On the Optimization of Outage Probability of Access Delay of MTDs in Cellular Networks for URLLC

Yunshan Yang*, Wen Zhan† and Lin Dai*
*Department of Electrical Engineering, City University of Hong Kong
†School of Electronics and Communication Engineering, Sun Yat-sen University
yunshyang3-c@my.cityu.edu.hk, zhanw6@mail.sysu.edu.cn, lindai@cityu.edu.hk

***Abstract*—This paper focuses on the outage probability of access delay of Machine-to-Machine (M2M) communications in cellular networks, which is an important performance indicator for Ultra-Reliable and Low-Latency Communication (URLLC). Specifically, by deriving the outage probability for given maximum allowable access delay as a function of system parameters, the outage probability is minimized by optimally tuning the Access Class Barring (ACB) factor. For given outage probability bound, the admission control and resource allocation for the random access channel are further discussed, where the maximum number of Machine-Type Devices (MTDs) that can be admitted with given number of preambles and the minimum number of preambles that should be allocated with given network size are obtained. Compared to the standard setting where the ACB factor is fixed, significant gains in outage probability are demonstrated by optimally tuning the ACB factor according to the number of MTDs and the traffic input rate of each MTD. It is also shown that for given required outage probability bound, the optimal tuning of ACB factor enables much more MTDs to be admitted for given preamble resource, and requires much fewer preambles for given network size.**

***Index Terms*—M2M communications, random access, LTE, outage probability, access delay, URLLC.**

I. Introduction

The Machine-to-Machine (M2M) communications is supposed to embrace mission-critical services, including telesurgery, intelligent transportation, and industry automation [1], with Ultra-Reliable and Low-Latency Communication (URLLC). The URLLC aims at providing those services with ultra-low latency of 1 ms to 10 ms and ultra-high reliability of more than 99.999% in terms of packet delivery performance [2]. However, studies have shown that as the main source of delay in cellular networks, the initial connection establishment procedure in the random access channel may take tens of milliseconds [3]. The delay performance becomes even worse when the network size grows [4], which fails to meet the stringent URLLC requirement. Therefore, how to improve the access delay performance and support URLLC in the random access channel of cellular networks has become a significant challenge.

Specifically, the access delay of each Machine-Type Device (MTD) is defined as the time spent from the generation of an access request until its successful transmission. To fulfill the stringent delay requirement of URLLC, various centralized control methods have been proposed in [5]–[8], where dedicated resource is reserved to each MTD. However, centralized control incurs significant coordination overhead when the number of MTDs is large, making it unscalable for large-scale networks. In contrast, distributed control, which has minimum coordination, attracts more attention and has been considered as the key approach to support URLLC [9].

With distributed control, nevertheless, severe congestion could occur if a massive number of MTDs contend for channel access at the same time, causing excessively long access delay due to low chances of successful access [10]. The outage probability of access delay, i.e., the probability that the access delay exceeds a given maximum allowable delay requirement, is an important performance indicator for URLLC, which has been studied in [11]–[13]. Specifically, in [11], [12], explicit expressions of outage probability of access delay were obtained based on the number of access requests in each time slot, which, however, varies with time and is difficult to track accurately in practice due to the uncoordinated nature of MTDs. In [13], an upper-bound of the outage probability of access delay was developed as a function of the traffic input rate of each MTD that can be easily measured. Nevertheless, the gap between the upper-bound and the actual outage probability was found to be increasing as the maximum allowable delay becomes larger.

To support URLLC in the random access channel, it is of crucial importance to minimize the outage probability for given access delay requirement by optimally tuning the access parameters of MTDs, which, unfortunately, has largely remained little understood due to the lack of proper characterization of access delay. In our recent work [14], a new analytical framework was established to optimize the access throughput of M2M communications in Long Term Evolution (LTE) cellular networks, and extended in [15] to minimize

This work was supported in part by National Natural Science Foundation of China under Grant 62001524, in part by the Shenzhen NSF Project under Grant JCYJ20200828214622001.

