Throughput Comparison of Random Access Methods for M2M Service over LTE Networks

Ki-Dong Lee, Sang Kim, and Byung Yi
LG Electronics Mobile Research USA
San Diego, CA, 92131 United States
Email: kidong.lee@lge.com

***Abstract*—It is one of the key issues in the 4-th generation (4G) cellular networks how to efficiently handle the heavy random access (RA) load caused by newly accommodating the huge population of Machine-to-Machine or Machine-Type Communication (M2M or MTC) customers/devices. We consider two major candidate methods for RA preamble allocation and management, which are under consideration for possible adoption in Long Term Evolution (LTE)-Advanced. One method, Method 1, is to completely split the set of available RA preambles into two disjoint subsets: one is for human-to-human (H2H) customers and the other for M2M customers/devices. The other method, Method 2, is also to split the set into two subsets: one is for H2H customers only whereas the other is for both H2H and M2M customers. We model and analyze the throughput performance of two methods. Our results demonstrate that there is a boundary of RA load below which Method 2 performs slightly better than Method 1 but above which Method 2 degrades throughput to a large extent. Our modeling and analysis can be utilized as a guideline to design the RA preamble resource management method.**

***Index Terms*—Machine-to-Machine (M2M), Long Term Evolution (LTE), Throughput.**

I. Introduction

Recently, the applications of machine-to-machine (M2M) communication [1] has drawn attention for its promising market opportunity ranging from ubiquitous healthcare service [2] to smart grid (www.oe.energy.gov/smartgrid.htm). Among those fora are 3GPP, ETSI M2M, IEEE 802.16p, and some minor groups. In cellular networks, “M2M” node refers to a communication node that has no or minimal interaction with human [3]. One of the main characteristics of M2M communication is that the fixed cost for communication is high: that is to say, the data that M2M nodes are transmitting every time is mostly small in size but the frequency of their making data connections is higher than ordinary human-type nodes due to their specific roles and functions: e.g. frequently reported small-size measurement data.

Due to this reason, one of the most important design considerations in M2M service networks is how to design an efficient method for handling the ‘random access’ (RA) load [4], [5], [6], [7] generated by a possibly huge population of M2M nodes attached to the network. In the whole process of RA, a UE shall go through a preliminary stage, called Access Barring Check (ABC) as depicted in Fig. 1, where the UE should check whether the access to the network is prevented or not, based on the received System Information (SI) and its random-number generation. Once the UE passes through the ABC procedure, it can start transmitting a preamble to the evolved NodeB (eNB), which will be the actual load to the random access channel (RACH). In the main leading standard body on the radio protocol of cellular networks, 3GPP, it is still on the table how to design the “access barring” mechanism specifically for M2M customers, called extended access baring (EAB). So far as agreed in the Radio Access Network (RAN) Working Group (WG), the features of EAB include Access Class (AC) based solution and “probability-based barring” solution, i.e., similar to the existing Access Class Barring (ACB) in LTE protocol specifications [6].

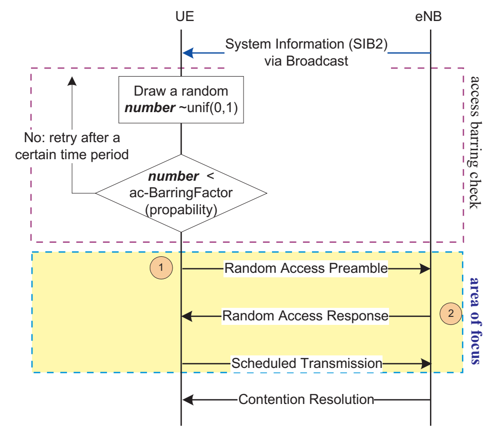

Fig. 1. The holistic procedure of Random Access and the Area of Focus in this work: The UE’s passing through Access Barring Check (ABC) will start transmitting a preamble, and the contention-based transmissions from multiple UE’s will generate RA load in Random Access Channel (RACH).

