On the Adaptive Determination of the Number of Preambles in RACH for MTC

Jinho Choi

Abstract—**Recently, in the Long Term Evolution-Advanced (LTE-A) system, a random access method, called random access (RACH) procedure, has been proposed for machine-type communications (MTC) to support a number of devices. Since the RACH procedure is similar to multichannel ALOHA, the fast retrial can be employed. In this letter, when the fast retrial is used for the RACH procedure, we propose an adaptive approach that can adaptively decide the number of preambles to maximize the throughput without knowing the number of devices and the access probability.**

Index Terms—**Machine-type communications, RACH procedure, access delay.**

I. Introduction

THERE HAS been a growing interest in machine-type communications (MTC) or machine-to-machine (M2M) communications that support a number of devices in a network [1], [2]. While there are various applications of MTC (e.g., smart grid, health care, and so on), technologies to enable MTC would be different from those for human-to-human (H2H) communications due to different characteristics of M2M communications. For example, there might be a number of devices that have low activity (in terms of transmitting data) in MTC. On the other hand, in H2H communications, the number of users (of high activities or high data rates) in a system would not be large due to the limitation of available spectrum.

In the long term evolution-advanced (LTE-A) system, a random access method, called random access (RACH) procedure, has been proposed for MTC [3]. This approach is extended in [4] to a hybrid system where H2H and M2M (or MTC) communications are combined in terms of resource allocations.

The RACH procedure is a contention-based random access method, which is similar to the slotted ALOHA protocol [5]. In the RACH procedure, there are multiple preambles in a preamble pool. A device can transmit a preamble from the preamble pool randomly and it can be connected if there is no collision. In this sense, the RACH procedure can be seen as a multichannel ALOHA [6], where congestion control is studied by estimating the number of arrivals (or the number of devices that send preambles).

In this letter, we model the RACH procedure as a multichannel ALOHA with the fast retrial [7]. To maximize the throughput when the number of devices and their access probability are not available, we propose an approach that adaptively decides the number of preambles or the size of preamble pool at an access point or base station (BS). This approach would be effective if devices in MTC cannot modify its access probability to maximize the throughput. Note that if devices can change their access probability, the approaches in [6], [8] could be used.

Manuscript received January 16, 2016; accepted March 19, 2016. Date of publication March 23, 2016; date of current version July 8, 2016. This work was supported by GIST (the title of the project is *Living Energy*). The associate editor coordinating the review of this paper and approving it for publication was D. Qiao.

The author is with the School of Information and Communications, Gwangju Institute of Science and Technology (GIST), Gwangju, South Korea (e-mail: jchoi0114@gist.ac.kr).

Digital Object Identifier 10.1109/LCOMM.2016.2546238

II. System Model

Suppose that a system consists of a BS and a number of devices. According to the RACH procedure [3], if a device wants to connect to the BS, it sends a preamble that is randomly chosen from a set of preambles or preamble pool. If the BS receives this preamble without collision, the connection from this device to the BS can be established. However, if multiple devices transmit the same preamble, the BS ignores the collided preamble.

Suppose that there are L preambles in the preamble pool, which is denoted by $\mathcal{C} = \{\mathbf{c}_1, \dots \mathbf{c}_L\}$. Note that in [3], L is fixed as 54. However, in this letter, we assume that the number of preambles, L, can be variable[1] and optimized in the RACH procedure as the key design parameter. In general, the duration of RACH, denoted by T_R, is proportional to the number of preambles, L. Thus, we have

$$T_R = \tau L, \tag{1}$$

where τ is constant. For convenience, we assume that $\tau = 1$ throughout the letter. In general, if L is large, the probability of collisions is low, which implies that a device can establish a connection to the BS when it needs without severe contention. However, there is a price to be paid for this benefit, which is the increase of the duration of RACH as shown in (1) in a system of fixed bandwidth. On the other hand, in a multicarrier system, T_R remains unchanged when L increases, while more subcarriers are to be allocated for RACH. Thus, in this case, the increase of L results in the increase of the bandwidth of RACH.

