

Partial Contention Free Random access Protocol for M2M communications in LTE Networks

Ahmed Samir¹, Mahmoud M. Elmesalawy², Ahmed Salah EL-Din Ali^{2,*}, Ihab Ali²

¹IT Dept., Etisalat-Egypt, Cairo, Egypt

²Electronics, Communications and Computers Department, Helwan University, Cairo, Egypt

Abstract The current specifications of the Random Access Channel (RACH) of the Long Term Evolution (LTE) and LTE-Advanced (LTE-A) is not appropriate for Machine-to-Machine (M2M) communications. This is because the massive number of M2M devices will lead to low access rate and large access delay. This paper presents a new protocol, based on the Distributed Queuing (DQ) algorithm, for improving the performance of the Random Access (RA) procedure of LTE to support the M2M services. The proposed protocol, Distributed Queuing Access for LTE (DQAL), minimizes the opportunity of collision in the access phase for M2M. The reduction in the collision will turn in enhancing both the access rate and the access delay. Furthermore the presented scheme is designed to keep the normal access procedures of Human-to-Human (H2H) communication without any impact. Results show that the access delay is decreased to half of the LTE access delay and the access success probability approaches unity.

Keywords Distributed Queuing, LTE, Machine-to-Machine, Random Access Channel

1. Introduction

The M2M communication introduced a huge number of low rate connections with long sleeping period in order to save the power [1]. The current implementation of the RACH in the LTE and LTE-A is not suitable for this type of communication. This is one of the major reasons that made the 3GPP emphasize the need to revise the design of RACH in the next-generation cellular networks. Accordingly, there are many research works are proposed to improve the performance of the RACH procedure in LTE to support M2M communications [2-6].

Optimized MAC which is based on transmitting the data of M2M devices either in RACH message 1 (preamble) or in message 3 (terminal identification message) [2]. This will be suitable for M2M applications having very small amount of data. Access Class Barring (ACB) which is based on defining 16 different classes with different priorities and back-off time for various traffic types [3]. Indeed, it is specified as a mechanism to control the access to the air interface in LTE and LTE-A [4]. On the other hand, some studies coincide in suggesting that the ACB should not be used as a stand-alone solution to solve the congestion problem in the MTC networks [5] [6]. In dynamic allocation of RACH resources, the network allocates more Random

Access (RA) slots for M2M devices if the radio access is under congestion [7]. Slotted access schema is based on assigning dedicated access resources for each single M2M terminal is proposed in [8]. Prioritized Random Access (PRA) is based on assigning different number of virtual resources for different traffic classes [3]. In case of overload, there are certain classes will be barred to minimize the collision probability. Self-Optimized Overload Control (SOOC) is continuously configure the RA resources based on the load condition [2]. Code-expanded RA which increases the access opportunities by using the code expanded RA instead of using more preambles or RA slots [9].

All of the above proposals are based on ALOHA and slotted ALOHA multiple access techniques even though they have low throughput and possible instability under heavy load. There is another approach for improving the RACH which is based on the tree algorithm and the Distributed Queuing Random Access Protocol (DQRAP). The main concept of DQRAP depends on using a number of control minislots as access opportunities. The protocol operates as random access protocol in the light traffic case and switches automatically to a reservation protocol when traffic becomes heavy [10]. DQRAP was suggested by [11] to enhance the random access channel in Code Division Multiple Access (CDMA). It is also used to improve the throughput of the WLAN as proposed in [12]. All of these studies proved the stability of the protocol and its near-optimum behavior in terms of access delay, channel utilization and power consumption.

* Corresponding author:

ahmed.elkomy87@gmail.com (Ahmed Salah EL-Din Ali)

Published online at <http://journal.sapub.org/jwnc>

Copyright © 2016 Scientific & Academic Publishing. All Rights Reserved

In this paper, a distributed queuing based access protocol for LTE (DQAL) is proposed to improve the random access performance for M2M while keeping the current LTE frame structure unchanged. As a result, the implementation will be very smooth with no need to change the existing RACH operation of the normal H2H in the LTE network.

The remainder of this paper is organized as follows. In section 2, the proposed DQAL protocol and how the preamble status is detected as well as the proposed RA algorithm are presented. In section 3, we mathematically formulate the access delay and access success probability. The simulation results are given and analyzed in section 4. Finally, section 5 concludes this paper.