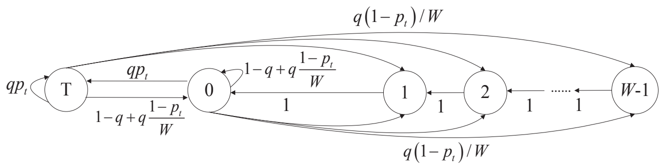


Fig. 1. State transition diagram of each individual access request [14].

the mean access delay. In this paper, we will further study how to optimize the outage probability of access delay for achieving URLLC requirements. Specifically, by deriving the explicit expression of outage probability for given maximum allowable access delay, we obtain the minimum outage probability with the Access Class Barring (ACB) factor (i.e., the initial transmission probability of each MTD) optimally chosen. For given outage probability bound, we further derive the maximum number of MTDs that can be admitted for given number of preambles, and the minimum number of preambles that should be allocated for given network size. Simulation results corroborate that compared to the standard setting where the ACB factor is fixed, significant gains in outage probability can be achieved by optimally tuning the ACB factor, and the outage probability bound can be satisfied by properly controlling the number of admitted MTDs or allocating preamble resource.

The remainder of this paper is organized as follows. Section II presents the system model and the preliminary analysis. The outage probability of access delay is characterized and optimized in Section III. For given bound of outage probability, the maximum number of MTDs and the minimum required number of preambles are obtained in Section IV. The analysis is verified by the simulation results presented in Section V.

II. System Model and Preliminary Analysis

Consider a single-cell LTE system with n MTDs attempting to access the Base Station (BS). Assume that for each MTD, one access request is generated once the MTD has data packets in its buffer. In the random access procedure, each MTD randomly selects one out of M orthogonal preambles and transmits its access request to the BS through the Physical Random Access CHannel (PRACH), which appears periodically [16]. We define the time slot as the interval between two consecutive PRACHs. In each time slot, an access request is successful if and only if there are no concurrent access request transmissions using the same preamble. Otherwise, a collision occurs and all of them fail.

To characterize the behavior of each access request, a discrete-time Markov process $\{X_j, j = 0, 1, ...\}$ has been established in [14] with state transition diagram of each individual access request shown in Fig. 1, where $q \in (0, 1]$ denotes the ACB factor and $W \in \{1, 2, ...\}$ represents the Uniform Backoff (UB) window size. A detailed description of the state transition process of Fig. 1 can be found in Section II-B in [14].

Let D_i denote the time spent from the beginning of State i until the service completion, where $i \in \{T, 0, 1, ..., W-1\}$. A fresh access request is initially in State T and its service completes when it shifts back into State T. Therefore, D_T is the service time of access request, which is also the access delay of each MTD. It was derived in [15] that the probability generating function of D_T is given by

$$G_{D_T}(z) = \frac{W(1-z)zqp}{W(1-z)(1-(1-q)z) - zq(1-z^W)(1-p)}, \tag{1}$$

where $p = \lim_{t\to\infty} p_t$ is the steady-state probability of successful transmission of access requests.

It was shown in [15] that when the UB window size $W = 1$, the first and second moments of access delay D_T can be optimized simultaneously. Therefore, in this paper, we set UB window size W to 1 and only consider the effect of ACB factor q on the access delay performance. Different from [15], which focused on the optimization of the mean access delay, in this paper, we will study the outage probability P_{out}, that is, the probability that the access delay of an access request exceeds a given maximum allowable delay D_b. In the following sections, we will first derive the minimum outage probability by optimally tuning the ACB factor, and then focus on the admission control and resource allocation for given bound on outage probability.

III. Minimum Outage Probability

For given maximum allowable access delay D_b, the outage probability that the access delay D_T exceeds D_b can be written as

$$P_{out} = \Pr\{D_T > D_b\} = \sum_{i=D_b+1}^{\infty} \Pr\{D_T = i\}. \tag{2}$$

With the UB window size $W = 1$, the probability generating function of the access delay D_T can be simplified to

$$G_{D_T}(z) = \frac{zpq}{1-z+zpq}. \tag{3}$$

By combining (2) and (3), we can obtain that

$$P_{out} = (1-pq)^{D_b}, \tag{4}$$

from which it can be seen that the outage probability P_{out} is crucially determined by the maximum allowable delay D_b, the ACB factor q, and the steady-state probability of successful transmission of access requests p.

