2021 Computing, Communications and IoT Applications (ComComAp) | 978-1-6654-2798-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/COMCOMAP53641.2021.9653000

Distributed Delay Optimization of Machine-Type Communications in 5G Networks

Feifan Luo*, Xinghua Sun*, Wen Zhan*, Jun Zou† and Hongzhou Tan‡

*School of Electronics and Communication Engineering, Shenzhen Campus of Sun Yat-sen University
†School of Electronic and Optical Engineering, Nanjing University of Science and Technology
‡School of Electronics and Information Technology, Sun Yat-sen University
luoff6@mail2.sysu.edu.cn, {sunxinghua, zhanw6}@mail.sysu.edu.cn
jun_zou@njust.edu.cn, issthz@mail.sysu.edu.cn

***Abstract*—Optimization of delay performance is crucial to support delay critical applications for Machine-type Communications (MTC) in 5G networks. Centralized approaches to optimize delay performance rely on information exchange between gNB and machine type devices (MTDs), which would lead to unaffordable overhead especially when the number of MTDs becomes large. To address this issue, a novel distributive method is proposed. Specifically, each MTD can estimate the number of MTDs by counting the number of Msg4 on the channel along with their own successful and total transmissions of access requests in each estimation interval and then obtain the optimal Access Class Barring (ACB) factor according to an explicit expression. It is shown that the proposed method can achieve the minimum mean access delay. Moreover, the proposed method can overcome the disparity of the estimated ACB factor caused by the difference of independent observations of each MTD, which can ensure that the delay jitter is optimized simultaneously.**

***Index Terms*—machine-type communications, random access, distributed delay optimization**

I. Introduction

Machine-type communications (MTC), which aims to provide access services for different machine type devices (MTDs) without human intervention, has attracted extensive attention in recent years. In fact, MTC has been regarded as one of three generic service categories for 5G system. For many scenarios, such as industrial automation and real-time monitoring/control, stringent requirements in terms of low latency are imposed [1]. However, the 5G random access (RA) procedure [2] inherits similar problems to that of LTE. In particular, when there are a large number of MTDs in the 5G network, congestion and overload will occur due to concurrent transmission on the shared channel. Severe collisions among MTDs would result in unexpected access delay. Therefore, in order to provide real-time services for the delay critical devices, important issues such as long access delay should be addressed in MTC.

The work was supported in part by the Key-Area Research and Development Program of Guangdong Province under Grant 2020B0101120003, in part by the State's Key Project of Research and Development Plan under Grants 2019YFE0114000, in part by Guangdong Basic and Applied Basic Research Foundation under Grant 2019A1515011906, in part by Industry-University-Research Cooperation Project in Zhuhai (ZH22017001200072PWC), in part by Guangdong Provincial Special Fund For Modern Agriculture Industry Technology Innovation Teams (No. 2021KJ122), in part by National Natural Science Foundation of China under Grant 62001524 and 61701234, and in part by the Science, Technology and Innovation Commission of Shenzhen Municipality under Grant 2021A04.

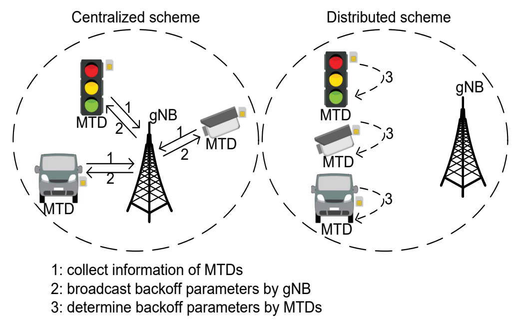

Fig. 1. Centralized optimization schemes and distributed optimization schemes.

There have been a great deal of studies concentrating on delay performance of MTC [3]–[5]. In particular, different schemes have been proposed to track the number of backlogged MTDs in each slot by gNB, based on which schemes to improve the delay performance were developed [3], [4]. In practice, however, the time-varying number of backlogged MTDs is hard to capture. Recently, a framework to optimize access delay performance in MTC was proposed [6], in which the optimal Access Class Barring (ACB) factor was formulated to be dependent on the number of registered MTDs instead of backlogged MTDs. In practice, the gNB can easily obtain the number of registered MTDs in the network and calculate the optimal ACB factor with the feedback from MTDs.