2. The Proposed DQAL

2.1. Protocol Description

In the current specification of LTE, the total number of preambles (N_{TL}) are divided into two parts, the contention based part (N_{CB}) and contention free (N_{CF}) part. The proposed access protocol is based on borrowing a part from the contention based preambles (N_{MD}) to be used by M2M devices (MDs) for access. In order to avoid any possible collision, H2H User Equipment (UE) is forced to see this part as a contention free using the notifications sent in System Information Block 2 (SIB2). This guarantees a normal RACH operation for H2H UEs. Figure 1 clarifies the preamble distribution from both H2H UEs and MDs perspectives.

Since original DQ protocol is based on using orthogonal minislots as access opportunities, direct implementation of it within LTE environment requires a change of the LTE frame structure. This is because the access opportunities in LTE (preambles) have not orthogonally in time. Therefore, the allocated number of preambles for MDs will be distributed among N_g virtual groups. Each virtual group has N_p preambles in which each preamble is equivalent to a minislot in the original DQRAP.

Figure 2 elaborates the LTE frame structure and the proposed DQAL virtual frame structure. As shown, each main frame consists from 10 subframes. On the other hand, the DQAL frame consisted from Collision Resolution Intervals (CRIs) at which the collision will be detected and reported by the eNodeB. When the MD wants to access the system, it selects and sends one of the preambles to the eNodeB. Once the preamble is detected by eNodeB, it will

send the Random Access Response (RAR) within time window of 5 subframes [13]. Then, the eNodeB should receive the corresponding message 3 within a time window of one frame; otherwise it will know that there is a collision in that preamble. Consequently, the system needs a time of more than one frame to detect the collision. Since the operation of the DQAL is based on broadcasting information about the detection status of preambles, two frames time interval is selected as a collision resolution interval as shown in Fig.2. Thus, the DQAL protocol is virtually see the LTE frame periodicity as a number of consecutive CRIs.

2.2. Preamble Status Detection

Since the main operation of DQAL protocol is based on identifying the status of each sent preamble, the eNodeB should broadcast the preamble status at the start of each CRI. There are three status listed below:

- i. **Successful:** If message 3 is sent successfully to eNodeB, then the eNodeB will know that the corresponding preamble is sent without collision.
- ii. **Collision:** During the RACH procedure, there will be two possibilities for collision either in the preamble (message 1) or in message 3. In the first case, the collision will be detected by eNodeB in which message 3 will not be received within the predefined window. In the second case, the MDs will detect the collision in which message 2 will not be received during the RAR window. This can be resolved by MD as it will randomly select one of the remaining groups that are not selected by the other MDs.
- iii. **Empty:** If a preamble is not used within the CRI by any MD, the eNodeB will consider its status as empty.

2.3. DQAL Algorithm

The main operation of the proposed DQAL protocol is based on the distributed collision resolution queue. All the MDs have two counters (RQ and pRQ), where RQ is the number of collisions waiting in the queue. The second counter pRQ is the position of each MD in the collision resolution queue. At the beginning of each CRI, each MD makes some initialization based on the broadcast information (RQ value and preambles detection status). This initialization called queuing discipline rule. Another rule is the request transmission rule, used to resolve the collision. The whole process of DQAL protocol is described in algorithm 1 below. And elaborated in fig. 3.