In this paper, we consider two most probable candidate methods for RA preamble allocation and management in the 3GPP LTE-Advanced network protocol (under development) [5]. One method, called “Method 1,” is to completely split the set of available RA preambles into two disjoint subsets: one is for human-to-human (H2H) customers and the other for M2M customers/devices [8]. The other method, called “Method 2,” is also to split the set into two subsets but one of the two sets are shared by H2H and M2M customers: namely, one is for H2H customers only whereas the other

The authors are with LG Electronics Mobile Research, San Diego, CA, 92131, United States (e-mail: kidong.lee@lge.com, iLoveLTE@gmail.com).
Most of this work is based on the first author’s work [4].

is for both H2H and M2M customers [9]. Method 2 can also provide a comparable throughput performance for M2M customers even if the preamble resources are shared with H2H customers. One of the most common feasible solutions is to allocate a properly chosen number of preambles to the preamble set shared by both H2H and M2M customers. In most cases, this number in Method 2 (for example, see x in Fig. 3) might be greater than the number of preambles in the set for M2M customers in Method 1 (for example, see m in Fig. 2). We analyze the throughput performance for the two methods through mathematical modeling. Our modeling and analysis has practical values in actual protocol design since it is simple and useful. Although Method 2 seemed to be capable of providing so-called *trunk efficiency*, our results demonstrated that the performance gain in throughput may depend on load situation. Therefore, by applying an adaptive method for choosing most appropriate method under the given conditions over time, it is possible to optimize or to improve the throughput performance in heavily loaded RACH.

The remainder of this paper is organized as follows. In Section II, we describe the random access model in LTE-Advanced networks (under development in 3GPP) and two methods under consideration. In Section III, we mathematically formulate the throughput based on some traffic parameters such as RA load, in other words, the arrival rate of RA attempts, and so on. Following the formulation, we compare the throughput performance of H2H RA attempts when the M2M RA throughput in both methods are tuned to be comparable to each other in Section IV. Following various numerical examples in Section V, we conclude our work in Section. VI.

II. Random Access Model

A. An Overview on Random Access Procedure

A user equipment (UE) should first perform Random Access (RA) procedure to obtain a new connection[1] to the network in a 3GPP case (e.g., Long Term Evolution, or LTE in short [5]) and this is a common procedure in most cellular systems. In 3GPP cellular systems, there are contention-based RA and non-contention-based RA. In this paper, we specifically consider contention-based RA and study throughput of RA channel (RACH) for two methods which will be described later.

An evolved NodeB (eNB), namely base station, broadcasts 'System Information', such as Master Information Block (MIB) and System Information Block Type k (SIBk, $k = 1, 2, \cdots$). SIB2 provides information for UE's how they can perform the RA procedure, where barring instruction is included: 'barring instruction' has a value *ac-BarringFactor* indicating the probability that an arbitrary UE is supposed to be prevented from attempting RA to the cell:

1) an ordinary UE[2] shall draw a random number from a uniform distribution $(0, 1)$;

[1]Data connection or signaling connection to a new node or either of such connections to the same node through different frequency channel resource

[2]There are several cases in Access Barring Check procedure in LTE but in this paper we consider the case where UE's are making an ordinary call, i.e., mobile-originating data call.

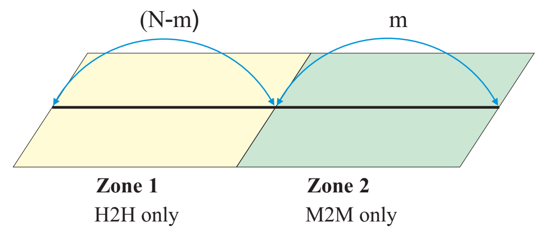

Fig. 2. Method 1: Completely disjoint separation of preambles [TR 37.868]. N: the number of preambles allocated to contention-based Random Access (RA). m: the number of preambles (out of the N preambles) dedicatedly allocated to M2M. $N-m$: the number of preambles (out of the N preambles) dedicatedly allocated to H2H.

2) if the random number drawn is smaller than *ac-BarringFactor*[3], then this UE is allowed to perform RA; otherwise, the UE is not allowed.