Above studies focused on centralized optimization schemes, i.e., the gNB determines the backoff parameters according to the collected information of the network and broadcasts them to each MTD, as shown in Fig. 1. However, these schemes will lead to the significant signaling overhead and extra access delay especially when short packets are usually transmitted in M2M communications. For example, for a range of vehicular application payload data sizes, the signaling overhead defined as the ratio of the signaling load to the sum of the signaling and traffic loads, is close to 100 percent [7]. Under such circumstance, distributed optimization schemes, a promising technique to reduce signaling overhead efficiently, become highly desirable where each MTD obtains the optimal backoff parameters and tunes its configuration independently.

There have been plenty of distributed schemes based on machine learning or game theory techniques [8], [9]. Without explicit expressions, most of them need iterations, with which might not guarantee the optimal delay performance of the network. Recently, distributed schemes focusing on the throughput performance were proposed in [10], [11]. In [10], each MTD determines the optimal backoff parameters by counting their own numbers of successful and total transmissions of access requests within an estimation interval. Nevertheless, this scheme will cause the ACB factor to differ among MTDs since the observation of each MTD is different, which may lead to deteriorated delay jitter. In fact, it is argued that the delay requirement in 5G system is not simply a pursuit of low delay but also the pursuit of deterministic delay. In some scenarios, such as remote surgery and driving, a large delay jitter is not allowed. In [11], the distributed scheme for IEEE 802.11 DCF network can overcome this issue by counting the number of busy intervals and ACK frames on the channel. Unfortunately, it cannot be applied to MTC since MTDs, unlike the nodes in IEEE 802.11 DCF network, usually do not have sensing capability.

In this paper, by leveraging explicit expressions of the mean access delay and the delay jitter (the second moment access delay) in [6], a distributed algorithm is proposed, with which each MTD can determine the optimal ACB factor independently to minimize the mean access delay and the delay jitter simultaneously. Specifically, by observing the number of Msg4 on the channel along with their own successful and total transmissions of access requests during an estimation interval, each MTD can obtain an estimation of the number of MTDs and then the optimal ACB factor according to an explicit expression of the observed statistics. Compared to the distributed schemes based on machine learning or game theory techniques, the computational complexity of the proposed algorithm is extremely low due to no iterations and therefore, no extra delay is introduced. In order to overcome the disparity of the estimated optimal ACB factor across MTDs, a threshold of the estimated network throughout referred to as the cutoff threshold is introduced. By this method, each MTD can have an identical estimated steady-state operating point, which can ensure that the delay jitter is optimized. It is validated by extensive simulations that when the estimation interval and the cutoff threshold are carefully tuned, the mean access delay and the delay jitter can be minimized simultaneously for a wide range of the network size.

II. System Model and Preliminary Analysis

Consider a 5G network where n MTDs transmit to a gNB over a common channel. Each MTD has Bernoulli[1] packet arrivals with rate $\lambda \in (0, 1)$. Besides, each MTD is equipped of a data buffer with infinite size. Once there are packets to be transmitted in its data buffer, the access request will be generated.

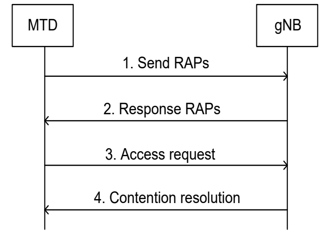

Fig. 2. Four steps of the RA procedure in 5G networks.

Note that the RA procedure in current 5G standard consists of four steps [2], as shown in Fig. 2. In Step 1, each MTD chooses one from M random access preambles (RAPs) to send to the gNB through the Physical Random Access CHannel (PRACHs). The time slot is defined as the interval between two consecutive PRACHs. In Step 2, the gNB will detect the preambles and then broadcast a random access response (RAR) for MTDs whose preambles are successfully received. In Step 3, those MTDs will transmit their access requests. If no more than one MTD select the same RAP, the access request will be successfully responded and all the packets in the data buffer will be delivered immediately. Otherwise, collision occurs, and the RA procedure will be restarted later. In Step 4, the contention resolution message, i.e., Msg4, will be sent to the MTDs whose transmitted messages are correctly decoded by the gNB.

The delay performance of MTC in terms of the mean access delay[2] $E[D_T]$ (in unit of time slots) and the delay jitter $E[D_T^2]$ has been analyzed in [6] and will be summarized in the following, based on which the distributed scheme will be proposed.