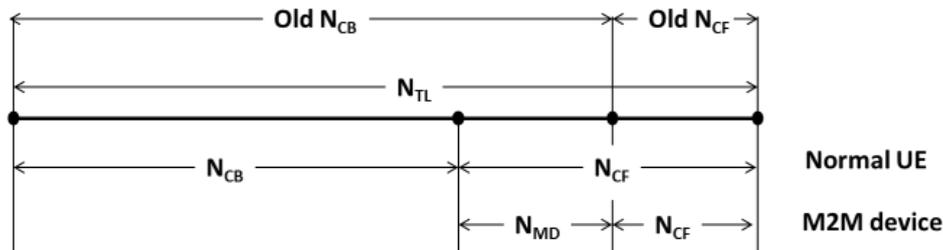


Figure 1. Preamble distribution over MDs and H2H UEs

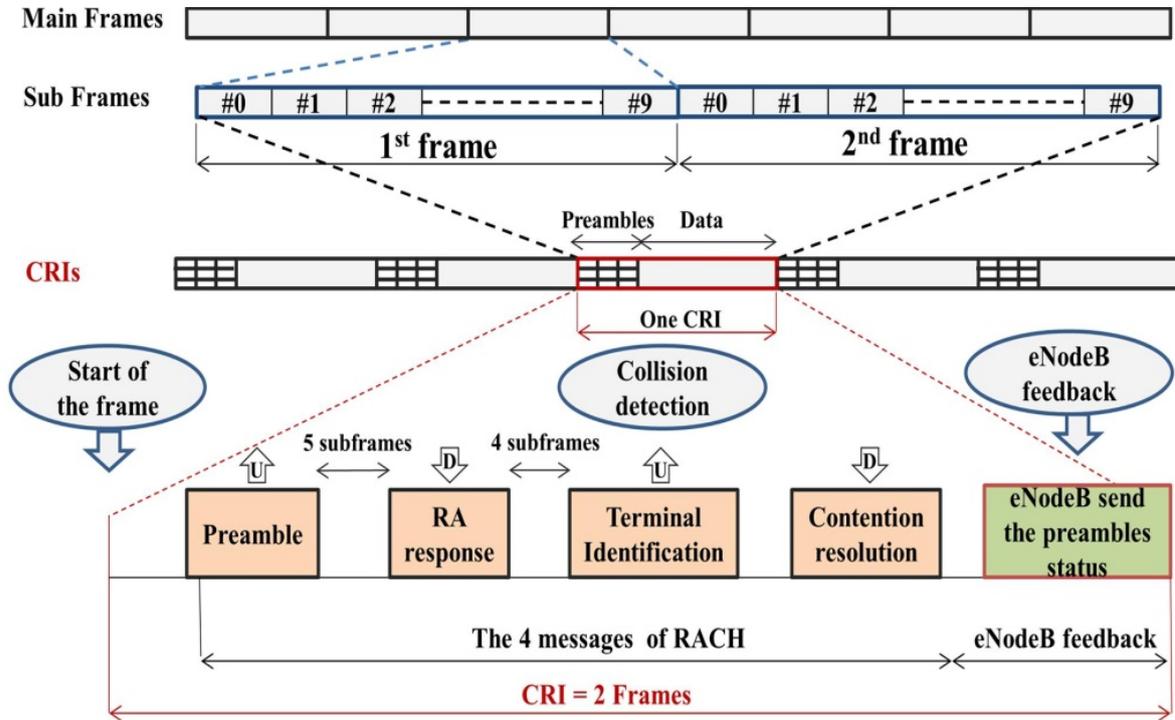


Figure 2. The Collision Resolution Interval (CRI) structure

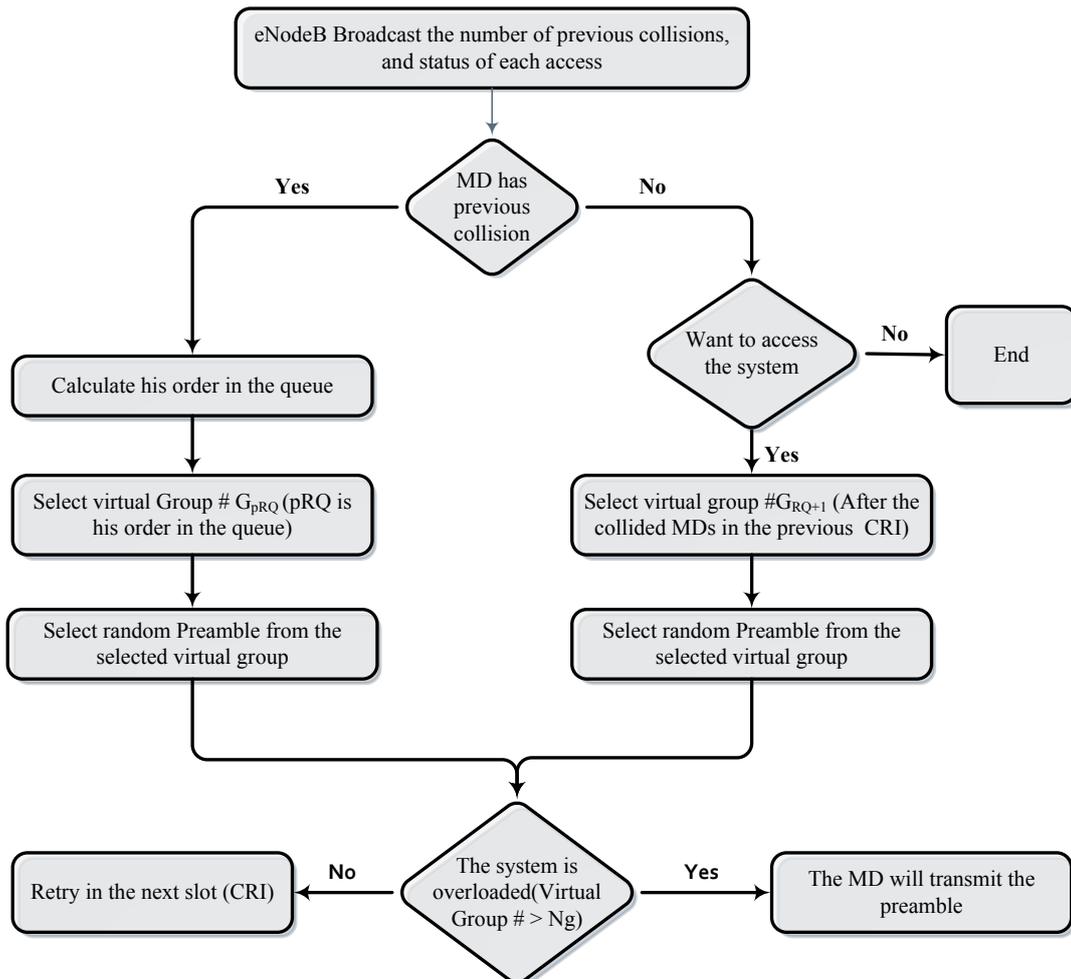


Figure 3. DQAL Flow chart

Algorithm: 1 Distributed queuing access protocol for LTE (DQAL)

- 1: Each step of the following will be repeated with every CRI
- 2: **Initialization**
- 3: eNodeB broadcast the RQ value and the preamble detection status
- 4: Each MD increment RQ by one for each collided preamble
- 5: **if** (RQ > 0) **then**
- 6: MDs Reduce the value of RQ by min (RQ, N_g)
- 7: **end if**
- 8: **if** (MD's preamble is collided) **then**
- 9: Calculate pRQ based on the preamble order
- 10: **end if**
- 11: **Preamble selection**
- 12: **if** (RQ < N_g) **do**
- 13: **if** (pRQ = 0) **then**
- 14: Randomly select one of preambles on group G_{RQ+1}
- 15: **else**
- 16: Randomly select one of preambles of group G_{pRQ}
- 17: **end if**
- 18: **else**
- 19: Wait to the next CRI
- 20: **end if**

