Throughput-Oriented Non-Orthogonal Random Access Scheme for Massive MTC Networks

Yichen Wang, *Member, IEEE*, Tao Wang, Zihuan Yang, Dawei Wang, *Member, IEEE*, and Julian Cheng, *Senior Member, IEEE*

***Abstract*— Machine-type communications (MTC) technology, which enables direct communications among devices, plays an important role in realizing Internet-of-Things. However, a large number of MTC devices can cause severe collisions. As a result, the network throughput is decreased and the access delay is increased. To address this issue, a throughput-oriented non-orthogonal random access (NORA) scheme is proposed for massive machine-type communications (mMTC) networks. Specifically, by employing the technique of tagged preambles (PAs), multiple MTC devices (MTCDs) choosing the same PA can be distinguished and regarded as a non-orthogonal multiple access (NOMA) group, which enables multiple MTCDs to share the same physical uplink shared channel for transmissions by multiplexing in the power domain. The Sukhatme's classic theory and the characteristic function approach are adopted to formulate an optimization problem. The aim is to maximize the throughput subject to the constraints on the power back-off factor, the number of MTCDs included in a NOMA group, and the successful transmission probability. Based on the particle swarm optimization (PSO) algorithm, the formulated optimization problem is efficiently solved. The derived solution can be used to adjust the access class barring factor such that more MTCDs can obtain the access opportunities. Moreover, a low-complexity suboptimal solution is also developed, which can achieve near-PSO performance under high data rate requirement. Simulation results show that the proposed scheme can efficiently improve the network performance and comparison is made with the existing schemes.**

***Index Terms*— Massive machine-type communications (MTC) networks, non-orthogonal random access, power back-off.**

Manuscript received June 17, 2019; revised October 24, 2019; accepted November 21, 2019. Date of publication December 5, 2019; date of current version March 18, 2020. This work was supported in part by the National Natural Science Foundation of China under Grant 61871314 and 61901379, in part by the Key Research and Development Program of Shaanxi Province under Grant 2019ZDLGY07-04, in part by the National Science and Technology Major Project under Grant 2018ZX03001003-004, in part by the Natural Science Basic Research Plan in Shaanxi Province of China under Grant 2019JQ-253, and in part by the open research fund of National Mobile Communications Research Laboratory, Southeast University under Grant 2020D04. The associate editor coordinating the review of this article and approving it for publication was Y. Li. *(Corresponding authors: Yichen Wang; Dawei Wang.)*

Y. Wang and T. Wang are with the School of Information and Communications Engineering, Xi'an Jiaotong University, Xi'an 710049, China (e-mail: wangyichen0819@mail.xjtu.edu.cn; wangtaot@stu.xjtu.edu.cn).

Z. Yang was with the School of Information and Communications Engineering, Xi'an Jiaotong University, Xi'an 710049, China. She is now with ZTE Corporation, Xi'an 710114, China (e-mail: huan1994@stu.xjtu.edu.cn).

D. Wang is with the Department of Communication Engineering, Northwestern Polytechnical University, Xi'an 710072, China, and also with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China (e-mail: wangdw@nwpu.edu.cn).

J. Cheng is with the School of Engineering, The University of British Columbia, Kelowna, BC V1V1V7, Canada (e-mail: julian.cheng@ubc.ca).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCOMM.2019.2957767

I. Introduction

With the development of the Internet-of-Things, machine-type communications (MTC) has been widely accepted as an efficient approach for future wireless communication systems. As the MTC technology enables machine-type devices exchange data with little human intervention [1], it has a wide range of applications and thus has attracted much research attention from both industry and academia [2]–[4].

In MTC networks, the widely used random access (RA) procedure is the four-message handshake process [5]. Specifically, the involved four messages include preamble (PA), random access response (RAR), initial layer-3 message, and contention resolution (CR) message. As the number of available PAs is usually much less than the number of active MTC devices (MTCDs), each PA will be selected by multiple MTCDs with high probability, which will likely cause large amounts of PA collisions. Moreover, the PA collisions can only be detected by the evolved node base station (eNB) in the third message handshake, i.e., the initial layer 3 message. Consequently, the rapid growth in the number of MTCDs will incur severe collisions and large signaling overheads, thus resulting in network congestion, large access delay, and the degradation of resource utilization efficiency. Because the RA scheme plays a critically important role in massive MTC (mMTC) networks, it should be carefully designed and a number of works have been dedicated to this important research topic [6]–[8].

Access class barring (ACB) mechanism, which can uniformly redistribute MTCDs in the time domain, is an efficient method to alleviate PA collisions [1]. In the ACB mechanism, the ACB factor is the most important parameter and is dynamically adjusted according to the number of active MTCDs in each RA slot and the number of available PAs. As the ACB factor can control the number of MTCDs that initiate the random access requests, the access success probability can be improved and the network congestion can be alleviated. To achieve this goal, we should carefully determine the value of the ACB factor [9]–[12]. On the one hand, a larger value of the ACB factor will allow more MTCDs to enter the RA procedure and thus cannot efficiently reduce the PA collisions. On the other hand, a smaller value of the ACB factor will lower resource utilization efficiency and increase access delay. To adapt the value of the ACB factor to the traffic loads and available PAs of the MTC network, many works have been conducted to estimate the number of active MTCDs that initiate the random access requests in the current RA slot [13]–[15]. Moreover, in order to improve further the

overall network throughput and decrease the total service time, several advanced ACB mechanisms have also been developed. In [16], the authors proposed a cooperative ACB mechanism to balance the load among multiple based stations. A pricing based ACB load control scheme was proposed [17] to provide service differentiation. The authors in [18] developed a two-phase ACB based protocol and the number of MTCDs that successfully access the network can be efficiently enhanced.

Limited wireless resource is another critical factor that affects the RA performance. Thus, resource allocation is also an important research topic. A number of resource allocation schemes have been developed to divide resources between the four-message handshake procedure and the data transmission phase [19]–[22]. In [23], a two-stage resource allocation scheme was proposed, and it allows MTCDs that fail to pass the ACB check in the first stage to contend for the remaining resources and thus improve the resource utilization efficiency. Moreover, since different MTCDs usually have diverse Quality-of-Service (QoS) requirements, such as latency, outage probability, and reliability, grouping multiple MTCDs having similar QoS requirements into one cluster is beneficial to increase the successful access probability. Correspondingly, cluster-based resource allocation schemes for mMTC network were also studied [24]–[27]. In addition, while using the traditional four-message handshake RA procedure, large amounts of PA collisions will occur in the first handshake and these collisions can only be detected in the third handshake, which not only waste the limited resource, but also increase the access delay. To improve further the resource utilization efficiency, different collision detection and solving methods are proposed, including the collision-aware resource access scheme [28], the splitting trees based collision resolution algorithm [29], and the distributed queueing random access mechanism [30]. By integrating the PAs with tag Zadoff-Chu (ZC) sequences generated by different root numbers, devices can transmit tagged PAs in the first step of RA procedure, which enables eNB to distinguish different MTCDs choosing the same PA. Thus, early PA collision detection can be realized and access delay can be reduced [31].

Based on the above discussions, we can find that the ACB-based RA schemes disperse the access requests of massive MTCDs to reduce the PA collision probability. The resource allocation based RA mechanisms mainly improve the resource utilization efficiency. However, for a large number of MTCDs, these methods cannot satisfy the performance requirements of future mMTC networks. As the successive interference cancelation (SIC) technique enables multiple MTCDs that choose the identical PA to have the opportunity for successful transmissions, non-orthogonal multiple access (NOMA) is a promising approach to enhance the throughput of the mMTC network [32]–[35]. A power domain multiplexing based uplink NOMA scheme was proposed in [37] by adopting the power back-off. Based on the SIC technique, the authors in [41] proposed a non-orthogonal RA (NORA) procedure and developed an analytical model to analyze the system performance. Although the NORA method has shown the potentials to improve the mMTC network performance, several crucial issues, such as assignment of multiple MTCDs to one particular PA and efficient utilization of the power domain resources, need to be addressed to improve further the network performance.