It was shown in [6] that when the uniform backoff (UB) window size $W = 1$, the probability of successful transmission p, which is referred to as the steady-state operating point, can be determined by the following fixed-point equation

$$p = \exp\left(-\frac{\hat{\lambda}/M}{p + \hat{\lambda}/(nq)}\right), \tag{1}$$

where $\hat{\lambda}$ denotes the aggregate input rate of the network, i.e., $\hat{\lambda} = n\lambda$ and q denotes the ACB factor. The network throughput, which is defined as the average number of successful transmissions per time slot, has been obtained in [6] as

$$\hat{\lambda}_{out} = \frac{\hat{\lambda}}{\frac{\hat{\lambda}}{n} \cdot \frac{1}{qp} + 1}. \tag{2}$$

For the delay performance, it was indicated that both the mean access delay and the delay jitter can be optimized

[1]Except for Bernoulli arrivals, many MTC nodes typically have packets with a fairly regular frequency. To this end, we have verified through additional experiments that the delay performance of periodic arrivals with a period of $T = 1/\lambda$ well agrees with the theoretical delay performance of Bernoulli arrivals [6] with rate λ. This is to say that the proposed scheme is also applicable for periodic arrivals.

[2]The access delay D_T is defined as the time consumed from the generation of an access request until its successful transmission.

simultaneously. To minimize mean access delay and the delay jitter, the ACB factor needs to be carefully tuned and its optimal value was obtained as

$$q_D^* = \begin{cases} \frac{\lambda}{\frac{n\lambda}{M} - e^{-1}} & \text{if } \lambda > \frac{M}{n}\hat{\lambda}_0, \\ \frac{4M\mathbb{W}_{-1}^2(-\sqrt{n\lambda/M}/2)}{n(-2\mathbb{W}_{-1}(-\sqrt{n\lambda/M}/2)-1)} & \text{otherwise}, \end{cases} \tag{3}$$

where $\hat{\lambda}_0 \approx 0.48$ is the single non-zero root of the equation $\hat{\lambda} - \hat{\lambda}(1 + 1/\mathbb{W}_{-1}(-\sqrt{\hat{\lambda}}/2))^2 = 4(\hat{\lambda} - e^{-1})$.

It can be seen from (3) that one should acquire the information of the number of MTDs n, the packet arrival rate λ of each MTD and the number of RAPs M to obtain the optimal ACB factor. Obviously, a centralized optimization method can be implemented by collecting these information and broadcasting the system information block carrying the optimal ACB factor through the gNB, which would nevertheless lead to substantial overhead [10] and extra delay.

III. Distributed Delay Optimization

In this section, we propose a distributed optimization scheme to minimize both the mean access delay and the delay jitter of MTC in 5G networks based on the results shown in section II.

Each MTD can easily know its packet arrival rate λ and the total number of RAPs M. The remaining problem lies in how each MTD estimates the number of MTDs n in the network in a distributive fashion. To provide support for our distributed algorithm, we first present two formulas in the following.

According to (1), the number of MTDs n can be given by

$$n = -\frac{M}{\lambda}\left(p + \frac{\lambda}{q}\right)\ln p. \tag{4}$$

By combining (2) and (4), the network throughput can be obtained as

$$\hat{\lambda}_{out} = -Mp\ln p. \tag{5}$$

According to (5), we have the network maximum throughput $\hat{\lambda}_{\max} = Me^{-1}$, which is achieved when the steady-state operating point $p = e^{-1}$. The above two equations are of importance for proposing distributed optimization scheme. In fact, (4) relates the number of MTDs n to the steady-state operating point p that can be measured run-time by each MTD. As for (5), it facilitates to estimate the number of MTDs of the network.