The main strength of the proposed algorithm inherits on allowing the new arrivals to choose preambles from the groups above G_{RQ} . This helps to avoid any further collision with the existing collided requests waiting in the queue. Furthermore, the algorithm helps the MDs to select the appropriate virtual group that can be used to reduce the probability of collision. This will enhance both the access rate and access delay.

3. DQAL Model and Analysis

The DQAL can be modelled as $M/G/N_g$ queuing system as long as we used exact discrete access time distribution. Each virtual group represents one server in this model as shown in Fig. 4. The feedback line indicates that any collide preamble must enter the queue again until it being transmitted successfully. The Enable Transmission Interval (ETI) represents the time taken by the MD while it is waiting till the start of the next CRI in order to transmit its preamble.

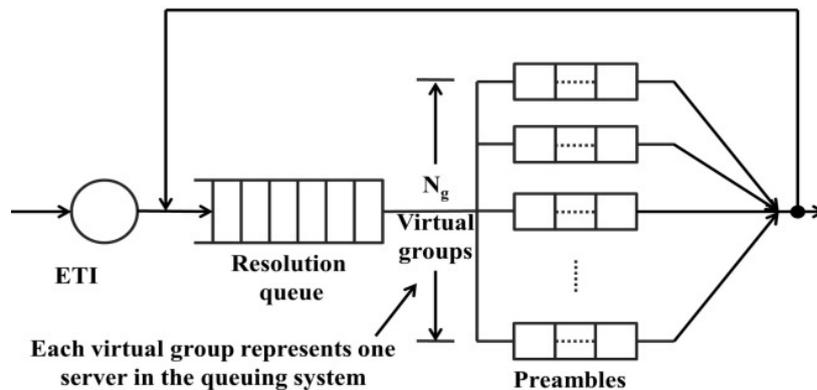


Figure 4. The DQAL model

3.1. Mean Access Delay

The total access delay (t_{TA}) in DQAL can be calculated based on the service time of the ETI (t_{ETI}) and the collision resolution delay (t_{RQ}) as follows,

$$E[t_{TA}] = E[t_{ETI}] + E[t_{RQ}] \quad (1)$$

Where $E[.]$ denotes for the expected value. In the following subsections, the analysis is provided to determine the value of each term in (1).

