Performance Enhancement of MTC in LTE Networks by Maximizing Random Access Procedure Throughput

Ibraheem M. Fayed
Eman S. El-Din

Network Planning Department, National Telecommunication Institute (NTI)
Cairo, Egypt.

Email: ibrahim.fayed@nti.sci.eg
Email: emanserag@gmail.com

Abstract—**A fundamental requirement for any cellular system is the possibility for the device to request a connection setup, commonly referred to as random access procedure. In Long Term Evolution (LTE) networks, the distribution of a limited number of radio resources among Human-to-Human (H2H) users and increasing number of Machine-Type-Communication (MTC) devices in Machine-to-Machine (M2M) communications is one of the main problems. An analytical model is conducted to compute the throughput for message 1 and message 2 using a Markov chain model for the four messages signaling flow with buffering for message 4 in LTE Third-Generation Partnership Project (3GPP) random access. The network performance will be enhanced by determining a dedicated arrival rate corresponding to maximum throughput of message 2 that will assist the network planner to optimize the network performance.**

Keywords-Machine Type Communication (MTC) ; Machine to Machine (M2M); LTE network; Random Access Procedure throughput.

I. INTRODUCTION

In the last decades, mobile communication has evolved from being an expensive technology for a few selected individuals to today's ubiquitous systems used by a majority of the world's population [1]. Modern wireless communication schemes like 3GPP LTE network, do not allow serving effectively M2M connections between a huge numbers of interacting MTC devices. A good background of this issue "when a huge number of machines request access" was explained in [2]. Increasing the network performance is not possible due to the finance approach. One possible solution of the problem is based on the use of random access procedure (RACH) procedure [3].

Increasing network capacity is the major objectives of LTE network. This increasing will provide high data rates for end users with low latency, ensure high Quality of Service (QoS), and reduce the cost. For channel access and ensuring the Quality of Service (QoS), there are two attributes of random access procedures in LTE: contention-free and contention-based random access.

In contention-free random access, the e-NodeB signals a reserved preamble for the mobile station, thus avoiding the contention. In contention-based random access, all mobile stations need to participate in contention for the resources [4].

A very huge number of machines, such as user equipment (UE) may dwell in coverage area of cell, which may request to access the network periodically or sporadically. They also have a small power that must be used as efficiently as possible. The 3GPP has carried in different studies [5] [6] that try to address the issues related to M2M communications in the present systems, as well as in the future releases of LTE. A detailed study and analysis of physical downlink control channel (PDCCH) performance for M2M traffic in LTE was introduced in [7].

The main aim of this paper is to obtain the upper band of the request arrival rate in order to sustain the network performance metric such as the throughput for MTC in LTE network. An explicit analysis for evaluation of throughput for message 1 and message 2 using a Markov chain model for the four messages signaling flow with buffering for message 4 in LTE 3GPP random access is presented.

This paper is organized as follows: random access procedure using four messages in LTE network is explained in Section 2. Markov chain model for transient states for the four messages with queuing buffer for message 4 and model assumptions and analysis is introduced in Section 3. In Section 4, results and verification are shown. Section 5 is devoted to conclusion.

II. RANDOM ACCESS PROCEDURE

A brief description of the contention-based random access procedure utilized, for example, by the MTC traffic, is explained as follow.

Step 1: The MTC device (UE) requests to initiate the RACH procedure by selecting one of the available RACH preambles randomly and then sending the preamble in Message 1 over the physical random access channel (PRACH) in the uplink. A collision occurs when two or more machines (UEs) get the same preamble in the same subframe. However, after preamble sending the UE waits for a random access response (RAR) (Message 2) from the e-NodeB within the time interval called a response window [3], i.e., even if two or more UEs use the same preamble for Message 1 and a collision occurs, the e- NodeB will detect this event and will not send reply to UE. The transmission of a random access preamble is restricted to certain subframes. Let b denote their periodicity, i.e., random access is possible in every b^{th} subframe. In addition, let K denote the total number of available preambles.