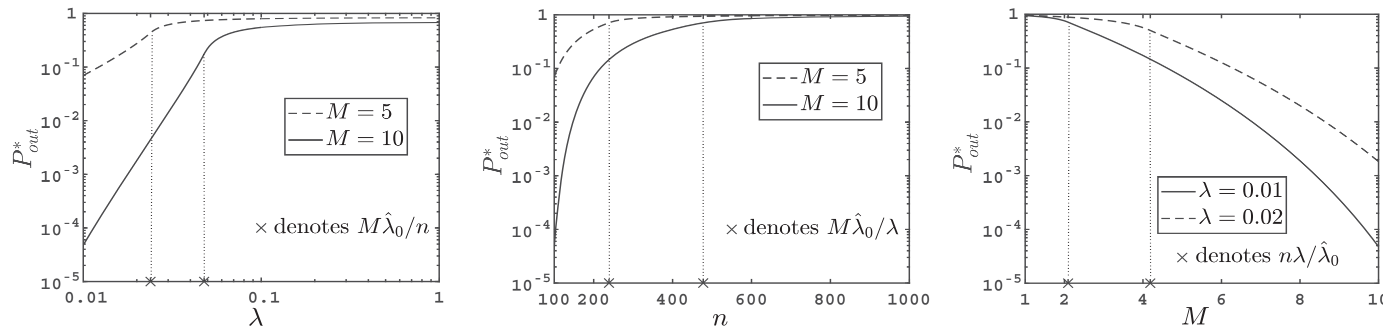
Note that the fixed-point equation of p was obtained in [14] as

$$p = \exp\left(-\frac{n\lambda/M}{\lambda/q+p}\right), \tag{5}$$

where M is the number of preambles, and λ is the traffic input rate of each MTD, i.e., the average number of arrived packets in each time slot per MTD. It was further obtained in [15] that the mean access delay $E[D_T] = \frac{1}{qp} + \frac{(1-p)(W-1)}{2p}$, which can be written as $E[D_T] = \frac{1}{qp}$ when $W = 1$. By further combining (4), the relationship between the outage probability P_{out} and the mean access delay $E[D_T]$ can be obtained as

$$P_{out} = \left(1 - \frac{1}{E[D_T]}\right)^{D_b}, \tag{6}$$

indicating that when the mean access delay $E[D_T]$ is minimized, the outage probability P_{out} is also minimized. The



(a) (b) (c)

Fig. 2. Minimum outage probability P^*_{out} versus (a) the input rate per MTD λ, (b) the number of MTDs n, and (c) the number of preambles M. $D_b = 10$ slots. $q = q^*_P$. (a) $n = 100$. $M \in \{5, 10\}$. (b) $\lambda = 0.01$. $M \in \{5, 10\}$. (c) $n = 100$. $\lambda \in \{0.01, 0.02\}$.

optimal ACB factor for minimizing mean access delay has been derived as Eq. (14) in [15]. Therefore, by denoting the optimal ACB factor for minimizing the outage probability as q^*_P, we have

$$q^*_P = \begin{cases} \frac{4M\mathbb{W}^2_{-1}\left(-\sqrt{n\lambda/M}/2\right)}{n\left(-2\mathbb{W}_{-1}\left(-\sqrt{n\lambda/M}/2\right)-1\right)} & \text{if } \lambda \leq \frac{M}{n}\hat{\lambda}_0, \\ \frac{\lambda}{\frac{n\lambda}{M}-e^{-1}} & \text{otherwise,} \end{cases} \tag{7}$$

according to Eq. (14) in [15], where $\hat{\lambda}_0 \approx 0.48$ is the single non-zero root of the equation $\hat{\lambda} - \hat{\lambda}\left(1 + 1/\mathbb{W}_{-1}\left(-\sqrt{\hat{\lambda}}/2\right)\right)^2 = 4\left(\hat{\lambda} - e^{-1}\right)$.

By substituting (7) into (4) and (5), the corresponding minimum outage probability, denoted by $P^*_{out} = \min_q P_{out}$, can then be obtained as

$$P^*_{out} = \begin{cases} \left(1 - \frac{4M\mathbb{W}^2_{-1}\left(-\sqrt{n\lambda/M}/2\right)p_L^{q=q_1}}{n\left(-2\mathbb{W}_{-1}\left(-\sqrt{n\lambda/M}/2\right)-1\right)}\right)^{D_b} & \text{if } \lambda \leq \frac{M}{n}\hat{\lambda}_0, \\ \left(1 - \frac{\lambda}{\frac{en\lambda}{M}-1}\right)^{D_b} & \text{otherwise,} \end{cases} \tag{8}$$

where $p_L^{q=q_1}$ is the larger non-zero root of the following fixed-point equation of p:

$$n\lambda\left(1 + 2\mathbb{W}_{-1}\left(-\sqrt{n\lambda/M}/2\right)\right) = \left(4Mp + \frac{4n\lambda}{\ln p}\right)\mathbb{W}^2_{-1}\left(-\sqrt{n\lambda/M}/2\right). \tag{9}$$