Fig. 1 presents the holistic view of RA procedure and the specific area of focus in this work. The traffic parameters, such as arrival rate of RA attempts, under consideration in this paper are observed after the Access Barring Check procedure. For example, the arrival rate of RA attempts is not the rate of original RA attempts but the rate of RA attempts passed through the Access Barring Check procedure. If the rate of original RA attempts is 100 with an *ac-BarringFactor* of 0.50, then the average observed rate of RA attempts we consider in this paper is around 50.

B. Detailed Description of RACH Opportunity

A RACH opportunity is defined as a 3-dimensional radio resource block on a time-frequency domain with a fixed number of preambles. Random Access Radio Network Temporary Identity (RA-RNTI) defined in the time-frequency domain is a time-frequency block. We denote by R the number of RA-RNTI's per unit time[4]. The number of available preambles for UE's is indicated in SIB2. We denote by N the number of available preambles. Then the total number of available RACH opportunities is $R \cdot N$. We next consider two methods for RA accommodating two types of UE's in a slightly different manner, respectively.

C. Two Methods under Consideration

We now describe two methods under consideration. The system accommodates human-type UE's (or H2H) and machine-type UE's (or M2M)[5]. The two methods are defined by the way of using preambles to these two types of UE's.

- Method 1: The set of all available preambles are separated into two subsets, where one set of preambles (e.g., Zone 1 in Fig. 2) are dedicatedly used by H2H UE's whereas the other set of preambles (e.g., Zone 2 in the same figure) are by M2M UE's. This method is what is currently described in 3GPP TR37.868 [3].
- Method 2: The set of all available preambles are separated into two subsets, where one set of preambles (e.g., Zone 1

[3]More specifically, it is the threshold value that ac-barringFactor indicates.

[4]Without loss of generality, we can assume that $R = 1$ by scaling the time-frequency domain in this paper.

[5]In this paper, MTC and M2M are used interchangeably.

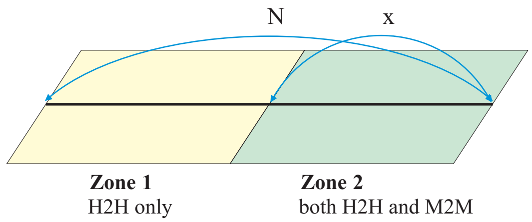

Fig. 3. Method 2: Separation of preambles with partial overlap [LGE R2-113343]. x: the number of preambles (out of the N preambles) shared by both M2M and H2H. $N-x$: the number of preambles (out of the N preambles) dedicatedly used by H2H. If x is properly chosen (i.e., a certain value greater than m), then M2M customers under Method 2 can experience a comparable throughput performance as those under Method 1. The chose of such x is not necessarily feasible at all times.

in Fig. 3) are dedicatedly used by H2H UE's whereas the other set of preambles (e.g., Zone 2 in the same figure) are shared by both H2H and M2M UE's.

III. Throughput Analysis

TABLE I
List of Symbols.

Parameter	Value
λ_k	arrival rate of RA requests of class k (including backoffs)
T_{ij}	RACH throughput of zone j in Method i
T_i	RACH throughput in Method i $(=\sum_j T_{ij})$
R	no. of available RA-RNTI's per unit time

(Class 1: H2H; Class 2: M2M)

For the arrival rate of RA attempts γ and the number of RACH opportunities S, the RACH throughput, denoted by T, can be modeled as:

$$\begin{aligned} T &= \text{(no. RACH oppt.)} \cdot \text{(per-opportunity thrput.)} \\ &= S \cdot \Big\{ \frac{\gamma}{S} \cdot \exp\big(-\frac{\gamma}{S}\big) \Big\} \\ &= \gamma \cdot \exp\big(-\frac{\gamma}{S}\big) \Big\} \end{aligned} \tag{1}$$

under the assumption that the arrival process is Poisson [7]. Thus, for Method 1, the RACH throughputs for H2H and M2M are given by

$$T_{11} = \lambda_1 \cdot \exp\Big\{ -\frac{\lambda_1}{(N-m)} \Big\}, \tag{2}$$

$$T_{12} = \lambda_2 \cdot \exp\Big\{ -\frac{\lambda_2}{m} \Big\}, \tag{3}$$

respectively. Therefore, the RACH throughput in Method 1 is given by

$$\begin{aligned} T_1 &= T_{11} + T_{12} \\ &= \lambda_1 \cdot \exp\Big\{ -\frac{\lambda_1}{(N-m)} \Big\} + \lambda_2 \cdot \exp\Big\{ -\frac{\lambda_2}{m} \Big\}. \end{aligned} \tag{4}$$