In random access, to resolve contention, each device involved in the collision may retransmit a preamble with a random back-off time [5]. Alternatively, exploiting the nature of multiple channels [7], a device involved in the collision can immediately retransmit another randomly chosen preamble in the next stage, which is called the fast retrial. In this letter, we employ the fast retrial for the retransmission of preamble by the devices in the collision.

[1] We may use random spreading sequences of low cross-correlation or Zadoff-Chu sequences to generate L preambles as in LTE systems. Alternatively, we can use L (orthogonal) time slots for a preamble pool.

III. Throughput

Suppose that there are M devices that transmit preambles randomly chosen from $\mathcal{C}$. The probability that a device can transmit its preamble without collision is given by $P_s = \left(1 - \frac{1}{L}\right)^{M-1}$. Denote by N the number of the devices that can successfully transmit their randomly chosen preambles. For a given M, the conditional mean of N becomes

$$\mathbb{E}[N \mid M] = M P_s = M \left(1 - \frac{1}{L}\right)^{M-1}, \quad (2)$$

where $\mathbb{E}[\cdot]$ denotes the statistical expectation. The throughout, which is denoted by T, can be defined by the ratio of N to the duration of RACH. For a given M, the conditional throughput becomes

$$\mathbb{E}[T \mid M] = \frac{\mathbb{E}[N \mid M]}{T_R} = \frac{M\left(1 - \frac{1}{L}\right)^{M-1}}{L}. \quad (3)$$

It can be shown that $\mathbb{E}[T|M]$ is a $\cap$-shape function of L for a fixed M. The maximum of $\mathbb{E}[T|M]$ is achieved when $L = M$. For a large L, the maximum becomes $\mathbb{E}[T \mid M = L] \approx e^{-1}$.

If each device is to transmit a preamble with a probability p_a, M becomes a random variable that can be expressed as $M = \sum_{k=1}^{K} X_k$, where K is the total number of devices in the system and X_k is the binary random variable that becomes 1 if device k transmits a preamble and 0 otherwise. Thus, we have

$$\Pr(M = m) = \pi_m = \binom{K}{m} p_a^m (1 - p_a)^{K-m}. \quad (4)$$

Note that p_a needs to include the backlogged arrivals in the previous stage. That is,

$$p_a = p + \frac{\beta}{K}, \quad (5)$$

where p is the access probability of each device and β is the average number of the collided preambles.

From (4), after some manipulations, the mean of N can be found as

$$\bar{N} = \mathbb{E}[N] = \mathbb{E}[\mathbb{E}[N \mid M]] = \sum_{m=0}^{K} m \left(1 - \frac{1}{L}\right)^{m-1} \pi_m = p_a K \left(1 - \frac{p_a}{L}\right)^{K-1}. \quad (6)$$

The average throughput, $\mathbb{E}[T]$, is now given by

$$\mathbb{E}[T] = \frac{\bar{N}}{L} = \frac{p_a K \left(1 - \frac{p_a}{L}\right)^{K-1}}{L}. \quad (7)$$

The average throughput is also a $\cap$-function of L and its maximum is achieved when

$$L = L^* = p_a K = \mathbb{E}[M].$$

For a large $L = L^*$, the maximum average throughput becomes $\mathbb{E}[T] \approx e^{-1}$.

In Fig. 1, we illustrate the difference between the average and conditional throughputs. We can see that the conditional throughput is slightly higher than the average throughput if $L = M = p_a K$ and the difference decreases with M. Note that since M is a random variable, it is not possible to set $L = M$ in advance. Thus, if K and p_a (or the product, $p_a K$) are known, L can be set to $\lceil p_a K \rceil$ to maximize the average throughput where $\lceil x \rceil$ represents the smallest integer that is greater than or equal to x. Unfortunately, in practice, it is not easy to know K and p_a. In addition, there is no closed-form expression for p_a as discussed in [7]. Thus, we may need to consider the estimation of K and p_a to decide L. Alternatively, we could estimate M from a quantity that is available at the BS.