Eq. (8) shows that the input rate per MTD λ, the number of MTDs n, and the number of preambles M are key system parameters that determine the minimum outage probability P^*_{out}. As Fig. 2 illustrates, P^*_{out} is a monotonic increasing function of λ and n, and a monotonic decreasing function of M. Specifically, Figs. 2a and 2b show that the minimum outage probability P^*_{out} increases with the traffic input rate per MTD λ and the number of MTDs n linearly and exponentially, respectively, when λ or n is small, but quickly approaches the limits, i.e., $P^*_{out}|_{\lambda=1} = \left(1 - \frac{1}{\frac{en}{M}-1}\right)^{D_b}$ and $P^*_{out}|_{n\to+\infty} = 1$. To reduce P^*_{out}, more resource, i.e., preambles, should be allocated. Fig. 2c shows that with $M \geq 4$, even a small increment of M, e.g., 2, can efficiently reduce the minimum outage probability P^*_{out} by an order of magnitude.

We can see from Fig. 2 that for given input rate per MTD λ, the minimum outage probability P^*_{out} could be extremely high when the number of MTDs n is too large or the number of preambles M is too small. In practice, for given maximum allowable access delay, the outage probability is usually required to be bounded. For instance, in smart factory, the maximum allowable access delay D_b ranges from 0.5 to few tens of milliseconds, and the outage probability of access delay is required to be no larger than a bound, denoted by P_b, ranging from 10^{-7} to 10^{-3} [12]. In the next section, we will further focus on admission control and preamble resource allocation for given outage probability bound.

IV. Admission Control and Resource Allocation for Bounded Outage Probability

In this section, we are interested in characterizing the maximum number of MTDs that can be admitted and the minimum required number of preambles when the outage probability is required to be below a certain bound.

Specifically, for given maximum allowable access delay D_b, to ensure that the outage probability P_{out} does not exceed a given bound, denoted by P_b, the maximum number of MTDs that can be admitted can be written as

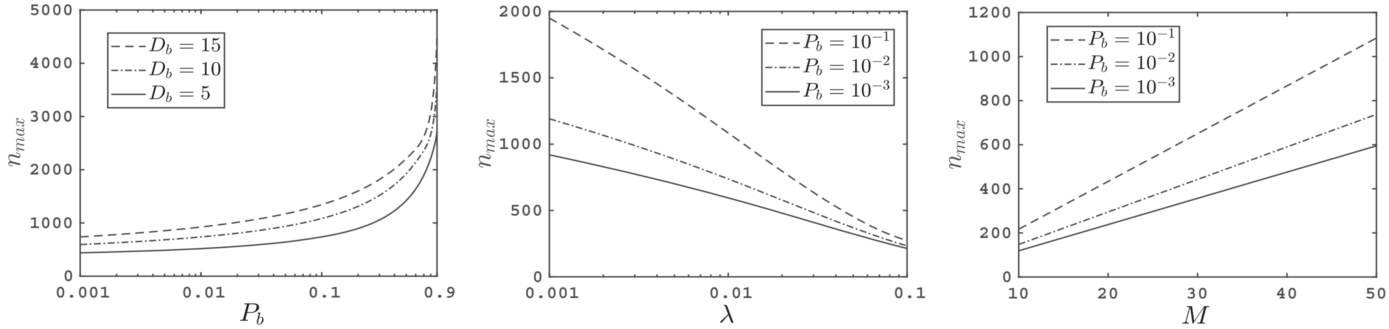
$$n_{max} = \max\{n | P_{out} \leq P_b\}. \tag{10}$$

Since P_{out} increases with the network size n, we can rewrite (10) as $n_{max} = \max\{n | P^*_{out} \leq P_b\}$, where P^*_{out} is the minimum outage probability and has been explicitly characterized in (8). Note that as P^*_{out} is also a monotonic increasing function of n, we have

$$n_{max} = n|_{P^*_{out} = P_b}, \tag{11}$$

that is, the maximum number of MTDs n_{max} should be the root of $P^*_{out} = P_b$, which can be obtained as

$$n_{max} = \begin{cases} \frac{4M}{\lambda}\left(\gamma^* e^{-\gamma^*}\right)^2 & \text{if } 0 \leq P_b \leq \left(1 - \frac{\lambda}{e\hat{\lambda}_0 - 1}\right)^{D_b}, \\ \frac{M}{e\lambda}\left(\frac{\lambda}{1 - \sqrt[D_b]{P_b}} + 1\right) & \text{if } \left(1 - \frac{\lambda}{e\hat{\lambda}_0 - 1}\right)^{D_b} < P_b < 1, \end{cases} \tag{12}$$



(a) (b) (c)