Step 2: e-NodeB replies with Message 2, which is also known as the RAR and it includes UL grant for Step 3. Message 2 is sent over the physical downlink shared channel (PDSCH). So, schedule the user is needed [7], i.e., send a downlink assignment control message over the PDCCH. There may be at most one RAR message in each subframe, but each may have more than one UL grant (each referring to a different preamble). Let c is defined as the maximum number of UL grants per RAR per subframe. Note that in this model, an UL grant is given for every uncollided preamble.

Step 3: In the case of successful preamble transmission after receiving Message 2 from the e-NodeB and RAR processing time, the UE sends a RRC connection request (Message 3) to the e-NodeB over Physical Uplink Shared Channel (PUSCH).

Step 4: RACH procedure is completed after the UE receiving a contention resolution message (Message 4) from the e-NodeB. Hybrid Automatic Repeat request (HARQ) procedure guarantees a successful transmission of Message 3/ Message 4. HARQ procedure provides a limit in Message 3/ Message 4 sequential transmission attempts. If the limit is reached UE should start a new RACH procedure by sending a preamble.

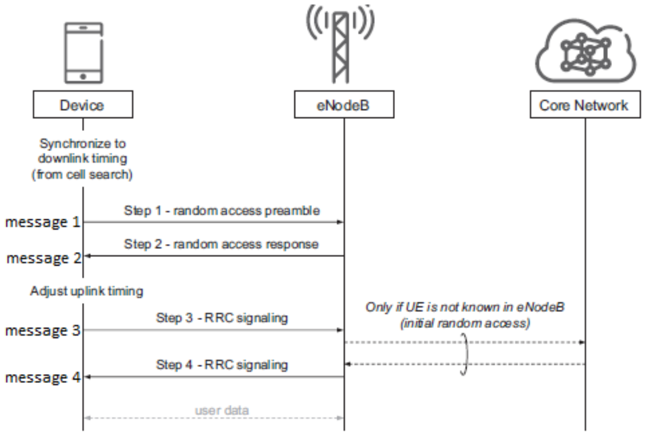

Figure 1. Four messages control signaling flow sequence in LTE random access. [1]

Let N be defined as the PDCCH resource size (in control channel elements CCEs), N^{Msg2} and N^{Msg4} are the number of CCEs used to send a Message 2 and a Message 4, respectively. So, a maximum of $M = N/N^{Msg4}$ Message 4's can be sent in one subframe if Message 2 is not present in that subframe [7]. When a Message 2 is sent in a subframe, then at most $m = (N - N^{Msg2})/N^{Msg4}$ Message 4's can be sent in that subframe. Although the parameters N, N^{Msg2}, and N^{Msg4} are closely connected from the system point of view, our model is greatly simplified when we use the derived parameters M and m. This messaging scheme is demonstrated in Figure 1.

III. MODEL PARAMETERS AND ASSUMPTIONS

In LTE there are K parallel Aloha channels used, and each time a UE makes a request of retransmission, the preamble is selected in random. By this random selection, the fresh random access requests and the retransmission attempts will be mixed. In addition, the input parameter of the model will be a, which is the aggregate rate of all requests (arrival of new requests and retransmission requests) not λ, which is the rate of fresh requests per sub-frame.

Let θ be defined as the throughput of successful requests. When the system is stable, it can be determined from the model as a function of a. The system is stable if the average input rate is the same as the average output rate, which means that the arrival rate λ of fresh requests is equal to the throughput θ of successful requests whenever the system is stable. This is how we get the functional relationship between the aggregate request rate a and the arrival rate λ of fresh requests.

Now, let A_{nk} represent the total number of random access requests with preamble k (including both the new requests and the retransmissions) in time slot n. Since the aggregate stream of requests (including the fresh ones and the retransmissions) is assumed to obey a Poisson process and the preambles are independently chosen from the uniform distribution, A_{nk} are independent and identically distributed random variables obeying a Poisson distribution with mean ab/K and point probabilities, i.e.,

$$P_i(a) = P_r\{A_{nk}=i\}=\frac{\left(\frac{ab}{k}\right)^i}{i!}e^{-ab/K} \qquad (1)$$

The probability of failure can be calculated as follow:

P_r\{failure\} = P_r\{collision occurs in Step 1\} +P_r\{no collision occurs in Step 1, loss occurs in Step 2\} +P_r\{no failure occurs in Steps 1 and 2, delay occurs in Step 4\}.