Specifically, consider a homogeneous network where each MTD is aware of its packet arrival rate λ and the total number of RAPs M. It's obvious that in order to obtain the number of MTDs n in a distributed manner, the network steady-state operating point p should be evaluated independently by each MTD. One direct scheme is to let each MTD record its own numbers of successful transmissions $n_s(T)$ and access attempts $n_t(T)$ during a certain period of time T. Then the network steady-state operating point p can be estimated as the ratio of these two quantities, i.e,

$$\overline{p} = \frac{n_s(T)}{n_t(T)}. \tag{6}$$

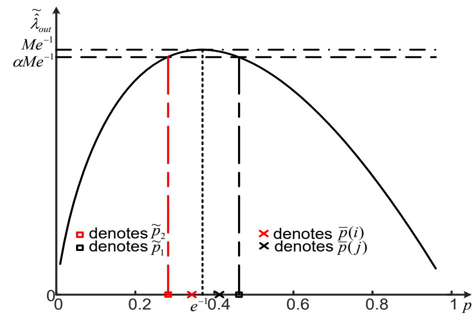

Fig. 3. Curve of (8). $\overline{p}(i)$ and $\overline{p}(j)$ are the network steady-state operating point estimated by MTDs i and j. $\tilde{p}_1$ and $\tilde{p}_2$ are two roots of (8).

By combining (3), (4) and (6), both the number of MTDs and the optimal ACB factor can be calculated. Nevertheless, this scheme will cause the estimated steady-state operating point and the optimal ACB factor to differ among MTDs, which leads to varied access delay and consequently large delay jitter.

To overcome this disparity, we let each MTD count the number of Msg4 $s(T)$ broadcast by the gNB on the channel during an estimation interval T. Then the network throughput can be estimated as

$$\tilde{\hat{\lambda}}_{out} = \min\left\{\frac{s(T)}{T}, Me^{-1}\right\}, \tag{7}$$

where the superscript ˜ denotes the estimation value. Note that the estimated throughput should not exceed the theoretical maximum network throughput, i.e., $\hat{\lambda}_{\max} = Me^{-1}$. By combining (5) and (7), the steady-state operating point can be estimated by the following equation

$$\tilde{\hat{\lambda}}_{out} = -M\tilde{p}\ln\tilde{p}, \quad \tilde{p} \in (0,1). \tag{8}$$

Fig. 3 presents how the estimated throughput $\tilde{\hat{\lambda}}_{out}$ varies with $\tilde{p}$. It is clear that for an estimated network throughput $\tilde{\hat{\lambda}}_{out}$, (8) has two roots, i.e., $\tilde{p}_1 = \exp\left\{\mathbb{W}_0(-\frac{1}{M}\min\{\frac{s(T)}{T}, Me^{-1}\})\right\}$ and $\tilde{p}_2 = \exp\left\{\mathbb{W}_{-1}(-\frac{1}{M}\min\{\frac{s(T)}{T}, Me^{-1}\})\right\}$. Only one of them is the steady-state operating point of the network. In order to determine which one, we propose to let each MTD record its own numbers of successfully transmitted packets and access attempts during a certain period of time T to assist to make its own decision. Naturally, each MTD should choose the one that is closer to $\overline{p}$, i.e.,

$$\tilde{p} = \arg\min_{\tilde{p}_1, \tilde{p}_2}\{|\tilde{p}_1 - \overline{p}|, |\tilde{p}_2 - \overline{p}|\}. \tag{9}$$

In practice, $\overline{p}$ is random variable with statistical errors that could be different across MTDs since it is only determined by their own observation. As shown in Fig. 3, the gap between $\tilde{p}_1$ and $\tilde{p}_2$ is determined by the estimated throughput $\tilde{\hat{\lambda}}_{out}$. When $\tilde{\hat{\lambda}}_{out}$ approaches $\hat{\lambda}_{\max} = Me^{-1}$, $\tilde{p}_1$ and $\tilde{p}_2$ are close to each other. In this case, as each MTD has a different $\overline{p}$, it may choose a different steady-state operating point. To prevent such disparity, we propose that when the gap between $\tilde{p}_1$ and $\tilde{p}_2$ is small, let each MTD set the estimated throughput to

$$\tilde{q}_D^* = \begin{cases} \dfrac{\lambda}{\frac{\left[\tilde{\hat{\lambda}}_{out}(\frac{1}{\lambda}+\frac{1}{q\tilde{p}})\right]\lambda}{M} - e^{-1}} & \text{if } \lambda > \dfrac{M}{\left[\tilde{\hat{\lambda}}_{out}(\frac{1}{\lambda}+\frac{1}{q\tilde{p}})\right]}\hat{\lambda}_0, \\ \dfrac{4M\mathbb{W}_{-1}^2(-\sqrt{\left[\tilde{\hat{\lambda}}_{out}(\frac{1}{\lambda}+\frac{1}{q\tilde{p}})\right]\lambda/M}/2)}{\left[\tilde{\hat{\lambda}}_{out}(\frac{1}{\lambda}+\frac{1}{q\tilde{p}})\right](-2\mathbb{W}_{-1}(-\sqrt{\left[\tilde{\hat{\lambda}}_{out}(\frac{1}{\lambda}+\frac{1}{q\tilde{p}})\right]\lambda/M}/2)-1)} & \text{otherwise .} \end{cases} \tag{12}$$