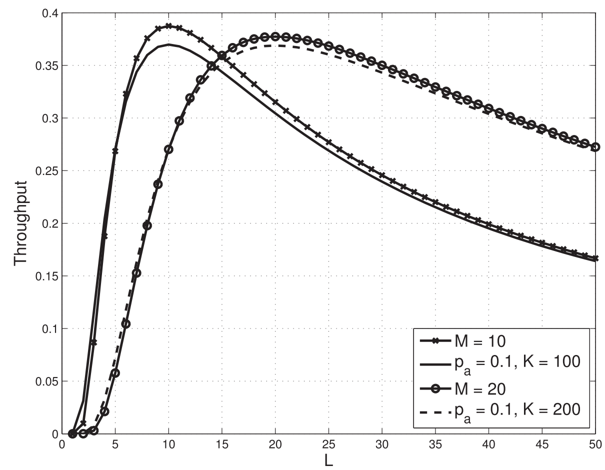

Fig. 1. The conditional and average throughputs for various values of L.

IV. Adaptive Preamble Pool

In this section, we first consider the estimation of M, which would be a random variable, at the BS. With the estimate of M, we then derive an approach to decide L.

A. Estimation of M

The conditional probability of N for a given M is

$$\Pr(N = n \mid M = m) = \binom{m}{n} P_s^n (1 - P_s)^{m-n} = \binom{m}{n} q^{(m-1)n} \left(1 - q^{m-1}\right)^{m-n}, \quad (8)$$

where $q = 1 - \frac{1}{L}$. Note that the number of the devices that can successfully transmit preambles, N, is limited as $N \le L$, since there are L preambles. In addition, N cannot be greater than M. Thus, the maximum likelihood (ML) estimate of M from N is given by

$$\hat{M}_{\text{ml1}}(n) = \underset{n \le m \le L}{\text{argmax}} \Pr(N = n \mid M = m). \quad (9)$$

Unfortunately, if $M > L$ (i.e., there are more active devices than preambles), this ML estimation does not work. To avoid this difficulty, we can use the number of collided preambles to estimate M. Let J denote the number of collided preambles. Clearly, $J = M - N$. The conditional probability of J for given $M = m$ becomes

$$\Pr(J = j \mid M = m) = \binom{m}{j} P_c^j (1 - P_c)^{m-j}, \quad (10)$$

where $P_c = 1 - P_s = 1 - q^{m-1}$. Thus, the ML estimate of M from J can be given by

$$\hat{M}_{\text{ml2}}(j) = \underset{j \le m < \infty}{\operatorname{argmax}} \Pr(J = j \mid M = m). \tag{11}$$

It is noteworthy that we do not need to know K and p_a (or p) in this ML estimation of M.

In order to find the ML estimate in (11), an exhaustive search method can be used. In this case, the complexity becomes high. In addition, for a large j, the binomial coefficient $\binom{m}{j}$ is too large and the complexity to find its value is also high. To avoid this difficulty, we can find the ML estimate using the mode of the binomial distribution. For a given m, there exists $j(m)$ that satisfies the following inequality:

$$\binom{m}{j} P_c^j (1 - P_c)^{m-j} \le \binom{m}{\psi(m)} P_c^{\psi(m)} (1 - P_c)^{m-\psi(m)}, \tag{12}$$

where $\psi(m)$ is the mode of the binomial distribution, which is $\psi(m) = \lfloor (m+1) P_c \rfloor$. For a given j, the ML estimate of M should satisfy the following equality:

$$j = \lfloor (m+1)(1 - q^{m-1}) \rfloor. \tag{13}$$

Since there can be multiple m's that satisfy (13), we need to find their likelihood values to choose the best m that maximizes the likelihood function.