In Method 2, not all the RA attempts from class 1 (i.e., H2H) are supposed to use a preamble belonging to Zone 1. Only $\frac{(N-x)}{N} \cdot 100$ (%) of the attempts are to arrive at Zone 1 whereas the others are to arrive at Zone 2 with the RA attempts from class 2 (i.e., M2M). Thus we have the arrival rates of RA attempts at Zones 1 and 2, respectively

$$\gamma_1 = \lambda_1 \cdot \frac{(N-x)}{N} \tag{5}$$

$$\gamma_2 = \lambda_1 \cdot \frac{x}{N} + \lambda_2. \tag{6}$$

Thus, we have

$$T_{21} = \gamma_1 \cdot \exp\Big\{ -\frac{\gamma_1}{(N-x)} \Big\}, \tag{7}$$

$$T_{22} = \gamma_2 \cdot \exp\Big\{ -\frac{\gamma_2}{x} \Big\}. \tag{8}$$

The M2M portion out of the RACH throughput T_{22} is only

$$T_{22}^{\text{M2M}} = \lambda_2 \cdot \exp\Big\{ -\frac{\gamma_2}{x} \Big\}, \tag{9}$$

whereas the H2H portion is the rest, namely,

$$\begin{aligned} T_{22}^{\text{H2H}} &= (\gamma_2 - \lambda_2) \cdot \exp\Big\{ -\frac{\gamma_2}{x} \Big\} \\ &= \lambda_1 \cdot \frac{x}{N} \cdot \exp\Big\{ -\frac{\gamma_2}{x} \Big\}. \end{aligned} \tag{10}$$

IV. Throughput Comparison

For any given number of preambles assigned to Zone 2 in Method 1, namely for a given m, we check how many preambles are needed for Zone 2 of Method 2 in order to provide the same RACH throughput for M2M. We can express the number, denoted by $x_{\min}$, as

$$x_{\min} = \inf\{x : T_{22}^{\text{M2M}}(x) = T_{12}\} \tag{11}$$

where T_{22}^{M2M} is represented as a function of x and T_{12} is a scalar for a given value of m. We obtain

$$\begin{aligned} x_{\min} &= \frac{\lambda_2}{\{\frac{\lambda_2}{m} - \frac{\lambda_1}{N}\}} \\ &= m \cdot \frac{1}{1 - \frac{m}{cN}} \quad (> m), \end{aligned} \tag{12}$$

where $c = \frac{\lambda_2}{\lambda_1}$.

More generally, if we have a stringent requirement that the RACH throughput for M2M should not be less than a certain threshold value, say η_T, the number of preambles required to meet the requirement is given by

$$x_{\min}^{\text{real}} = \inf\{x : T_{22}^{\text{M2M}}(x) \geq \eta_T\}, \tag{13}$$

and the integer solution can simply be obtained as

$$x_{\min}^{\text{int}} = \Big\lceil \inf\{x : T_{22}^{\text{M2M}}(x) \geq \eta_T\} \Big\rceil. \tag{14}$$

Supposing that $x_{\min}$ preambles are assigned to Zone 2 in Method 2 so that the M2M RACH throughput in Method 2 is equivalent to that in Method 1, we now compare the H2H RACH throughputs in both methods. Since the H2H RACH requests are supposed to use one of preambles assigned to both zones, the H2H RACH throughput is given by

$$T_2^{\text{H2H}} = T_{21} + T_{22}^{\text{H2H}}. \tag{15}$$

From the fact that

$$T_{21} = \lambda_1 \Big(1 - \frac{x}{N}\Big) \cdot \exp\Big\{ -\frac{\lambda_1}{N} \Big\} \tag{16}$$

and from (12), we have

$$T_{21}\Big|_{x=x_{\min}} = \lambda_1\left(1-\frac{1}{\left(\frac{N}{m}-\frac{\lambda_1}{\lambda_2}\right)}\right)\cdot\exp\left\{-\frac{\lambda_1}{N}\right\}. \quad (17)$$