To address the above mentioned issues, we propose a throughput-oriented NORA scheme for the mMTC networks. Specifically, the tagged PA technique is adopted such that the base station is enabled to detect the MTCDs that choose the same PA and thus the corresponding MTCDs choosing the identical PA can form a NOMA group. Then, the power domain resource is allowed to be multiplexed by the MTCDs in each NOMA group through the power back-off. As all MTCDs in each NOMA group employ different transmit powers to share the same resource for data transmission, SIC can be used by the base station for data reception. Based on the developed NORA procedures, we derive an analytical expression of the average throughput that can be achieved by one NOMA group by using the Sukhatme's classic theory and a characteristic function approach. Then we formulate the throughput maximization problem subject to the constraints on power back-off factor, the number of MTCDs in each NOMA group, and the successful transmission probability. Since the objective function is analytically intractable, we efficiently solve the throughput maximization problem by adopting the particle swarm optimization (PSO) algorithm. Based on the derived number of MTCDs in a NOMA group, the ACB factor is correspondingly adjusted such that more MTCDs can pass the ACB check and gain access opportunities. Moreover, by considering the high data rate requirement, we convert our formulated optimization problem to a simpler expression and obtain the low-complexity suboptimal solution, which can efficiently reduce the computation complexity and achieve near-PSO performance under the high data rate requirement. The main contributions of this paper are summarized as follows.

1) By employing the tagged PA technique, multiple MTCDs that choose the same PA can be distinguished and thus can form a NOMA group to share the same resource for transmissions through the power domain multiplexing.
2) By adopting the Sukhatme's classic theory and a characteristic function approach, closed-form expression of the average throughput for each NOMA group is obtained. Then, a throughput maximization problem is formulated and the optimized number of MTCDs included in a NOMA group and the power back-off factor are derived.
3) The analysis under high data rate requirement is conducted. Based on the analysis, the low-complexity suboptimal solution is obtained, which can efficiently reduce the computation complexity and achieve near-PSO performance under the high data rate requirement.
4) Based on the developed NORA schemes and the optimized optimized number of MTCDs included in a NOMA group, the ACB factor is jointly adjusted such that the potentials of the proposed schemes on supporting massive MTCDs can be efficiently exhibited.

The rest of this paper is organized as follows. Section II presents the system model. In Section III, we first describe

the proposed NORA scheme, then formulate and solve the optimization problem. In Section IV, the analysis under high data rate requirement is conducted and the low-complexity solution is developed. Simulation results are provided in Section V and the paper is concluded in Section VI.

II. System Model

We consider an mMTC network consisting of an eNB and N_m MTCDs. The arrival model of the number of MTCDs follows the time limited Beta distribution over a period of time T [1]. To be specific, all N_m MTCDs are activated between $t = 0$ and $t = T$ according to a Beta distribution. The probability density function (PDF) of the Beta distribution, denoted by $f(t)$, is given by [1]

$$f(t) = \frac{t^{\vartheta-1}(T-t)^{\kappa-1}}{T^{\vartheta+\kappa-1} \cdot Beta(\vartheta,\kappa)}, \quad \vartheta > 0,\ \kappa > 0 \tag{1}$$

where $Beta(\vartheta,\kappa)$ denotes the Beta function with parameters ϑ and κ. Note that the shaping parameters ϑ and κ can control the shape of the distribution. Then, the number of arrivals during the ith random access opportunity (RAO), denoted by $N_a(i)$, is determined by

$$N_a(i) = N_m \int_{t_i}^{t_{i+1}} f(t)dt \tag{2}$$

where t_i denotes the beginning time of the ith RAO.

As shown in Fig. 1, by employing the tagged PA technique and the NOMA approach, multiple devices choosing the same PA are allowed to share the identical physical uplink shared channel (PUSCH) for transmissions and thus can form a NOMA group. For each NOMA group, the channel gain between the ith MTCD and eNB, denoted by h_i, is given by

$$h_i = \frac{g_i}{l_i} \tag{3}$$

where l_i denotes the path-loss and g_i represents the Rayleigh fading channel gain [37]. Then, the PDF of g_i^2, denoted by $f_{g_i^2}(x)$, can be expressed as [38]

$$f_{g_i^2}(x) = \frac{1}{2u^2} e^{-\frac{x}{2u^2}}, \quad x \geq 0 \tag{4}$$

where $2u^2$ denotes the mean of exponentially distributed channel power gain g_i^2.

By employing the power back-off, the SIC is enabled at the eNB such that multiple MTCDs can share the same PUSCH for transmissions. Then, the transmit power of the ith MTCD in a NOMA group, denoted by $p_{t,i}$ with the unit of dB, can be written as [37]

$$p_{t,i}^{\mathrm{dB}} = \min\left\{p_{\max}, p_u^{\mathrm{dB}} - (i-1)\rho + 10\log_{10}M + PL_i\right\} \tag{5}$$

where $p_{\max}$ represents the maximum transmit power, p_u^{dB} is the target arrived power at eNB of the first MTCD in a NOMA group, ρ denotes the power back-off factor, M is the number of resource blocks (RBs) allocated for one PUSCH, and PL_i denotes the path loss of the ith MTCD. Eq. (5) indicates that the received power of MTCDs in a NOMA group decreases with the step of ρ, i.e., the received power of the former MTCD

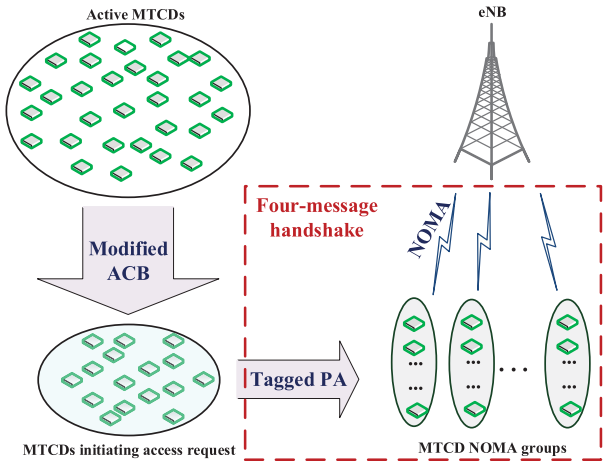


Fig. 1. System model of our proposed NORA scheme.

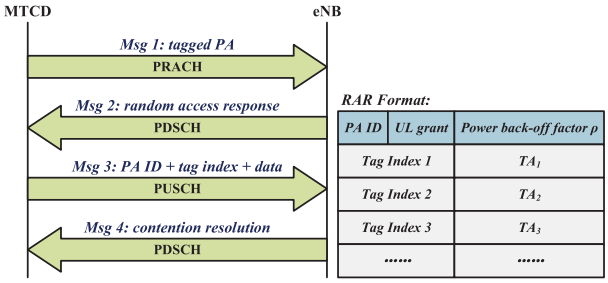


Fig. 2. Four-message handshake procedure of our proposed NORA scheme.

is ρ dB larger than that of the latter MTCD in a NOMA group, and thus the SIC can be used for the data reception. Since the transmit power of MTCD is usually lower than the maximum allowed power, we assume $p_{t,i}^{\mathrm{dB}} < p_{\max}$. Then, according to (5), the transmit power $p_{t,i}$ can be written as

$$p_{t,i} = p_u \cdot 10^{\frac{(1-i)\rho}{10}} \cdot M \cdot l_i^2. \tag{6}$$

Moreover, we assume perfect SIC is performed at the eNB. The notations used in this paper are summarized in TABLE I.

III. Throughput-Oriented Non-Orthogonal Random Access Scheme

A. Scheme Description

In the proposed NORA scheme, active MTCDs have to pass the ACB check before sending the access request, as shown in Fig. 1. To guarantee that more MTCDs can obtain the access opportunities by the proposed NORA scheme, we will also modify the ACB factor, which will be discussed in Section III-D. Then, the MTCDs that pass the ACB check will perform the four-message handshake procedure as illustrated in Fig. 2.

