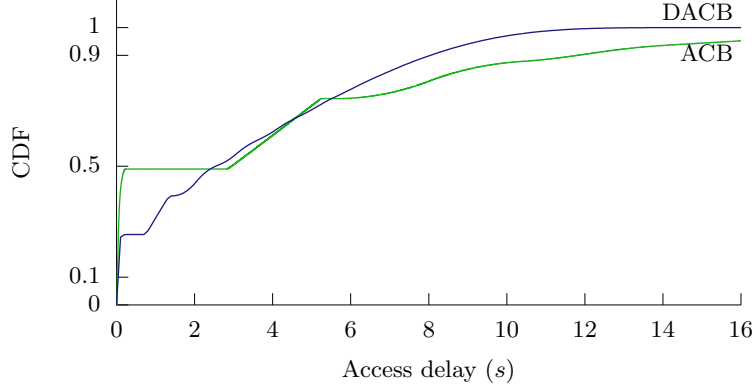
67

Fig. 5. Cumulative distribution function (CDF) of access delay for the static ($P_{\mathrm{ACB}} = 0.5$ and $T_{\mathrm{ACB}} = 4$ s) and for the DACB, traffic model 2.
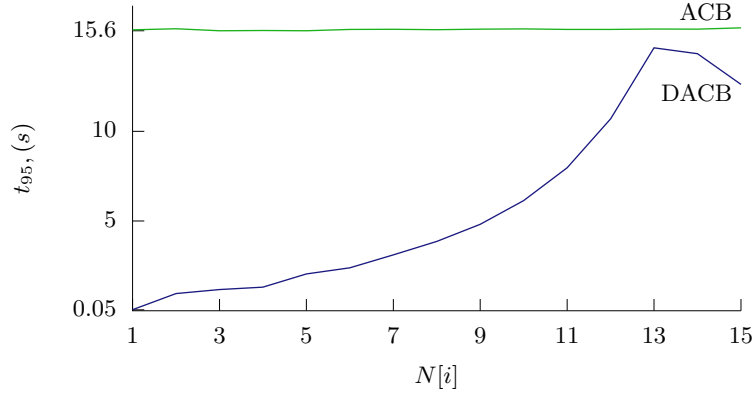

Fig. 6. 95th percentile of access delay, t_{95}, for the static ($P_{\mathrm{ACB}} = 0.5$ and $T_{\mathrm{ACB}} = 4$ s) and the dynamic ACB, traffic model 1.

6 Conclusion

The current random access procedure of LTE-A is not efficient at handling massive M2M communications. As such, congestion control methods must be implemented. Without a doubt, static ACB methods sharply enhance the performance of LTE-A during periods of extreme congestion. However, these methods greatly affect the access of UEs during periods of low or no congestion. By comparing our DACB method with the best possible implementation of a static ACB, we have observed that the former leads to a higher access success probability and lower access delay in extremely congested scenarios. Our DACB method also maintains an acceptable access delay in scenarios where a few UEs access the eNB, whereas implementing a static ACB sharply increases this parameter. Future work includes the performance analysis of our DACB method given that the calculation of the LC and the update of the barring rate at the UEs can only be

performed at certain intervals. In addition, the use of LCs from previous RAOs for the calculation of the barring rate may increase the reliability of the DACB.

Acknowledgment

References

1. 3GPP: TR 37.868, Study on RAN Improvements for Machine Type Communications (2011)
2. 3GPP: TS 22.368, Service Requirements for Machine-Type Communications (Dec 2014)
3. 3GPP: TS 36.331, Radio Resource Control (RRC), Protocol specification (2014)
4. 3GPP: TR 36.912 Feasibility study for Further Advancements for E-UTRA (2016)
5. Biral, A., Centenaro, M., Zanella, A., Vangelista, L., Zorzi, M.: The challenges of M2M massive access in wireless cellular networks. Digit. Commun. Netw. 1(1), 1–19 (2015)
6. Ferdouse, L., Anpalagan, A., Misra, S.: Congestion and overload control techniques in massive M2M systems: a survey. Trans. Emerg. Telecommun. Technol. (2015)
7. Ghavimi, F., Chen, H.H.: M2M Communications in 3GPP LTE/LTE-A Networks: Architectures, Service Requirements, Challenges, and Applications 17(2), 525–549 (May 2015)
8. Lin, G.Y., Chang, S.R., Wei, H.Y.: Estimation and Adaptation for Bursty LTE Random Access. IEEE Trans. Veh. Technol. 65(4), 2560–2577 (2016)
9. Lin, T.M., Lee, C.H., Cheng, J.P., Chen, W.T.: PRADA: Prioritized random access with dynamic access barring for MTC in 3GPP LTE-A networks. IEEE Trans. Veh. Technol. 63(5), 2467–2472 (2014)
10. Osti, P., Lassila, P., Aalto, S., Larmo, A., Tirronen, T.: Analysis of PDCCH Performance for M2M Traffic in LTE. IEEE Trans. Veh. Technol. 63(9), 4357–4371 (2014)
11. Suyang Duan, Shah-Mansouri, V., Wong, V.W.S.: Dynamic access class barring for M2M communications in LTE networks. In: Proc. IEEE Global Communications Conference (GLOBECOM). pp. 4747–4752 (2013)
12. Tavana, M., Shah-Mansouri, V., Wong, V.W.S.: Congestion control for bursty M2M traffic in LTE networks. In: Proc. IEEE International Conference on Communications (ICC). pp. 5815–5820 (2015)
13. Wei, C.H., Bianchi, G., Cheng, R.G.: Modeling and analysis of random access channels with bursty arrivals in OFDMA wireless networks. IEEE Trans. Wirel. Commun. 14(4), 1940–1953 (2015)