A call flow for successful and unsuccessful session setup establishment based on RACH procedure is shown in Figure 2.

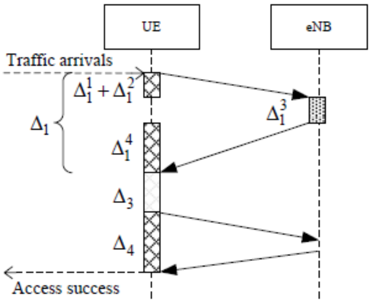

Figure 2. A request access success without collision

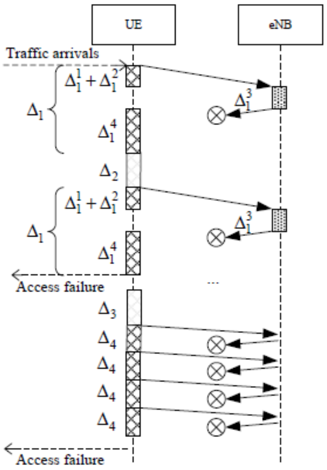

Figure 2b. request access failure

Figure 2. Four messages control signaling flow sequence to transmit data with retransmissions in LTE random access [3]

A mathematical model is proposed in the form of discrete Markov chain that follows the steps of RACH procedure and describes the evolution of Message 4 buffer and determines the number of preamble attempt collisions and the number of sequential Message 3/ Message 4 transmission attempts. With this model the access delay for each state of the Markov chain can be calculated by summing up the corresponding time intervals introduced as follow:

- Δ_{11}, Δ_{21}, Δ_{31}, and Δ_{41} are defined respectively as: waiting time for a RACH opportunity to transmit a preamble, preamble transmission time, preamble processing time at the e-NodeB, and RAR response window. Then $\Delta_1 = \Delta_{11} + \Delta_{21} + \Delta_{31} + \Delta_{41}$ is the time interval from the beginning of RACH procedure until sending message 3 or resending a preamble.

The back off window is defined as Δ_2 , the RAR processing time as Δ_3, and time for Message 3 transmission, waiting for Message 4, and Message 4 processing is Δ_4.

Let us consider inhomogeneous discrete Markov chain {ξ} over the state space

Ł={(0), (1), (2), (n,m), 0≤n≤N, 0≤m≤M) which determines the process of transitions between states. We assume that state (0) is the start point of RACH procedure, state (1) is the absorbing state denoting access success, state (2) is another absorbing state denoting access failure, the pair of (n,m) denotes the state when n Message 1 retransmissions and m Message 3/ Message 4 retransmissions occurred. The maximum number of retransmission for message 1equals (N-1) and the maximum number of retransmission of message (2) (M-1). The values of (N,M) will be taken from [9].

The start of transmission or retransmission of message 3 must be done after success in sending message 1. The two events are independent.

Let *g* denote the HARQ retransmission probability for Message 3/ Message 4. And *p* is the probability of retransmission failure for message 1 expressed in equation (1).

The probability of success *P* in state (n, m) will be:

$$P(n,m) = (1-p)(1-g)p^n g^m \tag{2}$$

P(n,m) will be shown and calculated as a future work.

Now, define

$$Y_n(1) := \#\{k : A_{nk} = 1, k = 1, \ldots, K\} \tag{3}$$

$$\hat{Y}_n(1) := \#\{k : A_{nk} \geq 1, k = 1, \ldots, K\} \tag{4}$$

where $\hat{Y}_n$ (1) and Y_n (1) are referring to the total number of preambles chosen in time slot n, and the number of successful (uncolliding) Message 1's respectively. The joint distribution of the random variables Y_n (1) and $\hat{Y}_n$ (1) is calculated as follows:

$$q_{ij}^{(1)}(a) := \Pr\{Y_n^{(1)} = i, Y_n^{\sim(2)} = j\} \tag{5}$$

$$= \binom{K}{i\ j-i} p_o^{K-J} p_1^i (1 - p_o - p_1)^{j-i}$$

$$0 \leq i \leq j \leq K$$

There is at most *b* Message 2's and so at most *bc* UL grants per time slot. Message 2's in time slot *n* are generated by Message 1's of the previous time slot. Let $\hat{Y}_n$ (2) and Y_n (2) denote the total number of UL grants included in Message 2 in time slot n and the number of successful (uncolliding) UL grants respectively. No losses appear in this step, i.e., $\hat{Y}_n$ (2) = $\hat{Y}_{n-1}$ (1) and Y_n (2) = Y_{n-1} (1), if the total number of preambles chosen in the previous time slot is sufficiently small, i.e., $\hat{Y}_{n-1}$ (1) ≤ *bc*, which is trivially true if $K \leq bc$. However, if $\hat{Y}_{n-1}$ (1)> *bc*, then losses occur so that $\hat{Y}_n$ (2)= *bc*. Let the preambles that are given a UL grant in the latter case are chosen in random by e-NodeB. Thus, we have (for the nontrivial case K >*bc*

$$q_{ij}^{(2)}(a) := Pr\{Y_n^{(2)} = i, \hat{Y}_n = j\} \tag{6}$$

$$\left\{\begin{array}{ll} q_{ij}^{(1)}(a) & 0 \leq i \leq j \leq bc \\ \displaystyle\sum_{k=bc}^{K} \sum_{l=i}^{k} q_{lk}^{(1)}(a) \frac{\binom{l}{i}\binom{k-l}{bc-i}}{\binom{k}{bc}}, & 0 \leq i \leq j = bc \end{array}\right\}$$

With the following marginal distributions:

$$q_i^{(2)}(a) := \Pr\{Y_n^{(2)} = i\} = \sum_{j=i}^{bc} q_{ij}^{(2)}(a) \tag{7}$$

$$0 \leq i \leq bc$$

$$\tilde{q}_j^{(2)}(a) := \Pr\{\tilde{Y}_n^{(2)} = j\} = \sum_{i=0}^{j} q_{ij}^{(2)}(a)$$

$$0 \leq j \leq bc$$

Utilizing the definition of $q_{ij}^{(2)}$ (*a*), we find that

$$\tilde{q}_j^{(2)}(a) = \begin{cases} \binom{K}{j} p_0(a)^{K-j} \left(1 - p_0(a)\right)^j, & 0 \leq j < bc \\ \displaystyle\sum_{l=bc}^{K} \binom{k}{l} p_0(a)^{K-l} \left(1 - p_0(a)\right)^l, & j = bc \end{cases} \tag{8}$$

So, $\tilde{Y}_n^{(2)} = min\{B(a), bc\}$

where *B (a)* is a binomially distributed random variable with parameters *K* and $1 - p_0(a)$.

For throughput analysis, the throughput *θ* of successful requests must be equal to the arrival rate λ of fresh requests.

This proposal finds conditions for stability in terms of the total traffic *a* and then to determine the throughput of successful requests *θ (a)* as a function of *a*, as well as the maximum throughput $\theta^* = max_a\ \theta(a)$. To simplify the notation, we assume here that *K* >*bc*. The generalization to the case $K \leq bc$ is straightforward.

By using the Slotted Aloha model in [8] for the RACH used in Step 1the throughput (per subframe) of successful Message 1's as a function of *a*, which is the arrival rate of all random access requests per subframe, can be calculated as follows:

$$\theta^{(1)}(a) = \frac{E\left[Y_n^{(1)}\right]}{b} = ae^{-\frac{ab}{K}} \tag{9}$$

To calculate the throughput in Step 2, Since *K* >*bc*, the throughput is then reduced by the limited number of UL grants in Message 2. The throughput (per subframe) of successful UL grants as a function of *a* is distinctly

$$\theta^{(2)}(a) = \frac{E\left[Y_n^{(2)}\right]}{b} = \frac{1}{b}\sum_{i=1}^{bc} i q_i^{(2)}(a) \tag{10}$$

Equations (9) and (10) can be solved using MATLAB software to show the throughput of message 1 and throughput of message 2

ISBN: 978-1-61208-572-2

IV. RESULTS ANALYSIS AND VERIFICATION

A typical values for the model parameters are taken from [7] [9] and summarized in Table 1.