Step 1: Each MTCD randomly selects one *tagged PA* which is identified by both *PA ID* and *tag index*, and then transmits the chosen tagged PA to the eNB via physical random access channel (PRACH).

Step 2: The eNB detects the received tagged PAs through two correlation operations [31]. As each tagged PA is associated with one particular tag index, MTCDs choosing the same

TABLE I
SUMMARY OF NOTATIONS AND DEFINITIONS

Notation	Defination	Notation	Defination
N_m	Total number of MTCDs	T	Time duration of Beta distributed arrival model
$N_a(i)$	Number of arrivals in the ith RAO	h_i	Channel gain between the ith MTCD and eNB
l_i	Pass-loss between the ith MTCD and eNB	g_i	Rayleigh fading channel gain between the ith MTCD and eNB
u^2	Variance of the zero-mean normal distribution	$p_{t,i}$	Transmit power of the ith MTCD in a NOMA group
p_u	Target arrived power at eNB of the first MTCD in a NOMA group	ρ	Power back-off factor
M	Number of RBs allocated for one PUSCH	$p_{r,i}$	Arrival power at eNB of the ith MTCD in a NOMA group
q	Equivalent power back-off factor	I	Number of MTCDs in a NOMA group
SINR_i	SINR at eNB of the ith MTCD	σ^2	Variance of AWGN
$R_{a,i}$	Achievable data rate of the ith MTCD	R_0	Target data rate of MTCD
$Q_i(I,q)$	Probability that the ith MTCD in the NOMA group is successfully decoded	β	SINR threshold
$T_{PA}(I,q)$	Average throughput achieved by one NOMA group including I MTCDs	J	Maximum number of RA attempts allowed for each data packet
p_s	Predefined successful transmission probability threshold	θ	Traditional ACB factor
$\tilde{\theta}$	Modified ACB factor	N_p	Number of available PAs

PA can be distinguished and form a NOMA group. Moreover, eNB can also obtain the channel power gains g_i^2 of MTCDs via the tagged PA detection and thus can sort MTCDs in each NOMA group in descending order. Then, the message 2 (Msg2), i.e. RAR, is sent to each NOMA group through the physical downlink shared channel (PDSCH). The RAR corresponding to each group includes the PA ID of the group, the allocated RBs for PUSCH, the power back-off factor ρ used for power domain multiplexing, and the timing advance (TA) for the uplink synchronization. Note that in Section III-C, we will determine the number of MTCDs that form one NOMA group. Without the loss of generality, we assume that each NOMA group includes I MTCDs. Then, when the number of MTCDs choosing the same PA is larger than I, only the former I MTCDs are selected to form the NOMA group and can receive Msg2. The remaining MTCDs are discarded and will perform a uniform back-off for the next access opportunity.

Step 3: MTCDs in each NOMA group obtain the power multiplexing order and the power back-off factor ρ from the received RAR. Then, each MTCD sends the message 3 (Msg3), which includes the PA ID, tag index and data packets, through the same PUSCH. As eNB can distinguish messages sent by different MTCDs through the associated tag index and SIC, data packets can be transmitted with Msg3, which is different from traditional RA schemes and can efficiently reduce the signaling overhead.

Step 4: The eNB will use SIC to decode the received data of each NOMA group and send the CR messages to MTCDs whose data are successfully decoded. Note that MTCDs that select the same tagged PA or choose different tagged PAs but experience the SIC failure cannot receive the CR message. In order to alleviate the congestion, MTCDs without receiving the CR message will perform a uniform back-off for the next access request.

Remark 1: As each tagged PA consists of one PA ZC sequence and one tag ZC sequence, MTCDs that select the same PA but different tags can be distinguished by eNB via two correlation operations [31]. Then, each pair of detected PA ID and tag index is associated with one particular MTCD. Moreover, since the transmit power of tagged PAs is fixed and can be known by eNB, the tagged PAs can be used for channel estimation to obtain channel gain h_i in (3). Furthermore, eNB can also obtain the time of arrival (TOA) information and TA value of each detected MTCD during the tagged PA detections [39], [40]. Then, eNB can calculate the distance between each detected MTCD and itself to estimate the path-loss. In this way, the small-scale fading channel gain can also be obtained.

Remark 2: By utilizing the TA values included in RAR messages or cell-specific reference signals (CRS) inserted in the PDSCH [39], [40], each detected MTCD can calculate the distance between itself and eNB to estimate the path-loss for determining the transmit power as given by (5) and (6).

Remark 3: As the tagged PA technique allows eNB to distinguish MTCDs that select the same PA but different tags, the tag indices of selected MTCDs will be included in the RAR message as shown in Fig. 2. Then, each MTCD can know whether it is selected or not by checking the listed tag indices in the RAR message. Moreover, the tag indices are sorted according to the descending order of channel power gains. Thus, selected MTCDs can also know their power back-off orders.

B. Optimization Problem Formulation

As we have already assumed, each NOMA group includes I MTCDs. Then, based on (6), the arrived power at eNB of the ith MTCD in a NOMA group, denoted by $p_{r,i}$, can be written as

$$p_{r,i} = p_{t,i} \cdot {h_i}^2 = {g_i}^2 10^{\frac{(1-i)\rho}{10}} M p_u, \quad i = 1, 2, \cdots, I. \quad (7)$$

Define the *equivalent power back-off factor* $q = 10^{-\frac{\rho}{10}}$, then (7) can be converted to

$$p_{r,i} = {g_i}^2 q^{i-1} M p_u, \quad i = 1, 2, \cdots, I. \quad (8)$$

Since $\rho > 0$, then $q \in (0, 1)$. Moreover, MTCDs in each group are sorted by eNB in the descending order based on the channel power gain g_i^2 as described in Section III-A, i.e., ${g_1}^2 \geq {g_2}^2 \geq \cdots \geq {g_I}^2$, then we have $\frac{{g_i}^2}{{g_{i+1}}^2} \geq 1$

$(i = 1, \cdots, I-1)$. Hence, we have

$$\frac{p_{r,i}}{p_{r,i+1}} = \frac{{g_i}^2}{{g_{i+1}}^2} \cdot q^{-1} > 1, \quad i = 1, \cdots, I-1. \tag{9}$$

Consequently, we can obtain that the detection order of SIC is the same as the order of MTCDs in the NOMA group.

By employing the SIC technique, if the former $i-1$ MTCDs are successfully decoded, the signal-to-interference plus noise ratio (SINR) at the eNB of the ith MTCD, denoted by SINR_i, is determined by

$$\mathrm{SINR}_i = \frac{p_{r,i}}{\sum_{m=i+1}^{I} p_{r,m} + \sigma^2} \tag{10}$$

where σ^2 is the variance of additive white Gaussian noise (AWGN). Then the achievable data rate of the ith MTCD, denoted by $R_{a,i}$, can be expressed as

$$R_{a,i} = \log(1 + \mathrm{SINR}_i). \tag{11}$$

Assuming that all MTCDs have the same target data rate R_0, then the event that eNB can successfully decode the message of the ith MTCD can be described by $R_{a,i} \geq R_0$.