Note that in [6], an approach to estimate M using the idle probability is considered, which is given by $\hat{M} = -L \ln \hat{P}_I$, where $\hat{P}_I$ is an estimate of the idle probability. The ratio of the number of ideal preambles to L is used as an estimate of P_I, $\hat{P}_I$. Unfortunately, this estimate does not provide a good performance as it is not optimal one and sensitive to a small change of $\hat{P}_I$ as mentioned in [6] (thus, there should be an additional step, fitting or smoothing).

B. Adaptive Determination of L

To maximize the conditional throughput in (3), we can consider a stochastic gradient ascending method [9]. To this end, we need to obtain the derivative of $\mathbb{E}[T \mid M]$ with respect to L, which is $\frac{d\mathbb{E}[T \mid M]}{dL} = \frac{M(M-L)}{L^3}\left(1 - \frac{1}{L}\right)^{M-2}$. Let L_t denote the value of L at stage t. In addition, $\hat{M}_t$ represents the ML estimate of M at stage t. Then, for given L_t and $\hat{M}_t$, the derivative becomes

$$\nabla_t(\hat{M}_t, L_t) = \frac{\hat{M}_t(\hat{M}_t - L_t)}{L_t^3}\left(1 - \frac{1}{L_t}\right)^{\hat{M}_t - 2}. \tag{14}$$

Note that since $\nabla_t(\hat{M}_t, L_t)$ is small for large L_t and M_t, the convergence rate of the stochastic gradient ascending method could be slow. To avoid it, we can normalize with the second order derivative, which is given by

$$\frac{d^2\mathbb{E}[T \mid M]}{dL^2} = \frac{M}{L^5}\left(1 - \frac{1}{L}\right)^{M-3} \times \left(L - L^2 - (M - L)(L - 3M + 1)\right).$$

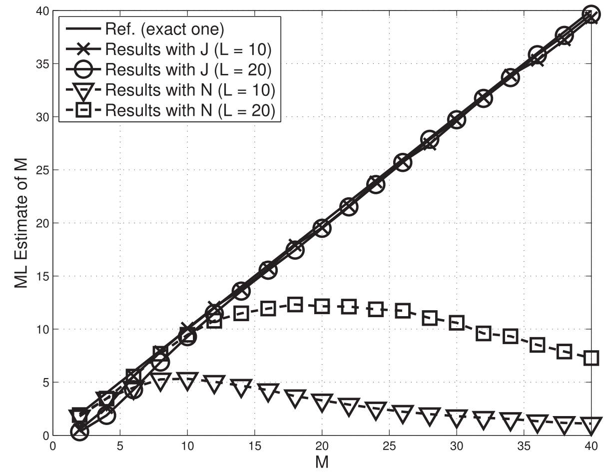

Fig. 2. The ML estimates of M from N and J.

It is noteworthy that the conditional throughput is not concave in L for given M. Thus, the second order derivative is not always negative. However, when $L \approx M$ (i.e., near to the solution), we can see that $\frac{d^2\mathbb{E}[T \mid M]}{dL^2} \approx H_o = -\frac{M}{L^3}\left(1 - \frac{1}{L}\right)^{M-2} < 0$. Thus, the ratio of the first order derivative to the approximate second order derivative becomes

$$\bar{\nabla}_t(\hat{M}_t, L_t) = \frac{\nabla_t(\hat{M}_t, L_t)}{H_o} = L_t - \hat{M}_t, \tag{15}$$

which implies that the stochastic gradient method with (15) is to find L that minimizes $\mathbb{E}[(M - L)^2]$. With the step size α, we can have the following iterative method to decide the value of L:

$$\begin{aligned} \hat{L}_{t+1} &= \max\{\hat{L}_t - \alpha \bar{\nabla}_t(\hat{M}_t, \hat{L}_t), L_{\min}\} \\ &= \max\{\hat{L}_t + \alpha(\hat{M}_t - \hat{L}_t), L_{\min}\}, \end{aligned} \tag{16}$$

where $L_{\min}$ is the minimum value of L. At the beginning of every stage, the BS can broadcast $L = \lceil \hat{L}_t \rceil$ to all devices so that a device can choose a preamble of the index between 1 and L. Note that any rounding operation should not be used in the recursion in (16) for an integer-valued $\hat{L}_{t+1}$ to avoid the accumulation of rounding errors. In addition, the step-size, α, can be decided to enjoy the tradeoff between convergence speed and steady-state performance [9].