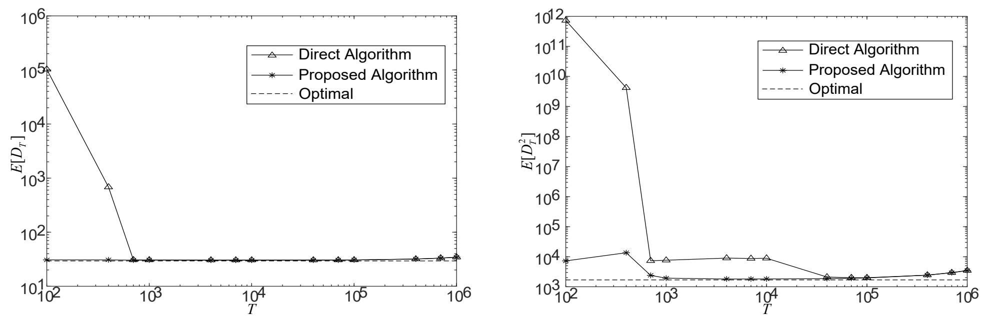

Fig. 4. The network delay performance comparison between direct algorithm and proposed algorithm. $M = 10$, $n = 200$, $W = 1$, $\lambda = 0.04$, $\alpha = 0.95$. (a) The mean access delay optimized $E[D_T]$ versus the estimation interval T. (b) The delay jitter optimized $E[D_T^2]$ versus the estimation interval T.

be $\tilde{\hat{\lambda}}_{out} = Me^{-1}$, i.e., each MTD has an identical estimated steady-state operating $\tilde{p} = e^{-1}$. To this end, we introduce an adjustable parameter $\alpha \leq 1$ as a threshold for judging whether the gap between $\tilde{p}_1$ and $\tilde{p}_2$ is small or not. Here we refer to α as the cutoff threshold. If the estimated throughput $\tilde{\hat{\lambda}}_{out} > \alpha Me^{-1}$, $\tilde{p}_1$ and $\tilde{p}_2$ is regarded to be close to each other. In this case, each MTD needs to modify the estimated throughput as $\tilde{\hat{\lambda}}_{out} = Me^{-1}$. Otherwise, when $\tilde{\hat{\lambda}}_{out} \leq \alpha Me^{-1}$, there is no disparity since $\tilde{p}_1$ and $\tilde{p}_2$ is far away from each other, in which case the estimated throughput will not be modified.

Overall, for each MTD, the estimated throughput in (7) can be rewritten as

$$\tilde{\hat{\lambda}}_{out} = \begin{cases} Me^{-1} & \text{if } \frac{s(T)}{T} > \alpha Me^{-1}, \\ \frac{s(T)}{T} & \text{otherwise .} \end{cases} \tag{10}$$

Note that this treatment may lead to a rounding error, yet it is marginal since both $\tilde{p}_1$ and $\tilde{p}_2$ are close to e^{-1} when α approaches 1.

Algorithm 1 Distributed Delay Optimization of MTC in 5G networks.

Input: λ: packet arrival rate; T: evaluation cycle; q: initial ACB factor; M: the total number of RAPs; α: cutoff threshold;

Output: $\tilde{q}_D^*$: estimated optimal ACB factor;

1: **for** Each MTD in the network after every interval T **do**
2: Record $n_s(T)$, $n_t(T)$ and $s(T)$;
3: Calculate $\tilde{\hat{\lambda}}_{out}$ according to (10);
4: Calculate $\tilde{p}$ according to (6), (8) and (9);
5: Calculate $\tilde{q}_D^*$ according to (12);
6: Update $q \leftarrow \tilde{q}_D^*$;
7: **end for**

Finally, the number of MTDs $\tilde{n}$ can be estimated as

$$\tilde{n} = \left[\tilde{\hat{\lambda}}_{out}(\frac{1}{\lambda} + \frac{1}{q\tilde{p}})\right], \tag{11}$$

where $\tilde{p}$ can be estimated by (8), (9) and (10), and $[\cdot]$ is the rounding operation. By substituting (11) into (3), each MTD can obtain their own optimal ACB factor $\tilde{q}_D^*$ presented in (12). By repeating above operations during each estimation interval T, the ACB factor of each MTD can be updated periodically to adapt to changes of the network.