Denote by $Q_i\,(I, q)$ the probability that the data transmitted by the ith MTCD in the NOMA group is successfully decoded when the data of the former $i-1$ MTCDs are decoded and perfectly removed. Then, $Q_i\,(I, q)$ $(i = 1, \cdots, I)$ can be written as

$$\begin{aligned} Q_i\,(I, q) &= \Pr\{R_{a,i} \geq R_0\} = \Pr\left\{\mathrm{SINR}_i \geq 2^{R_0} - 1\right\} \\ &= \Pr\left\{\frac{{g_i}^2 q^{i-1}}{\sum_{m=i+1}^{I} \left({g_m}^2 q^{m-1}\right) + \frac{\sigma^2}{p_u M}} \geq \beta\right\} \\ &= \Pr\left\{\frac{1}{\beta}{g_i}^2 q^{i-1} - \sum_{m=i+1}^{I} \left({g_m}^2 q^{m-1}\right) \geq \Phi\right\} \end{aligned} \tag{12}$$

where

$$\Phi = \frac{\sigma^2}{p_u M} \tag{13}$$

and β denotes the SINR threshold and is determined by

$$\beta = 2^{R_0} - 1. \tag{14}$$

Since ${g_1}^2 \geq {g_2}^2 \geq \cdots \geq {g_I}^2$, $\frac{1}{\beta}{g_i}^2 q^{i-1} - \sum_{m=i+1}^{I} \left({g_m}^2 q^{m-1}\right)$ in (12) can be viewed as *a linear weighted sum of order statistics*, which creates a challenge in calculating the probability $Q_i\,(I, q)$. However, we can employ the Sukhatme's classical theory to convert $\frac{1}{\beta}{g_i}^2 q^{i-1} - \sum_{m=i+1}^{I} \left({g_m}^2 q^{m-1}\right)$ to *a linear weighted sum of independent random variables*. We recall the following theorem.

Theorem 1: Define Λ_i $(i = 1, \cdots, I)$ as I independent identically distributed (i.i.d.) exponential random variables having the common mean $\overline{\Lambda}$. Let $\Lambda_{i:I}$ denote the exponential random variable related to the ith largest observation from the I original variables, i.e., $\Lambda_{1:I} \geq \Lambda_{2:I} \geq \cdots \geq \Lambda_{I:I}$. Then the spacings between ordered exponential random variables, denoted by $x_i = \Lambda_{i:I} - \Lambda_{i+1:I}$, $(i = 1, \cdots, I)$, are independently exponential random variables with probability density functions $f_{x_i}(x) = \frac{i}{\overline{\Lambda}} e^{-ix/\overline{\Lambda}}$, $x \geq 0$.

Proof: The proof of the theorem can be found in [42]. ■

Define the spacing random variables G_i $(i = 1, \cdots, I)$ as

$$G_i = g_i^2 - g_{i+1}^2, \quad i = 1, 2, \cdots, I-1 \tag{15}$$

and

$$G_I = g_I^2. \tag{16}$$

Based on *Theorem 1*, the spacing random variables $G_1, G_2, \cdots, G_I$ are independent exponential and the PDF of G_i is given by

$$f_{G_i}(x) = \frac{i}{2u^2} e^{-ix/2u^2}, \quad (i = 1, 2, \cdots, I). \tag{17}$$

According to (15), we have

$$g_i^2 = \sum_{r=i}^{I} G_r, \quad i = 1, \cdots, I. \tag{18}$$

Then, we can obtain

$$\begin{aligned} &\frac{1}{\beta}{g_i}^2 q^{i-1} - \sum_{m=i+1}^{I} \left({g_m}^2 q^{m-1}\right) \\ &= \frac{1}{\beta} q^{i-1} \sum_{r=i}^{I} G_r - \sum_{m=i+1}^{I} \left(q^{m-1} \sum_{r=m}^{I} G_r\right) \\ &= \frac{1}{\beta} q^{i-1} G_i + \frac{1}{\beta} q^{i-1} \sum_{m=i+1}^{I} G_m - \sum_{m=i+1}^{I} \left(q^{m-1} \sum_{r=m}^{I} G_r\right) \\ &= G_i \frac{1}{\beta} q^{i-1} + \sum_{m=i+1}^{I} G_m \left(\frac{1}{\beta} q^{i-1} - \sum_{r=i}^{m-1} q^r\right) \\ &= \sum_{n=i}^{I} G_n \left(\frac{1}{\beta} q^{i-1} - \frac{q^i - q^n}{1-q}\right). \end{aligned} \tag{19}$$

Substituting (19) into (12), we can derive

$$Q_i(I, q) = \Pr\left\{\sum_{n=i}^{I} G_n \left(\frac{1}{\beta} q^{i-1} - \frac{q^i - q^n}{1-q}\right) > \Phi\right\}. \tag{20}$$

Since G_n $(n = i, \cdots, I)$ are independent exponential random variables having mean $\frac{2u^2}{n}$, the probability $Q_i(I, q)$ can also be written as

$$Q_i(I, q) = \Pr\left\{\sum_{n=i}^{I} \frac{1}{n} \left(\frac{1}{\beta} q^{i-1} - \frac{q^i - q^n}{1-q}\right) Y_n > \Phi\right\} \tag{21}$$

where Y_n $(n = i, \cdots, I)$ are i.i.d. exponential random variables having mean $2u^2$. Define $Z_{I,\,q,\,i}$ as

$$Z_{I,\,q,\,i} = \sum_{n=i}^{I} \frac{1}{n} \left(\frac{1}{\beta} q^{i-1} - \frac{q^i - q^n}{1-q}\right) Y_n = \sum_{n=i}^{I} a_{i,\,n} Y_n \tag{22}$$

where

$$a_{i,\,n} = \frac{1}{n} \left(\frac{1}{\beta} q^{i-1} - \frac{q^i - q^n}{1-q}\right), \quad n = i, \cdots, I. \tag{23}$$

Then, we have $Q_i(I, q) = \Pr\{Z_{I,q,i} > \Phi\}$ and $Z_{I,q,i}$ can be viewed as *a linear weighted sum of independent exponential random variables*. Although the PDF of the linear weighted sum of independent exponential random variables

has already been investigated [36], the result obtained in the existing literature cannot be directly applied to determine the PDF of $Z_{I,q,i}$. To be specific, the existing literature requires that the weighted coefficients of independent random variables are nonnegative. However, the coefficients $a_{i,n}$ $(n = i, \cdots, I)$ can be either positive or negative. Since the characteristic function is the Fourier transform of the PDF, each characteristic function corresponds to a unique distribution function. Then, the characteristic function of the sum of independent random variables is equal to the multiplication of the corresponding characteristic functions. Therefore, by employing the characteristic function approach, we can obtain the PDF of $Z_{I,q,i}$, which is determined in the following theorem.

Theorem 2: The PDF of $Z_{I,q,i}$ given by (22) is determined by

$$f_{Z_{I,\,q,\,i}}(x) = \begin{cases} \sum_{n=i}^{I} \alpha_{I,i,n} \left[-\frac{1}{2u^2 a_{i,\,n}} e^{-\frac{x}{2u^2 a_{i,\,n}}} \right] \mathbb{I}\{a_{i,n} < 0\}, & x \le 0 \\ \sum_{n=i}^{I} \alpha_{I,i,n} \left[\frac{1}{2u^2 a_{i,\,n}} e^{-\frac{x}{2u^2 a_{i,\,n}}} \right] \mathbb{I}\{a_{i,n} > 0\}, & x > 0 \end{cases} \tag{24}$$

where $\mathbb{I}\{\cdot\}$ denotes the indicator function and

$$\begin{aligned} \alpha_{I,i,n} &= \prod_{k=i,k\neq n}^{I} \left(1 - \frac{a_{i,\,k}}{a_{i,\,n}}\right)^{-1} \\ &= \frac{{a_{i,\,n}}^{I-i}}{\prod_{k=i,k\neq n}^{I} (a_{i,\,n} - a_{i,\,k})}, \quad n = i, \cdots, I. \end{aligned} \tag{25}$$

Proof: The proof of the theorem is provided in Appendix A. ∎

By plugging (24) into (21), we can obtain

$$\begin{aligned} &Q_i(I,q) \\ &= \Pr\left\{ Z_{I,\,q,\,i} > \Phi \right\} \\ &= \int_{\Phi}^{\infty} \sum_{n=i}^{I} \alpha_{I,i,n} \left[\frac{1}{2u^2 a_{i,n}} \exp\left(-\frac{x}{2u^2 a_{i,n}} \right) \right] \mathbb{I}\{a_{i,n} > 0\} dx \\ &= \sum_{n=i}^{I} \alpha_{I,i,n} \exp\left(-\frac{\psi}{a_{i,n}} \right) \mathbb{I}\{a_{i,n} > 0\} \end{aligned} \tag{26}$$

where $\psi = \frac{\Phi}{2u^2}$. When $a_{i,\,n} > 0$, we have

$$q^{n-i+1} > q - \frac{1-q}{\beta}, \quad n = i, i+1, \cdots I \tag{27}$$

and thus the probability $Q_i(I,q)$ can be analyzed from the following two cases.