V. Simulation Results

In this section, we present simulation results of the estimation of M from J. Then, we show simulation results of the throughput with estimated M.

Fig. 2 shows the results of the ML estimation of M with J and N when $L \in \{10, 20\}$. We have 2000 runs for each value of M to find the mean values of the ML estimates. If $M \ll L$, we can see that the ML estimation of M from N might be reasonable. However, if M is approaching L, the ML estimate from N is biased. However, the ML estimate from J can provide good results for any value of M.

To decide the number of L, we use the iterative method in (16) with $L_{\min} = 10$ and $\alpha = 0.1$. It is assumed that $K = 1000$

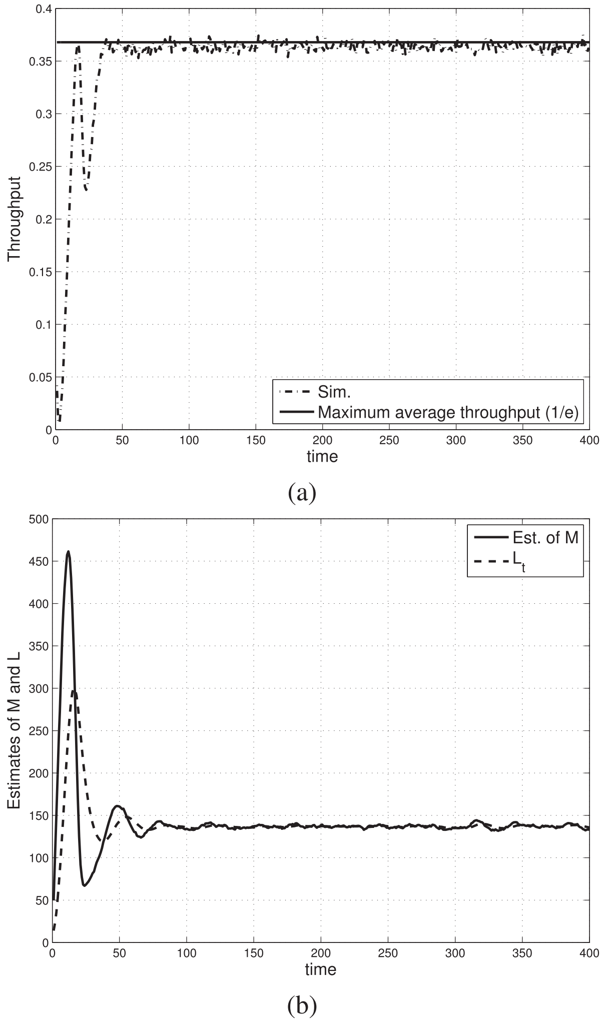

Fig. 3. Throughput and L_t with $K = 1000$ and $p = 0.05$: (a) throughput; (b) the values of L_t and $\hat{M}_t$.

and $p = 0.05$. The initial value of L is set to $L_0 = L_{\min}$. The simulation results are shown in Fig. 3, where each curve is obtained by averaging of 100 runs. In Fig. 3(a), we show the throughput. The maximum average throughput is $e^{-1} = 0.3679$. It is shown that by adaptively deciding the number of preambles, L_t, the maximum throughput can be achieved. Note that in deciding L, we do not need to know K and p_a.

In Fig. 3(b), we show the values of $\hat{M}_t$ and L_t. It is interesting to see that $\hat{M}_t$ increases in the beginning. Since the initial value of L is set to $L = L_{\min} = 10$, the number of preambles is too small to support active devices (there should be $Kp = 50$ active devices on average without backlogged ones). Thus, there might be more and more collided preambles, which increases M. However, L_t will also increase, which results in the decrease of M. After the transient behavior, we can see that $L_t \approx \hat{M}_t$. It is noteworthy that the steady state value of M is usually greater than $Kp = 50$ due to retransmissions.