The Algorithm 1 summarizes the approach to estimate the optimal ACB factor of MTC in 5G networks distributively. The accuracy of Algorithm 1 will be verified to be better than the direct algorithm in the next section. In fact, the accuracy of the optimal ACB factor obtained by the proposed algorithm is at the cost of more operations. However, the computational complexity of the proposed algorithm is still extremely low since we have explicit expressions for optimal configuration and no iterative operation is needed.

IV. SIMULATION RESULTS AND DISCUSSIONS

In this section, we will verify the effectiveness of the proposed algorithm by simulations. The simulation setting is in accordance with the assumptions presented in Section II. Each simulation run lasts for a fixed duration, namely 10^7 time slots and the initial ACB factor q_{init} is set to be $q_{init} = 0.2$.

The network delay performance comparison between the direct algorithm and proposed algorithm is presented in Fig. 4. In the direct algorithm, each MTD determines the optimal ACB factor by combining (3), (4) and (6). It is shown in Fig. 4 that both the mean access delay $E[D_T]$ and the delay jitter $E[D_T^2]$ of the direct algorithm deviate from the optimal value when the estimation interval T is small. In contrast, the mean access delay $E[D_T]$ of the proposed algorithm is close to the optimal value regardless of the value of the

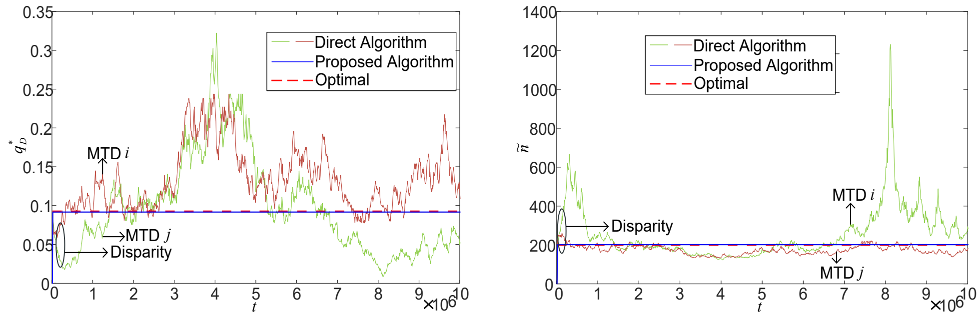

Fig. 5. Estimation performance of both the direct algorithm and proposed algorithm. $M = 10$, $n = 200$, $W = 1$, $\lambda = 0.04$, $\alpha = 0.95$, $T = 10^4$. (a) The evaluated optimal ACB factor $\tilde{q}_D^*$ of all MTDs versus time t. (b) The evaluated number of MTDs $\tilde{n}$ of all MTDs versus time t.

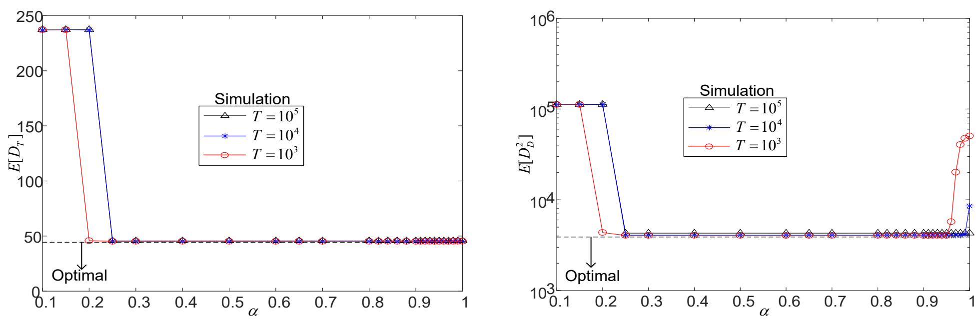

Fig. 6. The network delay performance versus the cutoff threshold α. $M = 10$, $n = 200$, $W = 1$, $\lambda = 0.1$, $T \in \{10^3, 10^4, 10^5\}$. (a) Mean access delay $E[D_T]$ versus the cutoff threshold α. (b) The delay jitter $E[D_T^2]$ versus the cutoff threshold α.