Case 1: $0 < q \le \frac{1}{\beta+1}$

In this case, we can obtain $q - \frac{1-q}{\beta} \le 0$, so (27) holds for $n = i, \cdots, I$ and thus $a_{i,\,n} > 0$ $(n = i, i+1, ...I)$. Then, we have

$$Q_i(I,q) = \sum_{n=i}^{I} \alpha_{I,\,i,\,n} \exp\left(-\frac{\psi}{a_{i,\,n}} \right). \tag{28}$$

Case 2: $\frac{1}{\beta+1} < q < 1$

In this case, we have $q - \frac{1-q}{\beta} > 0$ and thus $a_{i,\,n} > 0$ when $n < i - 1 + \log_q\left(q - \frac{1-q}{\beta}\right)$. Let

$$s_i = i - 1 + \left\lfloor \log_q\left(q - \frac{1-q}{\beta}\right) \right\rfloor, \quad (i = 1, \cdots, I). \tag{29}$$

Then, we can obtain

$$Q_i(I,q) = \sum_{n=i}^{\min\{I,s_i\}} \alpha_{I,\,i,\,n} \exp\left(\frac{-\psi}{a_{i,\,n}} \right). \tag{30}$$

Note that $\frac{1-q}{\beta} > 0$, then we have $0 < q - \frac{1-q}{\beta} < q$, and thus $\log_q\left(q - \frac{1-q}{\beta}\right) > 1$. Moreover, when $\beta > 1$ and $\frac{1}{\beta} < q < 1$, we have $\log_q\left(q - \frac{1-q}{\beta}\right) < 2$, then $\left\lfloor \log_q\left(q - \frac{1-q}{\beta}\right) \right\rfloor = 1$ and thus $s_i = i$, which results in a simple expression of $Q_i(I,q)$, i.e.,

$$Q_i(I,q) = \alpha_{I,\,i,\,i} \exp\left(\frac{-\psi}{a_{i,\,i}} \right). \tag{31}$$

Based on the above analysis, $Q_i(I,q)$ is given by

$$Q_i(I,q) = \begin{cases} \sum_{n=i}^{I} \alpha_{I,i,n} \exp\left(-\frac{\psi}{a_{i,n}}\right), & 0 < q \le \frac{1}{\beta+1} \\ \sum_{n=i}^{\min\{I,s_i\}} \alpha_{I,i,n} \exp\left(-\frac{\psi}{a_{i,n}}\right), & \beta \le 1, \frac{1}{\beta+1} < q < 1 \\ & \text{and } \beta > 1, \frac{1}{\beta+1} < q < \frac{1}{\beta} \\ \alpha_{I,\,i,\,i} \exp\left(\frac{-\psi}{a_{i,\,i}}\right), & \beta > 1, \frac{1}{\beta} < q < 1. \end{cases} \tag{32}$$

When a NOMA group has I MTCDs, the event that the $(I+1)$th MTCD is successfully decoded by eNB will never happen and thus the corresponding probability is equal to 0. Then, we can obtain $Q_{I+1}(I,q) = 0$. Note that $Q_i(I,q)$ is the conditional probability that the data transmitted by the ith MTCD in the NOMA group is successfully decoded when the data of the former $i-1$ MTCDs are decoded and perfectly removed. Then, $Q_i(I,q)$ $(i = 1, \cdots, I)$ might not be independent with each other. However, in order to facilitate the analysis, we ignore the effect of the SIC process on successful decoding probability of MTCDs, i.e., $Q_i(I,q)$ $(i = 1, \cdots, I)$ are assumed to be independent [37]. Thus, the probability that only former i MTCDs in the NOMA group are successfully decoded, denoted by $S_{PA}(i,I,q)$, is written by

$$S_{PA}(i,I,q) = \prod_{t=1}^{i} Q_t(I,q) \Big[1 - Q_{i+1}(I,q) \Big]. \tag{33}$$

As MTCDs can transmit data packets only when the four-message handshake procedure is successfully conducted, the network throughput is proportional to the number of MTCDs that successfully access the network. Consequently, we define the number of MTCDs that successfully finish

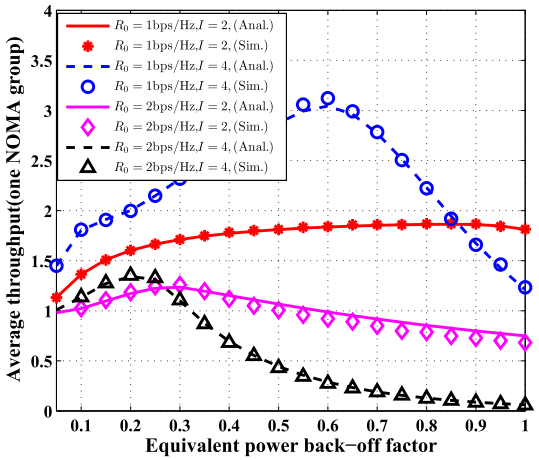


Fig. 3. Comparisons of the analytical and simulation results for the average throughput that can be achieved by one NOMA group.

the four-message handshake procedure as throughput. Then, the average throughput that can be achieved by one NOMA group including I MTCDs, i.e., the average number of MTCDs of one NOMA group that can successfully access the network in each RAO, denoted by $T_{PA}(I,q)$, is determined by

$$T_{PA}(I,q)=\sum_{i=1}^{I} i\cdot S_{PA}(i,I,q). \tag{34}$$

Theorem 3: The average throughput given by (34) can be further simplified to

$$\begin{aligned} T_{PA}(I,q) &= \sum_{i=1}^{I}\left\{ i\cdot\prod_{t=1}^{i} Q_t(I,q)\Big[1-Q_{i+1}(I,q)\Big]\right\} \\ &= \sum_{i=1}^{I}\left\{\prod_{t=1}^{i} Q_t(I,q)\right\}. \end{aligned} \tag{35}$$

Proof: The proof is provided in Appendix B. ∎

In Fig. 3, we compare the analytical result calculated by (35) with the Monte-Carlo simulation result that considers the effect of the SIC process on the successful decoding probability. We can observe from Fig. 3 that the analytical result agrees well with the simulation result under different power back-off factors, target data rates, and the number of MTCDs included in one NOMA group, which demonstrates the validity of our obtained analytical result.

Based on the above discussions, we can formulate following optimization problem. Specifically, we aim at maximizing the average throughput achieved by each NOMA group while meeting the constraints on the power back-off factor, the number of MTCDs contained in each NOMA group, and the successful transmission probability. Mathematically, our formulated optimization problem can be written as

$$\max_{q,I}\ T_{PA}(I,q) \tag{36a}$$

$$\text{s.t. } 0<q<1 \tag{36b}$$

$$1-\left(1-\frac{T_{PA}(I,q)}{I}\right)^{J}\geq p_s \tag{36c}$$

$$I\in N^{+} \tag{36d}$$

where J denotes the maximum number of RA attempts allowed for each data packet to transmit and p_s is the predefined successful transmission probability threshold. As $T_{PA}(I,q)$ given by (35) can also be viewed as the average number of MTCDs in the NOMA group that can be successfully decoded via one SIC process, $T_{PA}(I,q)/I$ represents the successful transmission probability of each MTCD for one access attempt. Thus, the constraint given by (36c) implies that the overall successful transmission probability of each MTCD cannot be below the predefined probability threshold.