Fig. 4 shows the throughput curve when the access probability, p, is periodically updated between 0.1 and 0.05 every 100 stages. We assume that $K = 1000$ and the step size is 0.2 to adjust the size of the preamble pool, L. The size of

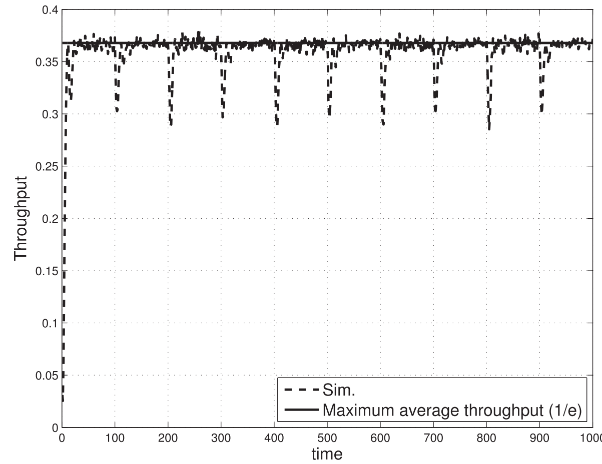

Fig. 4. Throughput when the access probability, p, is periodically updated between 0.1 and 0.05 every 100 stages with $K = 1000$.

the preamble pool needs to be updated to adapt to the change of p or time-varying traffic conditions. As shown in Fig. 4, the throughput is degraded when p changes. However, as L is adaptively adjusted, the throughput is recovered soon.

VI. Conclusions

In this letter, to maximize the throughput of the RACH procedure in LTE for MTC, we proposed to adaptively decide the size of the preamble pool, L. This approach could be effective when it is difficult to adjust the access probability of devices. A recursion to update the size of the preamble pool was derived using the estimate of the number of active devices from the number of collided preambles. From simulations, we can confirm that it is possible to maximize the throughput by adaptively deciding the size of the preamble pool at the BS without knowing the total number of devices and their access probability in a system.

References

[1] M. Hasan, E. Hossain, and D. Niyato, "Random access for machine-to-machine communication in LTE-Advanced networks: Issues and approaches," *IEEE Commun. Mag.*, vol. 51, no. 6, pp. 86–93, Jun. 2013.

[2] F. Ghavimi and H.-H. Chen, "M2M communications in 3GPP LTE/LTE-A networks: Architectures, service requirements, challenges, and applications," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 2, pp. 525–549, May 2015.

[3] 3GPP TR 37.868 V11.0, "Study on RAN improvements for machine-type communications," Oct. 2011.

[4] Y.-C. Pang, S.-L. Chao, G.-Y. Lin, and H.-Y. Wei, "Network access for M2M/H2H hybrid systems: A game theoretic approach," *IEEE Commun. Lett.*, vol. 18, no. 5, pp. 845–848, May 2014.

[5] B. Bertsekas and R. Gallager, *Data Networks*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1987.

[6] O. Arouk and A. Ksentini, "Multi-channel slotted aloha optimization for machine-type-communication," in *Proc. 17th ACM Int. Conf. Model. Anal. Simul. Wireless Mobile Syst.*, 2014, pp. 119–125.

[7] Y.-J. Choi, S. Park, and S. Bahk, "Multichannel random access in OFDMA wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 3, pp. 603–613, Mar. 2006.

[8] O. Galinina, A. Turlikov, S. Andreev, and Y. Koucheryavy, "Stabilizing multi-channel slotted aloha for machine-type communications," in *Proc. IEEE Int. Symp. Ing. Theory (ISIT)*, Jul. 2013, pp. 2119–2123.

[9] J. C. Spall, *Introduction to Stochastic Search and Optimizatin*. Hoboken, NJ, USA: Wiley, 2003.