estimation interval T. The delay jitter $E[D_T^2]$ of the proposed algorithm, nevertheless, is sensitive to T. As T increases, the delay jitter $E[D_T^2]$ first approaches and then deviates from its optimal value. Intuitively, a larger estimation interval T can make the estimation more accurate. On the other hand, with a too large estimation interval T, although the estimation is accurate, the ACB factor q is not adjusted in time during the long first round estimation interval. Therefore, the estimation interval T should be selected appropriately to achieve both the minimum of mean access delay $E[D_T]$ and delay jitter $E[D_T^2]$ simultaneously.

To see why the direct algorithm does not work well, Fig. 5 further shows the optimal ACB factor and the number of MTDs estimated by both the direct algorithm and the proposed algorithm. Here we present curves of the optimal ACB factor and the number of MTDs estimated by two random MTDs i and j over time for the direct algorithm. It can be seen that the proposed algorithm can estimate the number of MTDs n and the optimal ACB factor q_D^* accurately, and overcome the disparity well. In contrast, the number of MTDs n and the ACB factor q_D^* estimated by the direct algorithm is inaccurate and highly differentiated across MTDs. The disparity of the ACB factor among MTDs would enlarge the estimation error in the optimal ACB factor in the following estimation rounds, which has a great influence on the delay jitter $E[D_T^2]$. To sum up, the proposed algorithm is more robust and performs better in terms of delay performance than the direct algorithm.

To take a close look into how to choose the cutoff threshold α, Fig. 6 illustrates how the cutoff threshold α affects the delay performance. It is shown that the mean access delay $E[D_T]$ and the delay jitter $E[D_T^2]$ can achieve their optimal values simultaneously in a wide range of α, indicating robustness of the proposed algorithm. When α is too small, the delay performance is poor since it is easy to cause the MTDs to erroneously adjust the estimated network throughput as Me^{-1} according to (10). Here we suggest to set a large q_{init}, such as $q_{init} = 0.2$, which not only can make each MTD have more samples to estimate the optimal ACB factor in the first estimation interval, but also can enlarge the range of α with which the optimal delay performance can be achieved.[3] On the other hand, when α is too large, the delay jitter $E[D_T^2]$ is far away from its optimal value especially when the estimation interval $T = 10^3$ since a large α can not overcome the disparity caused by insufficient samples with a small T. To sum up, the cutoff threshold α should be selected appropriately, not only to alleviate disparity as much as possible, but also to control the estimation error within a tolerate range.

[3]With a large initial ACB factor q_{init}, the estimated throughput $\tilde{\tilde{\lambda}}_{out}$ in the first estimation interval is small. In this case, even with a small cutoff threshold α, proposed algorithm can estimate network throughput accurately in the first estimation interval since in this case, we have $\alpha Me^{-1} > \tilde{\tilde{\lambda}}_{out}$. And then the ACB factor is adjusted to its optimal value q_D^*, after which the network operates with the optimal delay performance. Therefore, in the following estimation rounds, the estimated throughput must be larger than αMe^{-1} (since α is small) and is modified as Me^{-1} which equals to the actual throughput, with which the network would be stabilized at the optimal delay performance.