TABLE I TYPICAL VALUES FOR THE MODEL PARAMETERS

Notion	Definition	Typical Values
K	Number of preambles	54
B	RACH Periodicity	5
C	Maximum number of UL grants per subframe	3

Using the above values in equations (1:10) and solving using MATLAB program which explained by the flowing flowchart in Figure 3, we get the results shown in Figures 4-6.

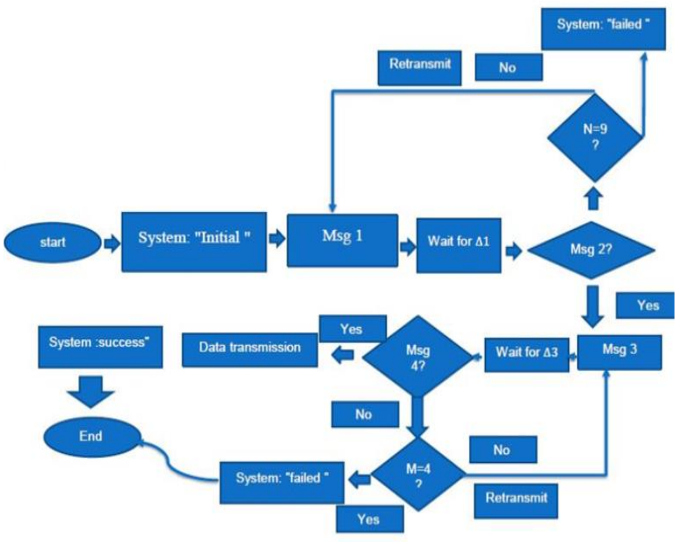

Figure 3. Flow chart explains the steps used in MATLAP program

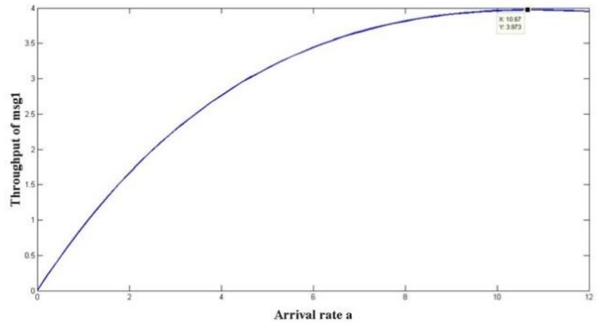

Figure 4. The throughput of message 1 with the increasing the aggregate arrival rate a

As shown in Figure 4 as the arrival rate for requests increases the throughput of message 1 increases till a certain value ($\theta^{(1)}(a) = 3.973$) at arrival rate a =10.67 $requests/ms$. Then the throughput approximately sets at this value because of the UE will know after a certain window time that the collision is occurred and retransmits message 1 again directly with new preamble in the uplink.

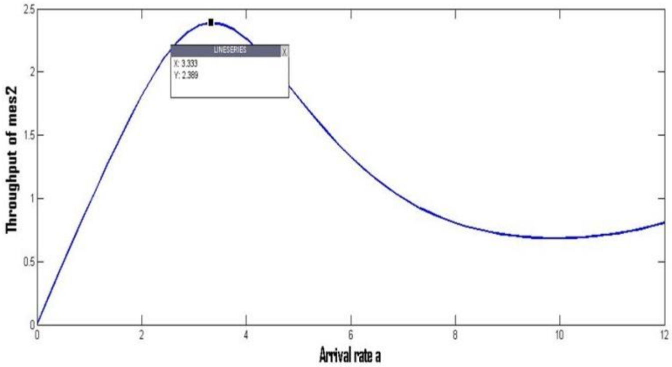

Figure 5. The throughput of message 2 with the increasing the aggregate arrival rate *a*

As shown in Figure 5, as the arrival rate for requests increases the throughput of message 1 increases till a certain value ($\theta^{(2)}(a) = 2.377$) at arrival rate a =3.333 $requests/ms$ and then throughput decreases because the message 2 contains information for more than one user up to 8 users and in case of collision lose downlink for message 2 the all users in this control channel will retransmit (requests) message 1 again directly and may be have the same preamble and collision in message 1 dose again.