C. PSO Algorithm Based Solution

Due to the complicated expression of the objective function $T_{PA}(I,q)$ as shown in (32) and (35), our formulated problem cannot be easily solved. Consequently, we employ a two-step scheme to solve efficiently the throughput maximization problem. Following the two-step scheme, the obtained power back-off factor and the number of MTCDs contained in one NOMA group are denoted by $q^{\star}$ and $I^{\star}$, respectively. To be specific, in the first step, we aim to determine the power back-off factor under the given I. Then, we can derive the corresponding average throughput under the given I. By performing a one-dimensional searching of I, we obtain the solution $q^{\star}$ and $I^{\star}$.

For given I, we develop a particle swarm optimization (PSO) algorithm to determine the corresponding power back-off factor $q^{*}(I)$. The basic idea of the PSO algorithm is to solve the optimization problem through cooperation and information sharing among individuals in a population of particles. Because the PSO algorithm can take the advantage of the parallel computing, simple operation and few parameters settings, the optimization problem can be efficiently solved within several iterations. Then, we define m back-off factors for the parallel computing. Denote $q_i^{(k)}(I)$ and $v_i^{(k)}(I)$ $(i=1,\cdots,m)$ as the value of ith back-off factor and the associated step-length, respectively, which are obtained in the kth iteration. Then, the step-lengths and the corresponding values of our defined m back-off factors are updated according to [44]

$$\begin{aligned} v_i^{(k)} = \varpi v_i^{(k-1)} &+ c_1 r_1^{(k)}\left(b_i^{(k)}-q_i^{(k-1)}\right) \\ &+ c_2 r_2^{(k)}\left(d^{(k)}-q_i^{(k-1)}\right) \end{aligned} \tag{37}$$

and

$$q_i^{(k)} = q_i^{(k-1)} + v_i^{(k)} \tag{38}$$

respectively, where ϖ is the inertia weight factor, c_1 and c_2 denote cognitive and social parameter, respectively, $r_1^{(k)}$ and $r_2^{(k)}$ are two independent numbers uniformly produced within $[0,1]$ for the kth iteration, $b_i^{(k)}$ is given by

$$b_i^{(k)} = \text{argmax}_{q_i^{(1)},\cdots,q_i^{(k-1)}} T_{PA}(I,q) \tag{39}$$

and $d^{(k)}$ is determined by

$$d^{(k)} = \text{argmax}_{b_1^{(k)},\cdots,b_m^{(k)}} T_{PA}(I,q). \tag{40}$$

Specifically, the inertia weight factor is used to adjust the search range of the solution space. The parameters c_1 and

Algorithm 1 PSO Algorithm for The Optimization Problem

Initialize: $I=1,\ \beta,\ \psi,\ m,\ \varpi,\ c_1,\ c_2,\ L$;
while $I \geq 1$ **do**
  initialize $q_i^{(0)}(I)$ and $v_i^{(0)}(I)$ $(i=1,\cdots,m)$;
  for k=*I to* L **do**
    for i=*1 to* m **do**
      update $v_i^{(k)}(I)$ based on (37);
      update $q_i^{(k)}(I)$ based on (38);
      calculate $T_{PA}(I,q)$ based on (32) and (35);
      calculate $b_i^{(k+1)}(I)$ based on (39) ;
    end
    calculate $d^{(k+1)}(I)$ based on (40);
  end
  $q^*(I)=q_m^{(L)}(I)$;
  if $1-\left(1-\frac{T_{PA}(I,q)}{I}\big|_{q=q^*(I)}\right)^J < p_s$ **then**
    $I_{upper}=I$;
    Break;
  else
    $I=I+1$;
  end
end
$I^\star=\text{argmax}_{1,\ldots,I_{upper}}\ T_{PA}(I,q)\big|_{q=q^*(I)}$;
$q^\star=q^*(I^\star)$;

c_2 are both positive constants and are usually equal to two. As shown in (37), the equation used for updating the step-length includes three parts. To be specific, the first part $\varpi v_i^{(k-1)}$ is related to the step-length of the corresponding back-off factor in the last iteration, where ϖ is brought into the first term to balance the ability between the global and local search. The second term $c_1 r_1^{(k)}\left(b_i^{(k)}-q_i^{(k-1)}\right)$ denotes the cognitive part, in which $b_i^{(k)}$ given by (39) is the best local result of the ith back-off factor in the prior $k-1$ iterations. The third part $c_2 r_2^{(k)}\left(d^{(k)}-q_i^{(k-1)}\right)$ represents the social part, where $d^{(k)}$ determined by (40) is the best global result that has been found in the prior $k-1$ iterations by all the back-off factors.

In our developed PSO algorithm, each back-off factor uses its local optimal result $b_i^{(k)}$ and the global optimal result $d^{(k)}$ which is shared by all particles to adjust its step-length and the corresponding value for the next iteration, as shown in (37) and (38). Consequently, all back-off factors converge after a number of iterations, and the corresponding value of the power back-off factor $q^*(I)$ for the given I can be obtained by adopting the developed PSO algorithm.

Theorem 4: $T_{PA}(I,q)/I$ is a decreasing function of I.

Proof: The proof is provided in Appendix C. ∎

According to *Theorem 4*, the left-hand side of (36c) is also a decreasing function of I, which implies that I must be upper-bounded by constraint (36c). Therefore, the optimal number of MTCDs contained in one NOMA group $I^\star$ can be determined by a one-dimensional search. Denoting this upper-bound by I_{upper}, then we have

$$I^\star=\underset{1,\ldots,I_{upper}}{\text{argmax}}\ T_{PA}(I,q)\big|_{q=q^*(I)}. \tag{41}$$

Thus, we can derive that the power back-off factor $q^\star = q^*(I^\star)$. The algorithm for determining the PSO-based solution $q^\star$ and $I^\star$ is shown in Algorithm 1, where L is the maximum number of iterations for the PSO algorithm.

D. Modified ACB Mechanism

Based on our derived PSO-based solution, the ACB mechanism can also be correspondingly adjusted such that more MTCDs can acquire the access opportunities. Specifically, in conventional ACB mechanisms, the ACB factor denoted by θ is given by

$$\theta=\min\left\{1,\frac{N_p}{N_a}\right\} \tag{42}$$

where N_p is the number of available PAs and N_a is the number of currently active MTCDs that can be estimated by eNB [11]. It should be noted that in the ith RAO, the number of currently activated MTCDs N_a in (42), which is a random variable, includes not only the newly arrived MTCDs, but also the MTCDs which experience access failure in the former RAOs and re-initiate the RA procedure in current RAO. However, $N_a(i)$ determined by (2) represents the statistically newly activated number of MTCDs in the ith RAO. Thus, N_a is generally not equal to $N_a(i)$. In the ACB mechanism, eNB broadcasts the ACB factor θ to all active MTCDs at the beginning of each RAO, meanwhile each active MTCD generates a random number denoted by $\hat{\theta}$, which is uniformly distributed between 0 and 1. Then, each MTCD will compare the received ACB factor θ with its produced random number $\hat{\theta}$ to decide whether to transmit the access request message. To be specific, if $\hat{\theta}\leq\theta$, the corresponding MTCD is allowed to enter the conventional four-message handshake procedure. Otherwise, the MTCD will perform a uniform back-off to wait for the next RAO. Because the ACB mechanism can control the number of contending MTCDs for each RAO and scatter the active MTCDs in the time domain uniformly, the network congestion can be efficiently alleviated.