Fig. 3. Maximum number of MTDs n_{max} versus (a) the outage probability bound P_b, (b) the input rate per MTD λ, and (c) the number of preambles M. (a) $\lambda = 0.01$. $M = 50$. $D_b \in \{5, 10, 15\}$ slots. (b) $M = 50$. $D_b = 10$ slots. $P_b \in \{10^{-1}, 10^{-2}, 10^{-3}\}$. (c) $\lambda = 0.01$. $D_b = 10$ slots. $P_b \in \{10^{-1}, 10^{-2}, 10^{-3}\}$.

where γ^* is the root of $\frac{e^{2\gamma} p_L^{q=q_L}}{2\gamma - 1} = \frac{1}{\lambda}\left(1 - \sqrt[D_b]{P_b}\right)$, and $p_L^{q=q_1}$ is the larger non-zero root of the fixed-point equation (9). The detailed derivation of (12) is presented in Appendix A.

Eq. (12) shows that the maximum number of MTDs n_{max} is determined by the maximum allowable access delay D_b, the outage probability bound P_b, the input rate per MTD λ, and the number of preambles M. As Fig. 3 illustrates, n_{max} is a monotonic increasing function of P_b and M, and a monotonic decreasing function of λ. Specifically, Figs. 3a and 3b show that the maximum number of MTDs n_{max} superlinearly increases with the outage probability bound P_b and logarithmically decreases with the input rate per MTD λ, when P_b or λ is small. Fig. 3c shows that n_{max} linearly increases with the number of preambles M. For small M, e.g., $M = 10$ and $P_b = 10^{-3}$, only $n_{max} = 119$ MTDs can be admitted. For a large-scale network where thousands of MTDs may need access, the number of preambles should be allocated properly.

For given outage probability bound, the minimum number of preambles that should be allocated can be written as

$$M_{min} = \min\{M | P_{out} \le P_b\}. \tag{13}$$

Note that as shown in (12), the maximum number of MTDs n_{max} is a linear function of the number of preambles M. Accordingly, if the number of MTDs n is given, the minimum number of preambles M_{min} should be the root of $n_{max} = n$, based on which we can derive M_{min} from (12) as

$$M_{min} = \begin{cases} \frac{n\lambda}{4\left(\gamma^* e^{-\gamma^*}\right)^2} & \text{if } 0 \le P_b \le \left(1 - \frac{\lambda}{e\hat{\lambda}_0 - 1}\right)^{D_b}, \\ \frac{en}{\frac{1}{\lambda} + \frac{1}{1 - \sqrt[D_b]{P_b}}} & \text{if } \left(1 - \frac{\lambda}{e\hat{\lambda}_0 - 1}\right)^{D_b} < P_b < 1, \end{cases} \tag{14}$$

where γ^* is the root of $\frac{e^{2\gamma} p_L^{q=q_L}}{2\gamma - 1} = \frac{1}{\lambda}\left(1 - \sqrt[D_b]{P_b}\right)$, and $p_L^{q=q_1}$ is the larger non-zero root of the fixed-point equation (9).

V. Simulation Results

In this section, simulation results will be presented to verify the above analysis. The simulation setting is the same as the system model described in Section II, and each simulation is carried out for 10^8 time slots. In the simulation, the outage probability of access delay is obtained by calculating the ratio of the number of successful access requests with access delay exceeding D_b to the total number of successful access requests. The maximum allowable access delay $D_b = 10$ time slots, which corresponds to 10 ms as required in URLLC smart factory applications [12][1].

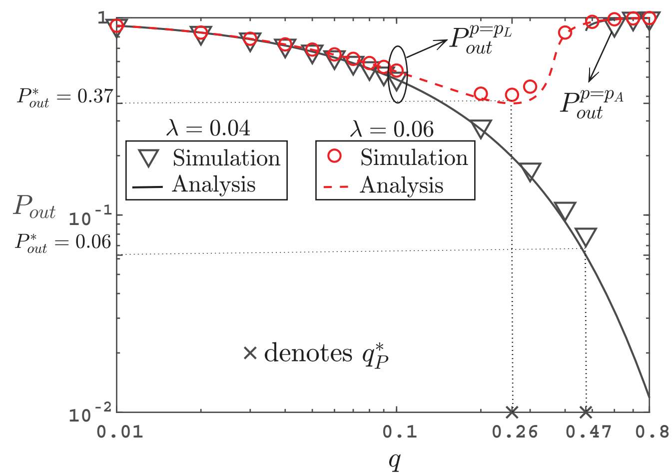


Fig. 4. Outage probability P_{out} versus the ACB factor q. $D_b = 10$ slots. $n = 500$. $M = 50$. $\lambda \in \{0.04, 0.06\}$.