The failure in message 2 affects directly in message 1 so to get the maximum throughput of message 1 and message 2, the intersection between throughput curve of message 2 and message 1 (which is the maximum throughput for message 2) will be considered as shown in Figure 6.

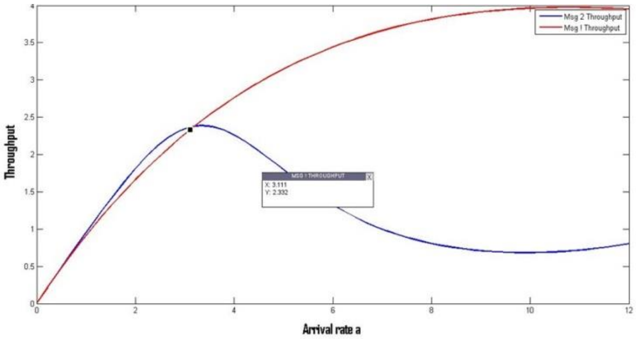

Figure 6. The intersection between throughput of message 1and throughput of message 2 with the increasing the aggregate arrival rate a

As shown in Figure 6, the maximum throughput is equal 2.332 at arrival rate a equals a =3.333 $requests/ms$ and this results is agree with [7].

V. CONCLUSION

The main aim of this paper is to obtain the upper band of the arrival request rate in order to sustain the network performance metric such as the throughput for MTC in LTE network. Explicit analysis for evaluation of throughput for message 1 and message 2 using a Markov chain model for the four messages signaling flow with buffering for message 4 in LTE 3GPP random access was presented.

In LTE planning, it is important to take into consideration the requests arrival rate (as forecasting plan) and the planner must select the network parameters (such as the base station locations, antennas height, etc.) to achieve the maximum throughput. So in this paper, it was found that, the planner has to design the system with arrival rate less than $3.333\ requests/ms$ to maximize network throughput. If the arrival rate increases above this value, the collisions in message 1 will increase, causing a degradation of the network throughput and also a decrease of the network performance.

The proposed analytical model will be assigned as a future work to calculate the state probability by using the Markov chain model explained in the proposed model.

REFERENCES

[1] E. Dahlman, S. Parkvall, and J. Skold, 4G, LTE-Advanced Pro and The Road to 5G. 3rd Edition, London, United Kingdom, Academic Press, Elsevier, 2016.

[2] 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, Apr. 2011.

[3] V. Y. Borodakiy, K. E. Samouylov, Y. V. Gaidamaka, P. O. Abaev, I. A. Buturlin, and S. A. Etezov, " Modelling a Random Access Channel with collisions for M2M Traffic in LTE Networks ", The reported study was partially supported by the RFBR, research projects No. 13-07-00953, 14-07-00090, 2014.

[4] A. N. Khan, J. Khalid, H. K. Qureshi, " Performance Analysis of Contention-Based Random Access Procedure in Clustered LTE Networks ", Seventh International Conference on Next Generation Mobile Apps, Services and Technologies,pp 203-209, 2013.

[5] 3GPP R2 112863, “Backoff enhancements for RAN overload control,” ZTE, Barcelona, Spain, May 2011.

[6] 3GPP TR 23.888 V1.6.0, Third-Generation Partnership Project; Technical Specification Group Services and System Aspects; System Improvements for Machine-Type Communications (MTC), Sophia-Antipolis, France, Nov. 2011.

[7] P. Osti, P. Lassila, S. Aalto, A. Larmo, and T. Tirronen, "Analysis of PDCCH Performance for M2M Traffic in LTE", IEEE Transactions on Vehicular Technology, Vol. 63, No. 9, pp. 4357– 4371, November 2014.

[8] D. Bertsekas and R. Gallager, Data Networks, 2nd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 1992.

[9] 3GPP TR 37.868 – Study on RAN Improvements for Machine-type Communications. (Re-lease 11). 2011.