1) Service time of the enable transmission interval

The time axis can be normalized in CRI units, consequently the service time can be considered as uniformly distributed random variable in the interval from 0 to 1. Since the arrival rate of MDs access requests is independent on the CRI timing, therefore the expected value of service time $E[t_{ETI}] = 0.5$.

2) Collision resolution delay

Let λ be the arrival rate of access requests to the system. The probability that the MD will find a free preamble in a given virtual group to access the system can be represented as,

$$p_f(\lambda) = \sum_{m=0}^{\infty} P(\text{free preamble} | m) P(m) \quad (2)$$

Where $P(\text{free preamble} | m)$ is the probability that an MD will choose free preamble in a given CRI and $P(m)$ is the probability that there are m MDs trying to access the system in that CRI. Thus,

$$p_f(\lambda) = \sum_{m=0}^{\infty} P(m) N_p \left(\frac{1}{N_p}\right) \left(1 - \frac{1}{N_p}\right)^{m-1} \quad (3)$$

Since the arrival rate of access requests is modeled with Poisson distribution for some M2M application [14] [15], equation (3) can be rewritten as follows,

$$p_f(\lambda) = \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} e^{-\lambda} \left(1 - \frac{1}{N_p}\right)^{m-1} = e^{-\left(\frac{\lambda}{N_p}\right)} / \left(1 - \frac{1}{N_p}\right) \quad (4)$$

Since the MDs have the same probability to successfully send their preamble, so the service time can be modelled as a discrete geometric random variable with probability distribution function as follows,

$$F_{RQ}(t) = p(T \leq t) \quad (5)$$

According to [11], the geometrical distribution can be approximated to the corresponding exponential distribution for a continuous service time. As a result, the system can be modelled as $M/M/N_g$ with a probability density function for the collision resolution system delay as follows,

$$f_{RQ}(t) = \frac{\partial F_{RQ}(t)}{\partial t} = \left(1 - p_f(\lambda)\right)^t \ln\left(\frac{1}{1 - p_f(\lambda)}\right) \quad (6)$$

Hence, the average service time is given by,

$$\mu_{RQ}^{-1} = 1 / \ln\left(1 - 1 / \left(e^{-\frac{\lambda}{N_p}} / \left(1 - \frac{1}{N_p}\right)\right)\right) \quad (7)$$

Where μ is the service rate that represents the rate of served MDs per CRI. Provided that the collision resolution system was modelled by Markov chain as shown in Fig. 5, the steady state probability p_n can be obtained using [16] as follows:

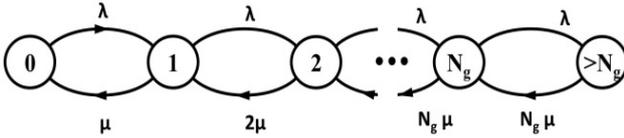


Figure 5. Discrete-time Markov chain for the $M/M/N_g$ system

$$p_n = \begin{cases} p_0 \frac{(N_g \rho)^n}{n!}, & n \leq N_g \\ p_0 \frac{N_g^{N_g} \rho^n}{N_g!}, & n > N_g \end{cases} \quad (8)$$

Where, $\rho = \frac{\lambda}{N_g \mu_{RQ}}$ is the utilization factor. Given the condition that, $\sum_{n=0}^{\infty} p_n = 1$, we can obtain the value of p_0 as follows,

$$p_0 = \left[\sum_{n=0}^{N_g-1} \frac{(N_g \rho)^n}{n!} + \sum_{n=N_g}^{\infty} \frac{N_g^{N_g} \rho^n}{N_g!} \right]^{-1} \quad (9)$$