In the previous sections, we have described the proposed NORA scheme for the mMTC network, which employs the tagged PA technique and the NOMA technology to enable more MTCDs to obtain the access opportunities. Moreover, by formulating the average throughput maximization problem, we also have determined the PSO-based power back-off factor $q^\star$ and the number of MTCDs contained in each NOMA group $I^\star$. By integrating the power-domain multiplexing, the proposed NORA scheme allows $I^\star$ MTCDs to share one particular PA. Consequently, to improve further the network performance, the ACB factor corresponding to our proposed scheme should also be adjusted. Specifically, different from the traditional ACB mechanism, our modified ACB factor denoted by $\tilde{\theta}$ is determined by

$$\tilde{\theta}=\min\left\{1,\frac{N_p I^\star}{N_a}\right\}. \tag{43}$$

Comparing (42) and (43), we can observe that our modified ACB factor $\tilde{\theta}$ is $I^\star$ times as large as the conventional ACB factor θ. Such a modification statistically guarantees that each PA can be shared by $I^\star$ MTCDs, which can efficiently increase the average throughput in each NOMA group and thus improves the system performance.

IV. Analysis Under High Data Rate Requirement

According to the previous analysis, we can find that the target data rate R_0 plays a critically important role on the

achievable network throughput. If the MTCD requires high target data rate, e.g., future smart grid requires data rate larger than 3.072 bps/Hz [45], we have that $\beta = 2^{R_0} - 1$ is usually much larger than one. In this case, we can know that the interval $[1/\beta, 1)$ occupies most part of the domain for the back-off factor q as indicated by (32). Consequently, the optimal solution for $q \in [1/\beta, 1)$ should achieve the near-optimal performance when the target data rate is high, i.e., we can regard the optimal solution within the interval $q \in [1/\beta, 1)$ as the suboptimal solution within the entire interval $q \in (0, 1)$ under the high data rate requirement.

When $\beta > 1$ and $q \in [1/\beta, 1)$, we can know that $Q_t(I, q)$ $(t = 1, \cdots, I)$ is given by (31). Then, when the target data rate R_0 is high, our optimization problem can be approximately simplified to

$$\max_{q,I} \; T_{PA}(I, q) = \sum_{i=1}^{I} \left\{ \prod_{t=1}^{i} \alpha_{I,t,t} \exp\left(\frac{-\psi}{a_{t,t}}\right) \right\} \tag{44a}$$

$$\text{s.t. } \frac{1}{\beta} \le q < 1 \tag{44b}$$

$$\text{Eqs. (36c) and (36d)}$$

Note that the following analyses in this section are conducted under the conditions $\beta > 1$ and $q \in [1/\beta, 1)$, i.e., $Q_t(I, q)$ is described by (31).

Define $M_i(I, q)$ as

$$M_i(I, q) = \prod_{t=1}^{i} Q_t(I, q) = \prod_{t=1}^{i} \alpha_{I,\,t,\,t} \exp\left(\frac{-\psi}{a_{t,t}}\right). \tag{45}$$

Then, the average throughput $T_{PA}(I, q)$ given by (44a) can be expressed by

$$T_{PA}\,(I, q) = \sum_{i=1}^{I} M_i(I, q). \tag{46}$$

Based on (25), $\alpha_{I,\,t,\,t}$ can be written as $\alpha_{I,\,t,\,t} = \prod_{k=t+1}^{I} \frac{k}{k-t+t\beta\sum_{e=1}^{k-t} q^e}$. Then $M_i(I, q)$ is equal to $\prod_{t=1}^{i} \Big(\prod_{k=t+1}^{I} \frac{k}{k-t+t\beta\sum_{e=1}^{k-t} q^e} \Big) \exp\left(\frac{-\psi}{a_{t,t}}\right)$. Moreover, $M_i(I+1, q)$ $(i = 1, \cdots, I)$ can be determined by

$$\begin{aligned} &M_i(I+1, q) \\ &= \prod_{t=1}^{i} \left(\prod_{k=t+1}^{I+1} \frac{k}{k-t+t\beta\sum_{e=1}^{k-t} q^e} \right) \exp\left(\frac{-\psi}{a_{t,t}}\right) \\ &= \left(\prod_{t=1}^{i} \frac{I+1}{I+1-t+t\beta \sum\limits_{e=1}^{I+1-t} q^e} \right) \\ &\quad \cdot \left[\prod_{t=1}^{i} \left(\prod_{k=t+1}^{I} \frac{k}{k-t+t\beta \sum\limits_{e=1}^{k-t} q^e} \right) \exp\left(\frac{-\psi}{a_{t,t}}\right) \right] \\ &= X_i(I, q) \cdot M_i(I, q), \quad (i = 1, \cdots, I) \end{aligned} \tag{47}$$

where

$$X_i(I, q) = \prod_{t=1}^{i} \frac{I+1}{I+1-t+t\beta\sum_{e=1}^{I+1-t} q^e}. \tag{48}$$

Note that $\alpha_{I,\,I,\,I} = 1$ for any given I. Thus, $M_{I+1}(I+1, q)$ can be expressed as

$$\begin{aligned} &M_{I+1}(I+1, q) \\ &= \prod_{t=1}^{I+1} \Big(\prod_{k=t+1}^{I+1} \frac{k}{k-t+t\beta\sum_{e=1}^{k-t} q^e} \Big) \exp\left(\frac{-\psi}{a_{t,t}}\right) \\ &= \left(\prod_{t=1}^{I} \frac{I+1}{I+1-t+t\beta \sum\limits_{e=1}^{I+1-t} q^e} \right) \\ &\quad \times \left[\prod_{t=1}^{I} \left(\prod_{k=t+1}^{I} \frac{k}{k-t+t\beta \sum\limits_{e=1}^{k-t} q^e} \right) e^{\frac{-\psi}{a_{t,t}}} \right] \exp\left(\frac{-\psi}{a_{I+1,I+1}}\right) \\ &= X_I(I, q) M_I(I, q) e^{-\psi\beta\frac{I+1}{q^I}} \end{aligned} \tag{49}$$

By plugging (47) and (49) into (46), we can further obtain the expression of $T_{PA}(I+1, q)$ within the interval $[1/\beta, 1)$ as

$$T_{PA}(I+1, q) = F(I, q) M_I(I, q) + \sum_{i=1}^{I-1} X_i(I, q) M_i(I, q) \tag{50}$$

where

$$F(I, q) = \left(1 + e^{-\psi\beta\frac{I+1}{q^I}}\right) X_I(I, q). \tag{51}$$

Then, we can obtain the following theorem.

Theorem 5: For any given I, $F(I, q)$ is a decreasing function of q within the interval $q \in [1/\beta, 1)$.

Proof: The proof is provided in Appendix D. ∎

Based on *Theorem 5*, the following theorem can be further obtained.

Theorem 6: For any given I, the average throughput of the NOMA group $T_{PA}(I, q)$ is a decreasing function of q in $[1/\beta, 1)$.

Proof: The proof is provided in Appendix E. ∎

According to *Theorem 6*, we can know that the optimal back-off factor in the interval $[1/\beta, 1)$ is equal to $1/\beta$. As $[1/\beta, 1)$ occupies most part of the interval $q \in (0, 1)$ when $\beta \gg 1$, the suboptimal back-off factor denoted by q_{sub} under the high data rate requirement is

$$q_{\text{sub}} = \frac{1}{\beta}. \tag{52}$$

Since I is upper-bounded by constraint (36c), the suboptimal number of MTCDs included in one NOMA group, denoted by I_{sub}, is determined by

$$I_{\text{sub}} = \underset{I \in \mathcal{I}}{\operatorname{argmax}} T_{PA}(I, q)\big|_{q=\frac{1}{\beta}} \tag{53}$$

where

$$\mathcal{I} = \left\{ I \in N^+ \Big| 1 - \left(1 - \frac{T_{PA}\,(I, q)}{I}\Big|_{q=\frac{1}{\beta}}\right)^J \ge p_s \right\}. \tag{54}$$

By employing the developed suboptimal scheme given by (52) and (53), we can efficiently reduce the algorithm complexity and achieve near-PSO performance under the high data rate requirement, which will be illustrated in the following Section.