Similarly, we can obtain

$$T_{22}^{\mathrm{H2H}}\Big|_{x=x_{\min}} = \lambda_1\left(\frac{1}{\left(\frac{N}{m}-\frac{\lambda_1}{\lambda_2}\right)}\right)\cdot\exp\left\{-\frac{\lambda_2}{m}\right\}. \quad (18)$$

From (17) and (18), we can rewrite (15) as

$$T_2^{\mathrm{H2H}} = \lambda_1\left(1-\frac{1}{\left(\frac{N}{m}-\frac{\lambda_1}{\lambda_2}\right)}\right)\cdot\exp\left\{-\frac{\lambda_1}{N}\right\} + \lambda_1\left(\frac{1}{\left(\frac{N}{m}-\frac{\lambda_1}{\lambda_2}\right)}\right)\cdot\exp\left\{-\frac{\lambda_2}{m}\right\}. \quad (19)$$

We now compare T_{11} and T_2^{H2H}.

TABLE II
H2H RACH THROUGHPUT COMPARISON.

Method	H2H RACH Throughput
Method 1	$\lambda_1\cdot\exp\left\{-\frac{\lambda_1}{(N-m)}\right\}$
Method 2	$\lambda_1\left[(1-\alpha)\cdot\exp\left\{-\frac{\lambda_1}{N}\right\}+\alpha\cdot\exp\left\{-\frac{\lambda_2}{m}\right\}\right]$

Note: $\alpha^{-1} = \left(\frac{N}{m}-\frac{\lambda_1}{\lambda_2}\right)$.

V. Numerical Examples

We examine the throughput performance of the two methods for various settings of parameters values. We assume that the eNB assigns 54 preambles for UE's to use. In Method 1, H2H UE's can use one of 49 preambles whereas M2M UE's can use one of the rest, namely, one of the other 5 preambles. In Method 2, H2H UE's can use one of the 54 preambles whereas M2M UE's can use one of a certain number of preambles determined by (11).

Next three figures, Figs. 4-6, present the behavior of RACH throughput for H2H UE's according to the variation of the arrival rate of RA attempts of M2M UE's. The arrival rate of RA attempts of H2H UE's is 1, 10, and 100 for Fig. 4, 5, and 6, respectively.

When the arrival rate of RA attempts of H2H UE's is one, i.e., $\lambda_1 = 1$, it is shown in Fig. 4 that Method 2 has a slightly better throughput when the arrival rate of M2M RA attempts is smaller than 0.1 (Erlangs). However, it is also shown that this method suffers from a huge degradation in throughput when the arrival rate of M2M RA attempts is greater than 0.1. In the case that we use Access Barring Check as part of the RA procedure, the variation of arrival rate changes over time due to the Binomial trials observed in the Access Barring Check procedure.

In Fig. 5, it is presented that the throughput of Method 2 is approximately 2.5% better than that of Method 1 when $\lambda_2 < 1$ whereas the throughput of Method 2 is approximately 9% worse than that of Method 1. In the previous two cases where $\lambda_1 = 1, 10$, the degradation of throughput is larger than its improvement if we use Method 2.

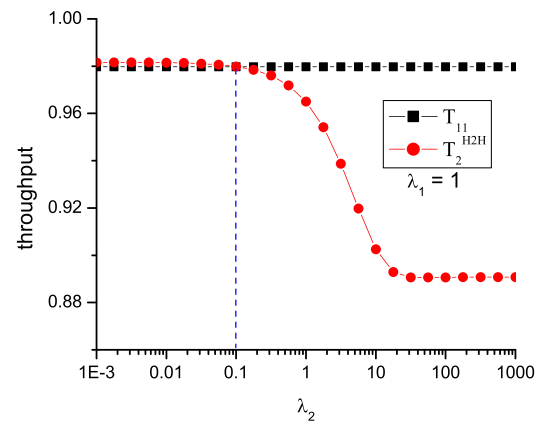

Fig. 4. Comparison of throughput for H2H customers. Throughput vs. λ_2 when $\lambda_1 = 1$.