As the queuing in the system starts once the preambles in all virtual groups are already used, therefore the queuing probability can be obtained as below,

$$P_{RQ} = p_0 \frac{(N_g \rho)^{N_g}}{N_g! (1-\rho)} = \frac{(N_g \rho)^{N_g} / N_g! (1-\rho)}{\sum_{n=0}^{N_g-1} \frac{(N_g \rho)^n}{n!} + \frac{(N_g \rho)^{N_g}}{N_g! (1-\rho)}} \quad (10)$$

The expected number of MDs waiting in the queue is defined as (N_Q), which can be obtained as follows,

$$N_Q = p_0 \frac{(N_g \rho)^{N_g}}{N_g!} \frac{\rho}{(1-\rho)^2} = P_{RQ} \frac{\rho}{1-\rho} \quad (11)$$

Using Little's theorem, the average time (W) that the MD has to wait in the queue can be expressed as follows,

$$W = \frac{N_Q}{\lambda} = P_{RQ} \frac{\rho}{\lambda(1-\rho)} \quad (12)$$

By substituting with the service rate in the utilization factor equation, we can obtain,

$$W = \frac{P_{RQ}}{N_g \ln\left(1 / \left(1 - e^{-\frac{\lambda}{N_p}} / \left(1 - \frac{1}{N_p}\right)\right)\right) (1-\rho)} \quad (13)$$

Using the calculated values of $E[t_{TA}]$, average service time, and the average waiting time, the expected value for the

total access delay of the system can be expressed as follows,

$$E[t_{TA}] = 0.5 + \frac{1}{\ln\left(1 / \left(1 - e^{-\frac{\lambda}{N_p}} / \left(1 - \frac{1}{N_p}\right)\right)\right)} + \frac{P_{RQ}}{N_g \ln\left(1 / \left(1 - e^{-\frac{\lambda}{N_p}} / \left(1 - \frac{1}{N_p}\right)\right)\right) (1-\rho)} \quad (14)$$

3.2. Access Success Probability

The 3GPP in [13] defines the access success probability in LTE as the probability of successfully completing the random access procedure within the maximum number of preamble retransmissions. We will refer to the maximum number of preamble retransmissions with N_{max} . Since the random access response window is 5 subframes as defined in [13], then each retrial takes 10ms (one frame length). Accordingly, the success probability can be defined as the probability of successfully completing the random access procedure within $10 * N_{max}$ ms, which is the equivalent time period for $N_{max} / 2$ CRIs. Hence, by substituting with t equal $N_{max} / 2$ in (6), we can obtain the access success probability $p_s(\lambda)$ as follow,

$$p_s(\lambda) = F_{RQ}(5) = 1 - (1 - p_f(\lambda))^{N_{max}/2} \quad (15)$$

By substitute with $p_f(\lambda)$ in (15), the access success probability can be obtained as,

$$p_s(\lambda) = 1 - \left[1 - \left[e^{-\frac{\lambda}{N_p}} / \left(1 - \frac{1}{N_p}\right) \right] \right]^{N_{max}/2} \quad (16)$$

4. Performance Evaluation and Comparison

The 3GPP recommended settings defined in [13] and depicted in Table 1 are used to evaluate the performance of the proposed DQAL protocol.

Table 1. The 3GPP suggested parameters for comparison

Parameter	Setting
Total number of preambles per RA slot	54
Number of dedicated preambles (contention free)	10
Maximum number of preamble retransmission	10
ra-ResponseWindowSize	5 ms
PRACH Configuration Index	0, 3, 6 and 9

In order to support highest possible arrival rate, we should select the best value for N_p and N_g that maintain the system stability (utilization factor to be less than one). Under fixed

capacity (fixed number of preambles reserved for MDs), there is one tuning parameter left (N_p). As the N_p value increases, the probability that MD will find free preamble in the virtual group will increase, and accordingly the total access delay will decrease as shown in Fig. 6. However, the choice of N_p is limited by the stability constraint. Figure 7 demonstrates the relation between the utilization factor and the arrival rate with different values for N_p . It can be observed that at $N_p = 6, 18,$ and 30 preambles, the system can support up to $500, 550,$ and 500 requests/sec respectively. This means that the maximum arrival rate is directly proportional to the N_p till certain value, and then the relation is inversed which is an anticipated partitioning phenomenon.