Fig. 4 shows how the outage probability P_{out} varies with the ACB factor q, where the number of MTDs $n = 500$ and the number of preambles $M = 50$. The expressions of minimum outage probability P_{out}^* and the corresponding optimal ACB factor q_P^* have been given in (8) and (7), respectively, as segmented functions of the traffic input rate per MTD λ. Therefore, we consider two cases depending on whether $\lambda \le \frac{M}{n}\hat{\lambda}_0 \approx 0.048$ or not. When $\lambda = 0.06 > \frac{M}{n}\hat{\lambda}_0$, the analysis shows that the outage probability is minimized when the ACB factor q is set as $q_P^* = 0.26$. We can observe from the simulation results that the minimum outage

[1]Note that the length of a time slot is assumed to be 1 ms, as stipulated in the LTE standard [17].

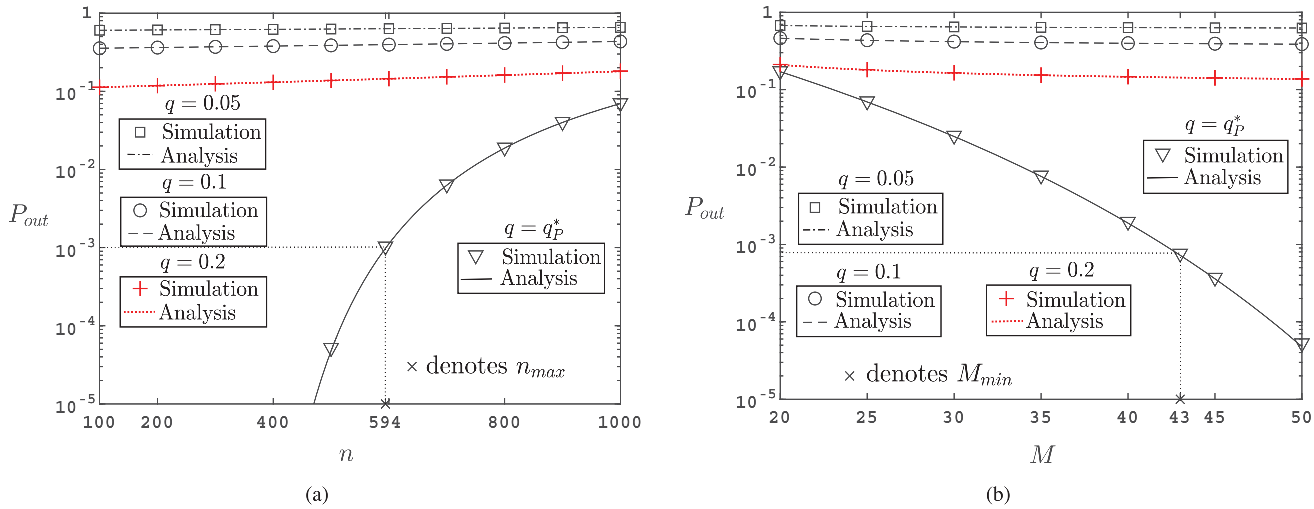


Fig. 5. Outage probability P_{out} versus (a) the number of MTDs n and (b) the number of preambles M. $q \in \{q_P^*, 0.05, 0.1, 0.2\}$. $\lambda = 0.01$. $D_b = 10$ slots. (a) $M = 50$. (b) $n = 500$.

probability $P_{out}^* = 0.37$ is indeed achieved with $q_P^* = 0.26$. On the other hand, when $\lambda = 0.04 < \frac{M}{n}\hat{\lambda}_0$, the optimal ACB factor is calculated as $q_P^* = 0.47$, with which the minimum outage probability $P_{out}^* = 0.06$ is achieved according to (8). The simulation results presented in Fig. 4 well agree with the analysis.

Fig. 5a further shows how the minimum outage probability varies with the number of MTDs n with the ACB factor q optimally tuned, where the number of preambles $M = 50$ and the input rate per MTD $\lambda = 0.01$. For comparison, the outage probabilities with the ACB factor q fixed to 0.05, 0.1, and 0.2 are also presented, which are chosen from the default values of ACB factor as the current standard specifies [18]. It can be seen from Fig. 5a that when the ACB factor is fixed, the outage probability is high even with the number of MTDs as small as 100, with which the outage probability bound $P_b = 10^{-3}$ given in smart factory applications [12] cannot be satisfied. In sharp contrast, significant gains in outage probability can be achieved by optimally tuning the ACB factor as $q = q_P^*$ according to (7), which corroborates that the adaptive tuning of the ACB factor is of great importance for optimizing the access delay performance. It can be obtained from (12) that for given outage probability bound $P_b = 10^{-3}$, the maximum number of MTDs that the system can support is $n_{max} = 594$, which is verified by the simulation results presented in Fig. 5a.