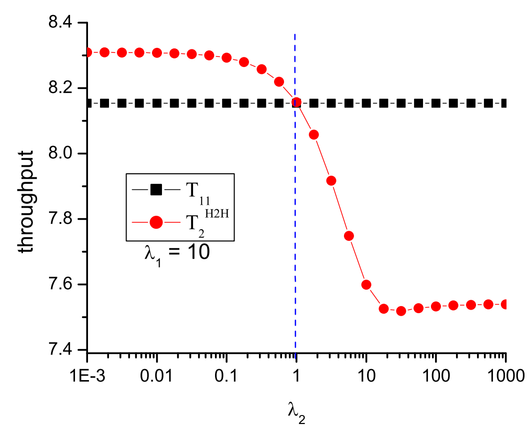

Fig. 5. Comparison of throughput for H2H customers. Throughput vs. λ_2 when $\lambda_1 = 10$.

Lastly in the first set of examples, we consider a very scarce case where the arrival rate of H2H RA attempts is very large, namely, $\lambda_1 = 100$. In this unusual case, it is shown that Method 2 performs better than Method 1 at all times. In Fig. 6, it is observed that the throughput of H2H customers has a valley, namely, the minimum around $\lambda_2 = 10$. This is because the overall throughput consists of the throughput in Zone 1 and that in Zone 2, which have different decreasing and increasing behaviors in x (Please see Eq. (19) for more details).

Next, we examine the throughput performance according to the variation of the arrival rate of H2H RA attempts when the arrival rate of M2M RA attempts is stationary. Figs. 7-9 present the results. It is observed that Method 2 provides throughput performance comparable to or slightly worse than Method 1 when the arrival rate of H2H RA attempts is smaller than 10 (Erlangs), which is a practical case. In Fig. 7 and Fig. 9, Method 2 performs better than Method 1 only when the rate of H2H RA attempts is greater than the rate of M2M RA attempts. However, this is not a practical situation even if it can happen with an extremely low probability.

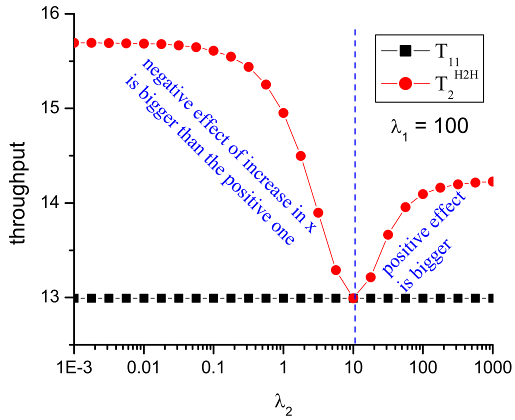

Fig. 6. Comparison of throughput for H2H customers. Throughput vs. λ_2 when $\lambda_1 = 100$.

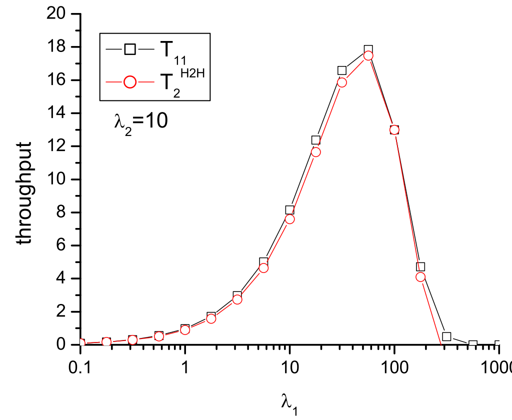

Fig. 8. Comparison of throughput for H2H customers. Throughput vs. λ_1 when $\lambda_2 = 10$.