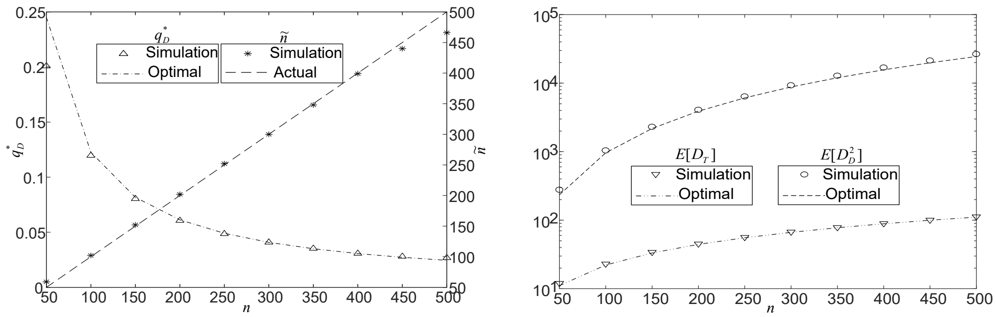

Fig. 7. The performance of proposed algorithm versus the number of MTDs n. $M = 10$, $W = 1$, $\hat{\lambda} = 20$, $T = 10^4$, $\alpha = 0.95$. (a) The evaluated optimal ACB factor $\tilde{q}_D^*$ and the evaluated number of MTDs $\tilde{n}$ versus the number of MTDs n. (b) The delay performance versus the number of MTDs n.

Fig. 7 further shows how the performance of proposed algorithm varies with the number of MTDs n in the network. According to previous simulations, we set the estimation interval $T = 10^4$ and the cutoff threshold $\alpha = 0.95$. It can be observed from Fig. 7 that with our proposed algorithm, both the number of MTDs n and the optimal ACB factor q are estimated accurately. Moreover, the mean access delay $E[D_T]$ and the delay jitter $E[D_T^2]$ can be optimized simultaneously regardless of the number of MTDs n in the network. It indicates this configuration of the estimation interval T and the cutoff threshold α can be applied for a wide range of the number of MTDs n.

V. Conclusion

In this paper, we propose a novel distributed delay optimization algorithm for MTC in 5G networks. It is found that the achievability of the minimum of mean access delay is not sensitive to the estimation interval and the cutoff threshold. However, both of them have a crucial impact on the disparity of the estimated ACB factor across MTDs, due to which the delay jitter will be affected greatly. To minimize the mean access delay and delay jitter simultaneously, both the estimation interval and the cutoff threshold should be carefully tuned. The proposed method focuses on delay performance in the homogeneous scenarios. How to satisfy delay requirements in the heterogeneous scenarios where each MTD adopts a distinct set of access parameters distributively is of great practical significance, which deserves future study.

References

[1] M. H. Kafi, P. Mitchell, and D. Grace, "Prioritised dynamic RACH (PD-RACH) scheme for delay-critical MTC communication," Proc. IEEE 31th Int. Symp. Pers. Indoor Mob. Radio Commun., 2020, pp. 1-6.

[2] 3rd Generation Partnership Project (3GPP), "NR; Medium Access Control (MAC) protocol specification," 3GPP, TS 38.321, v16.4.0, 03 2021.

[3] S. Duan, V. Shah-Mansouri, and V. W. S. Wong, "Dynamic access class barring for M2M communications in LTE networks," Proc. IEEE Globecom, 2013, pp. 4747-4752.

[4] C. Y. Oh, D. Hwang, and T. J. Lee, "Joint access control and resource allocation for concurrent and massive access of M2M devices," IEEE Trans. Wireless Commun., vol. 14, no. 8, pp. 4182-4192, Mar. 2015.

[5] M. Koseoglu, "Lower bounds on the LTE-A average random access delay under massive M2M arrivals," IEEE Trans. Commun., vol. 64, no. 5, pp. 2104-2115, 2016.

[6] 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, 2019.

[7] Z. Dawy, W. Saad, A. Ghosh, J. G. Andrews and E. Yaacoub, "Toward Massive Machine Type Cellular Communications," IEEE Wireless Commun., vol. 24, no. 1, pp. 120-128, February 2017.

[8] C. Chang and R. Y. Chang, "Design and analysis of multichannel slotted Aloha for machine-to-machine communication," Proc. IEEE Globecom, 2015, pp. 1-6.

[9] M. N. Soorki, W. Saad, M. H. Manshaei, and H. Saidi, "Stochastic coalitional games for cooperative random access in M2M communications," IEEE Trans. Wireless Commun., vol. 16, no. 9, pp. 6179-6192, 2017.

[10] C. Zhang, X. Sun, J. Zhang, X. Wang, S. Jin, and H. Zhu, "Distributive throughput optimization for massive random access of M2M communications in LTE networks," IEEE Trans. Veh. Technol., vol. 69, no. 10, pp. 11828-11840, 2020.

[11] X. Sun and Y. Gao, "Distributed throughput optimization for heterogeneous IEEE 802.11 DCF networks," Wireless Netw., vol. 24, pp. 1-11, 2018.