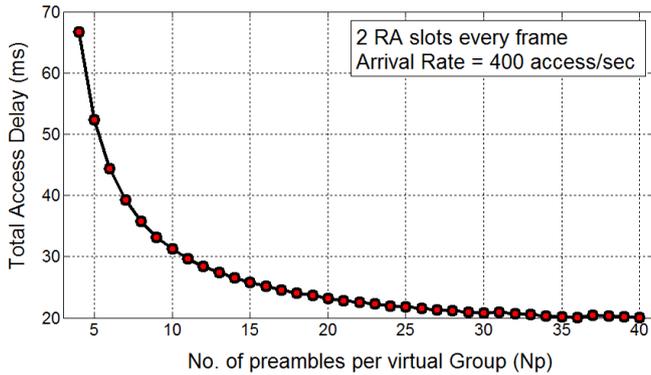


Figure 6. Average access delay with N_p

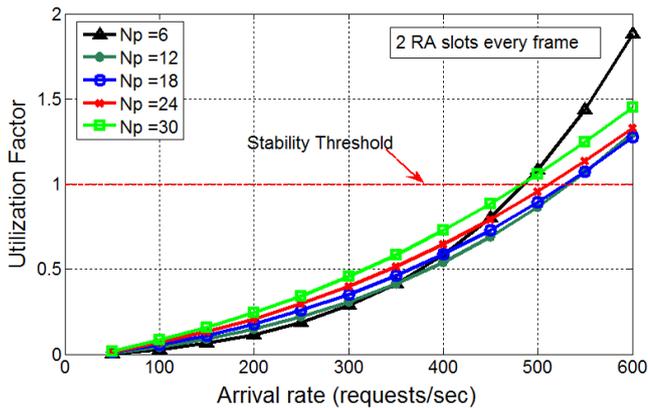


Figure 7. Utilization factor with different values for N_p

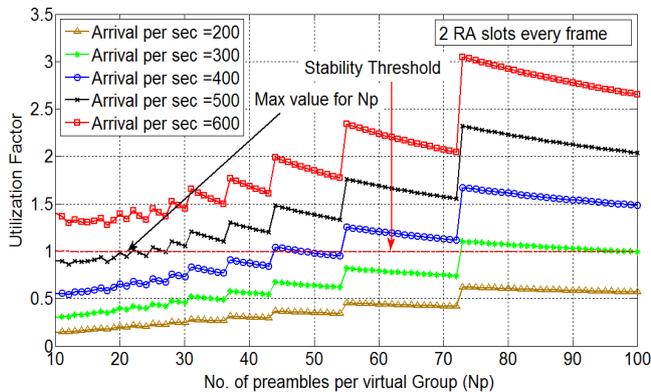


Figure 8. Utilization factor vs. N_p

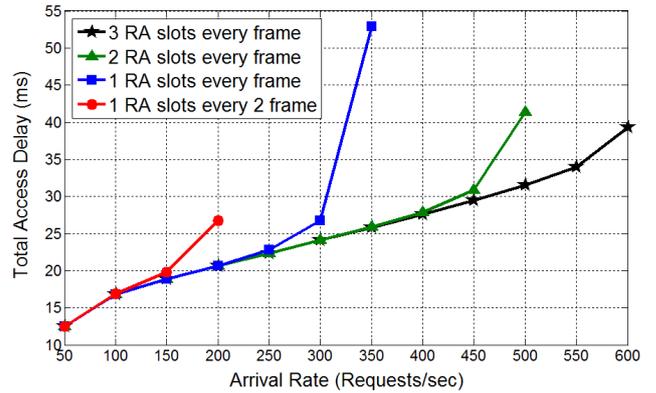


Figure 9. Average access delay for MD

Figure 8 shows the maximum value of N_p that allows the highest arrival under system stability constraint. The non-monotonic relation between the utilization factor and N_p can be explained due to the lossless distribution of the preambles on the virtual groups. For different PRACH Configuration Indices (0, 3, 6 and 9), the obtained maximum N_p values are 18, 22, 18, and 30. For a while, $N_p = 18$ is taken as optimum value where the other values lead to left unused preambles.

The average access delay was shown in Fig. 9 at different numbers for RA slots. As the number of RA slots increases, the system can support much more arrival rate with lower delay.

The 3GPP in [13] define the access success probability as the probability of completing the random access procedure within the maximum number of preamble retransmissions. This is equivalent to 100ms maximum access delay [13]. According to Fig. 9, the maximum obtained access delay for DQAL is less than 55ms which is less than 100ms. Hence; we can consider the access success probability equal one for DQAL. It is also important to note that the average access delay for DQAL is less than half of slotted ALOHA's delay at different values for RA slots as shown in Table 2.