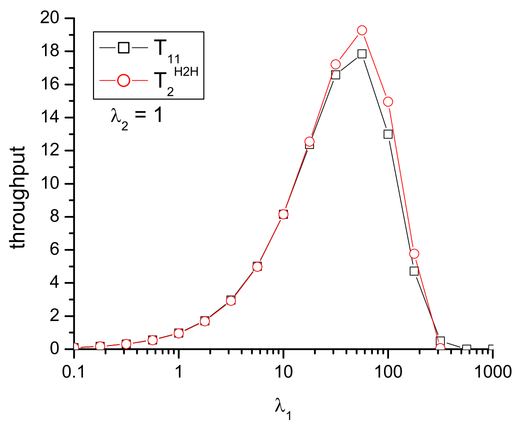

Fig. 7. Comparison of throughput for H2H customers. Throughput vs. λ_1 when $\lambda_2 = 1$.

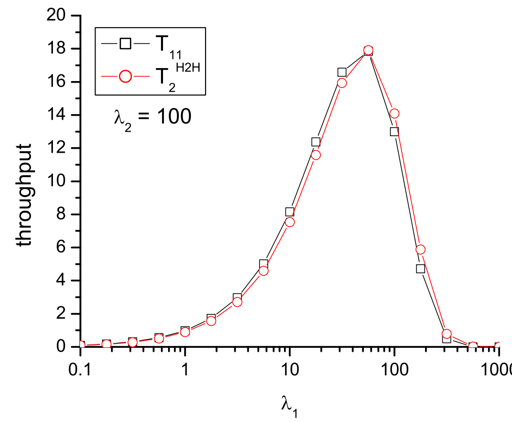

Fig. 9. Comparison of throughput for H2H customers. Throughput vs. λ_1 when $\lambda_2 = 100$.

VI. Conclusions

We have discussed the throughput performance of two potential candidate methods for RA preamble allocation and management, which are under discussion for possible adoption in Long Term Evolution (LTE)-Advanced. Method 1 was to completely split the set of available RA preambles into two disjoint subsets: one is for human-to-human (H2H) customers and the other for M2M customers/devices. Method 2 was also to split the set into two subsets: one is for H2H customers only whereas the other is for both H2H and M2M customers. Our results demonstrated that there is a boundary of RA load below which Method 2 performs slightly better than Method 1 but above which Method 2 degrades throughput to a large extent. Since it is one of the hot issues in 4G cellular networks how to efficiently handle the heavy RA load caused by the huge M2M population, our modeling and analysis can be utilized as a guideline to design the RA preamble resource management method.

References

[1] K.-D. Lee, "MAC PDU Signaling and operating methods for access class barring and back-off control for large-scale radio access network," US Patent Application 61/377,470, 2010.

[2] K.-D. Lee, S. Kim, and B.K. Yi, "Packet-Level Scheduling for Implant Communications Using Forward Error Correction in an Erasure Correction Mode for Reliable u-Healthcare Service," *J. Commun. Netw.*, vol. 13, no. 2, pp. 160-166, Apr. 2011.

[3] 3GPP TR 37.868, *Study on RAN Improvements for Machine-type Communications; (Release 10)* , v.0.7.0, Oct. 2010.

[4] K.-D. Lee, "Efficient Management for Preamble Set Separation for Random Access Control in a Large Scale Cellular Network with Both Human-Type Users and Machine-Type Communication Devices," US Patent Application, 61/510,443, 2011.

[5] 3GPP TS 36.300, *Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2*, v.10.4.0, Jun. 2011.

[6] 3GPP RAN WG2 Meeting 75, Athens, Greece, Aug. 2011.

[7] W. Luo and A. Ephremides, "Stability of N interacting queues in random-access systems," *IEEE Trans. Info. Theory*, vol. 45, no. 5, pp. 1579–1587, Jul. 1999.

[8] K.-D. Lee, "Preamble set separation for random access control in large scale cellular networks," US Patent Application 61/387,008, 2010.

[9] K.-D. Lee, "Study on throughput performance behaviour in preamble separation with partial overlap among differet user classes," *LG Electronics Mobile Research Technical Memo*, May 2011.