Fig. 5b shows how the minimum outage probability P_{out}^* varies with the number of preambles M, where the number of MTDs $n = 500$ and the input rate per MTD $\lambda = 0.01$. It can be seen from Fig. 5b that the minimum outage probability can be substantially reduced by increasing M. Yet if the ACB factor is fixed, the outage probability remains at a high level regardless of the increase of M. It can be obtained from (14) that for given required outage probability bound $P_b = 10^{-3}$, the minimum number of preambles that should be allocated is $\lceil M_{min} \rceil = 43$, which can be verified by the simulation results presented in Fig. 5b.

VI. Conclusion

In this paper, the outage probability for given maximum allowable access delay of M2M communications in LTE cellular networks is characterized and minimized by optimally tuning the ACB factor. By establishing the relationship between the outage probability of access delay and the mean access delay, it is shown that the outage probability and the mean access delay can be simultaneously minimized, and the minimum outage probability crucially depends on the network size, the traffic input rate of each MTD, and the number of preambles. The maximum number of MTDs that can be admitted for given outage probability bound is further derived, and found to be a linearly increasing function of the number of preambles, based on which the minimum number of preambles that should be allocated for given network size is also obtained. Simulation results corroborate that compared with the standard setting where the ACB factor is fixed, the outage probability of access delay can be substantially reduced by optimally tuning the ACB factor according to the number of MTDs and the traffic input rate of each MTD.

The analysis offers important insights for admission control and resource allocation for supporting M2M communications with URLLC requirements. Specifically, to satisfy the outage probability bound required by the URLLC applications, the analysis reveals the admission limit of MTDs for given preamble resource. When the network size is given, it also indicates how many preambles should be allocated at least. The proposed optimal tuning of ACB factor enables us to support more MTDs or use much less preamble resource when meeting the URLLC requirements on outage probability of access delay in cellular networks.

Appendix A
Derivation of (12)

It has been shown in (11) that the maximum number of MTDs n_{max} can be solved as the root of $P_{out}^* = P_b$, where P_{out}^* is given in (8) as a segmented function of the number of MTDs n depending on $n \leq M\hat{\lambda}_0/\lambda$ or not. Let n_1 and n_2 denote the root of $P_{out}^* = P_b$ for $n \leq M\hat{\lambda}_0/\lambda$ and $n > M\hat{\lambda}_0/\lambda$, respectively.

For $n \leq M\hat{\lambda}_0/\lambda$, we have

$$P_{out}^*|_{n \leq M\hat{\lambda}_0/\lambda} = \left(1 - \frac{4M\mathbb{W}_{-1}^2\left(-\sqrt{n\lambda/M}/2\right)p_L^{q=q_1}}{n\left(-2\mathbb{W}_{-1}\left(-\sqrt{n\lambda/M}/2\right)-1\right)}\right)^{D_b} = P_b. \tag{15}$$

By letting $\gamma = -\mathbb{W}_{-1}\left(-\sqrt{n\lambda/M}/2\right)$, (15) can be written as

$$\frac{e^{2\gamma}p_L^{q=q_L}}{2\gamma-1} = \frac{1}{\lambda}\left(1 - \sqrt[D_b]{P_b}\right), \tag{16}$$

where $p_L^{q=q_1}$ is the larger non-zero root of the fixed-point equation (9). Denote the root of (16) as γ^*, and n_1 can then be written as

$$n_1 = \frac{4M}{\lambda}\left(\gamma^* e^{-\gamma^*}\right)^2. \tag{17}$$

On the other hand, for $n > M\hat{\lambda}_0/\lambda$, we have

$$P_{out}^*|_{n > M\hat{\lambda}_0/\lambda} = \left(1 - \frac{\lambda}{\frac{en\lambda}{M}-1}\right)^{D_b} = P_b, \tag{18}$$

based on which n_2 can be obtained as

$$n_2 = \frac{M}{e\lambda}\left(\frac{\lambda}{1-\sqrt[D_b]{P_b}} + 1\right). \tag{19}$$