Table 2. Results comparison for LTE and DQAL

No. of RA Slots	λ (Req/s)	Avg. Access Delay (ms)		Avg. Success Probability	
		LTE [2]	DQAL	LTE [2]	DQAL
1	350	119	53	0.78	1
2	500	102	42	0.98	1
3	600	98	39	0.9	1

5. Conclusions

In this paper we introduced a new protocol (DQAL) to solve the access issue of M2M communications over LTE. The protocol is designed in such a way to be used by the MDs while not affecting the normal access operation of H2H communication. The original DQRAP is adapted to fit with the frame structure and specifications of LTE. The optimum design parameters in terms of the number of virtual groups

and the number of preambles per each group are determined. The results show that the performance of the proposed DQAL protocol outperforms the original access protocol used by LTE in both access delay and success probability.

REFERENCES

- [1] P. Popovski, "Reliable Reporting for Massive M2M Communications With Periodic Resource Pooling," *IEEE Wirel. Commun. Lett.*, vol. 3, no. 4, pp. 429–432, 2014.
- [2] A. Laya, L. Alonso, and J. Alonso-Zarate, "Is the random access channel of LTE and LTE-A suitable for M2M communications? A survey of alternatives," *IEEE Commun. Surv. Tutorials*, vol. 16, no. 1, pp. 4–16, 2014.
- [3] J. P. Cheng, C. H. Lee, and T. M. Lin, "Prioritized Random Access with dynamic access barring for RAN overload in 3GPP LTE-A networks," *2011 IEEE GLOBECOM Work. GC Wkshps 2011*, pp. 368–372, 2011.
- [4] 3GPP TS 36.331 V10.5.0, "Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC)," Mar. 2012.
- [5] S.-Y. Lien, K.-C. Chen, and Y. Lin, "Toward ubiquitous massive accesses in 3GPP machine-to-machine communications," *IEEE Commun. Mag.*, vol. 49, no. 4, pp. 66–74, April 2011.
- [6] M.-Y. Cheng, G.-Y. Lin, H.-Y. Wei, and A.-C. Hsu, "Overload control for Machine-Type-Communications in LTE-Advanced system," *IEEE Commun. Mag.*, vol. 50, no. 6, pp. 38–45, June 2012.
- [7] 3GPP TSG RAN WG2 #71 R2-104662, "MTC simulation results with specific solutions," ZTE, Madrid, Spain, 23rd Aug. 2010.
- [8] 3GPP TR 37.868 V11.0.0, "Study on RAN Improvements for Machine-Type Communications," Sep. 2011.
- [9] Pratas, Nuno K., et al. "Code-expanded random access for machine-type communications." 2012 IEEE Globecom Workshops. IEEE, 2012.
- [10] W. Xu and G. Campbell, "A Near Perfect Stable Random Access Protocol for a Broadcast Channel," in *IEEE Proc. ICC92*, vol. 1, 1992, p. 370374.
- [11] L. Alonso and O. Sallent, "A Near-Optimum MAC Protocol Based on the Distributed Queuing Random Access Protocol (DQRAP) for a CDMA Mobile," in *IEEE J. Sel. Areas Commun.*, vol. 18, no. 9, Sep. 2000, pp. 1701–1718.
- [12] L. Alonso, R. Ferrus, and R. Agusti, "WLAN throughput improvement via distributed queuing MAC," *IEEE Commun. Lett.*, vol. 9, no. 4, pp. 310–312, 2005.
- [13] 3GPP TSG RAN WG2 #71 R2-104663, "[100bis#11] LTE: MTC LTE simulations," ZTE, Madrid, Spain, 23rd Aug. 2010.
- [14] Zeng, Qi, Husheng Li, and Daiyuan Peng. "Frequency-hopping based communication network with multi-level QoS in smart grid: code design and performance analysis." *Smart Grid, IEEE Transactions on* 3.4 (2012): 1841-1852.
- [15] Cao, Yang, et al. "QoS-oriented wireless routing for smart meter data collection: Stochastic learning on graph." *Wireless Communications, IEEE Transactions on* 13.8 (2014): 4470-4482.
- [16] D. Bertsekas and R. Gallager, *DATA NETWORKS*. Englewood Cliffs, NJ: Prentice-Hall International, 1987.