It has been shown that the minimum outage probability P_{out}^* is a monotonic increasing function of the number of MTDs n. When $n = M\hat{\lambda}_0/\lambda$, the minimum outage probability can be obtained as $P_{out}^*|_{n=M\hat{\lambda}_0/\lambda} = \left(1 - \frac{\lambda}{e\hat{\lambda}_0 - 1}\right)^{D_b}$. Accordingly, we have $P_{out}^*|_{n \leq M\hat{\lambda}_0/\lambda} \leq \left(1 - \frac{\lambda}{e\hat{\lambda}_0 - 1}\right)^{D_b}$ and $P_{out}^*|_{n > M\hat{\lambda}_0/\lambda} > \left(1 - \frac{\lambda}{e\hat{\lambda}_0 - 1}\right)^{D_b}$. By further combining (17) and (19), we can obtain the expression of n_{max} in (12).

References

[1] H. Chen, R. Abbas, P. Cheng, M. Shirvanimoghaddam, W. Hardjawana, W. Bao, Y. Li, and B. Vucetic, "Ultra-Reliable low latency cellular networks: use cases, challenges and approaches," *IEEE Commun. Mag.*, vol. 56, no. 12, pp. 119–125, Dec. 2018.

[2] *Study on Scenarios and Requirements for Next Generation Access Technologies,* TR 38.913 V0.3.0, 3GPP, Mar. 2016.

[3] 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. Surveys Tuts.*, vol. 16, no. 1, pp. 4–16, 1st Quart., 2014.

[4] F. Ghavimi and 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, 2nd Quart., 2015.

[5] C. She, C. Yang, and T. Q. S. Quek, "Joint uplink and downlink resource configuration for ultra-reliable and low-latency communications," *IEEE Trans. Commun.*, vol. 66, no. 5, pp. 2266–2280, May 2018.

[6] C. She, C. Yang, and T. Q. S. Quek, "Cross-Layer optimization for ultra-reliable and low-latency radio access networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 127–141, Jan. 2018.

[7] A. Anand and G. de Veciana, "Resource allocation and HARQ optimization for URLLC traffic in 5G wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 11, pp. 2411–2421, Nov. 2018.

[8] S. E. Elayoubi, P. Brown, M. Deghel, and A. Galindo-Serrano, "Radio resource allocation and retransmission schemes for URLLC over 5G networks," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 4, pp. 896–904, Apr. 2019.

[9] G. J. Sutton, J. Zeng, R. P. Liu, W. Ni, D. N. Nguyen, B. A. Jayawickrama, X. Huang, M. Abolhasan, Z. Zhang, E. Dutkiewicz, and T. Lv, "Enabling technologies for ultra-reliable and low latency communications: From PHY and MAC layer perspectives," *IEEE Commun. Surveys Tuts.*, vol. 21, no. 3, pp. 2488–2524, 3rd Quart., 2019.

[10] A. Biral, M. Centenaro, A. Zanella, L. Vangelista, and M. Zorzi, "The challenges of M2M massive access in wireless cellular networks," *Digit. Commun. Netw.*, vol. 1, no. 1, pp. 1-19, 2015.

[11] L. Zhao, X. Chi, L. Qian, and W. Chen, "Analysis on latency-bounded reliability for adaptive grant-free access with multipackets reception (MPR) in URLLCs," *IEEE Commun. Lett.*, vol. 23, no. 5, pp. 892–895, May 2019.

[12] M. Centenaro, L. Vangelista, and S. Saur, "Analysis of 5G radio access protocols for uplink URLLC in a connection-less mode," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3104–3117, May 2020.

[13] L. Zhao, X. Chi, and Y. Zhu, "Martingales-Based energy-efficient D-ALOHA algorithms for MTC networks with delay-insensitive/URLLC terminals co-existence," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 1285–1298, Apr. 2018.

[14] W. Zhan and L. Dai, "Massive random access of machine-to-machine communications in LTE networks: Modeling and throughput optimization," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2771–2785, Apr. 2018.

[15] W. Zhan and L. Dai, "Access delay optimization of M2M communications in LTE networks," *IEEE Wireless Commun. Lett.*, vol. 8, no. 6, pp. 1675–1678, Dec. 2019.

[16] *Evolved Universal Terrestrial Radio Access (E-UTRA); Medium Access Control (MAC) protocol specification,* document TS 36.321 V12.5.0, 3GPP, Apr. 2015.

[17] *Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation,* document TS 36.211 V14.9.0, 3GPP, May 2019.

[18] *Evolved Universal Terrestrial Radio Access (E-UTRA); Radio Resource Control (RRC) protocol specification,* document TS 36.331 V13.3.0, 3GPP, Jan. 2017.