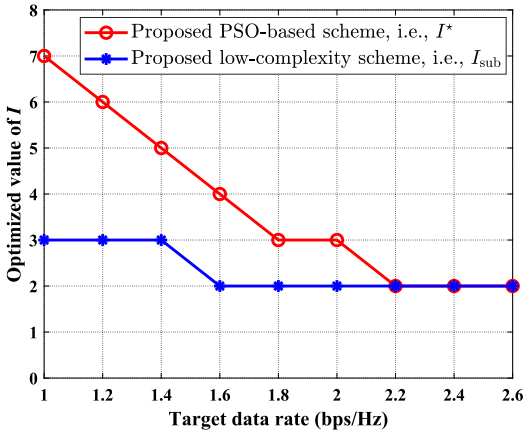


Fig. 4. Optimized value of I versus the target data rate under the proposed PSO-based and low-complexity NORA schemes.

V. Simulation Results

In this section, we evaluate our proposed PSO-based and low-complexity suboptimal NORA schemes through simulations. Specifically, we set the time duration of Beta distributed arrival model as $T = 10$ s and the shape parameters of the Beta function $\vartheta = 3$ and $\kappa = 4$ [1]. Moreover, we set the parameter of the exponentially distributed channel power gain $u^2 = 1$ and the number of RBs allocated for each PUSCH $M = 1$. Besides, we assume that the number of available PAs N_p is equal to 54, the duration of each RAO is 5 ms, the target arrived SNR $\frac{p_u}{\sigma^2} = 10$ dB, the maximum number of RA attempts for each data packet $J = 10$, the successful transmission probability threshold $p_s = 99.5\%$, and the back-off window in uniform back-off mechanism for failing packets is 20ms, i.e., 4 RAOs. Moreover, we also assume that the eNB can perfectly estimate the number of active MTCDs for each RAO such that the ACB factor can be correspondingly adjusted.

To illustrate the superiority of our proposed schemes, we compare the performances of the traditional ACB-based NORA scheme [41] and the traditional ACB-based orthogonal RA (ORA) scheme [11]. Specifically, in the traditional ACB-based NORA scheme, MTCDs that choose the same PA to access will form a NOMA group after the traditional ACB check, and the power back-off factor ρ is set as 3 dB for PUSCH transmission in each NOMA group. In our simulations, three performance metrics, which are total service time, average throughput, and average access delay, respectively, are evaluated. The total service time is defined as the required number of RAOs to serve all MTCDs within J PA transmissions. The average throughput is defined as the ratio of successful MTCDs that have completed the four-message handshake process and access the network within J PA transmissions to the total service number of RAOs. The average access delay is the average number of RAOs for each MTCD to successfully access the network.

Figure 4 shows the optimized number of MTCDs in each NOMA group versus the target data rate under the proposed PSO-based and low-complexity NORA schemes. We can observe from Fig. 4 that the number of MTCDs included in one NOMA group under both the proposed PSO-based and low-complexity NORA schemes are decreasing functions of the target data rate. Specifically, when the target data rate is low, more MTCDs are allowed to form a NOMA group. However, when the target data rate is high, each NOMA group only includes less MTCDs. Moreover, we can also observe that the optimized number of MTCDs in one NOMA group under the proposed PSO-based and low-complexity NORA schemes,

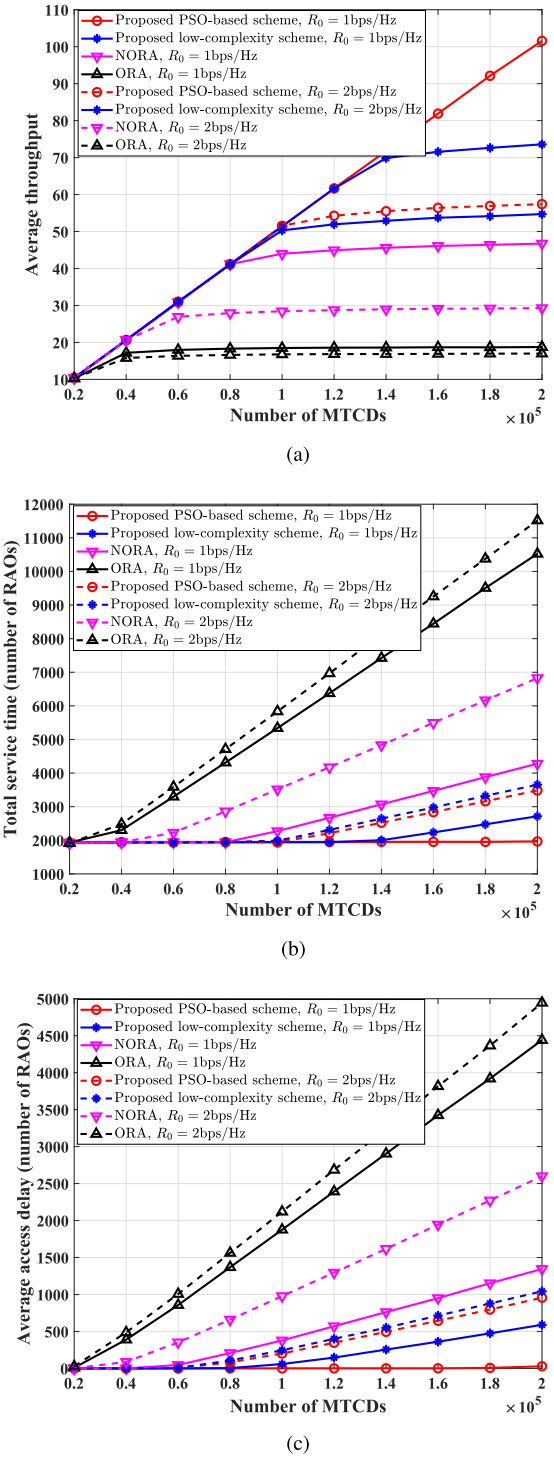


Fig. 5. Performance evaluations and comparisons versus the number of MTCDs N_m in our proposed PSO-based scheme, proposed low-complexity scheme, NORA scheme and ORA scheme for mMTC networks. (a) The average throughput. (b) The total service time. (c) The average access delay.

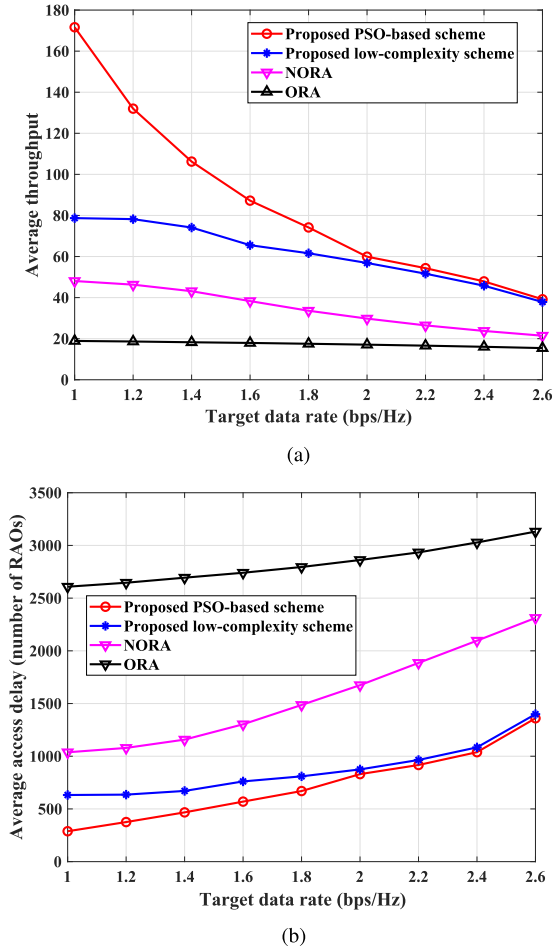


Fig. 6. Performance evaluations of our proposed PSO-based and low-complexity NORA schemes as well as the existing NORA and ORA schemes as the function of the target data rate R_0, where 10^5 MTCDs are activated simultaneously. (a) The average throughput. (b) The average access delay.

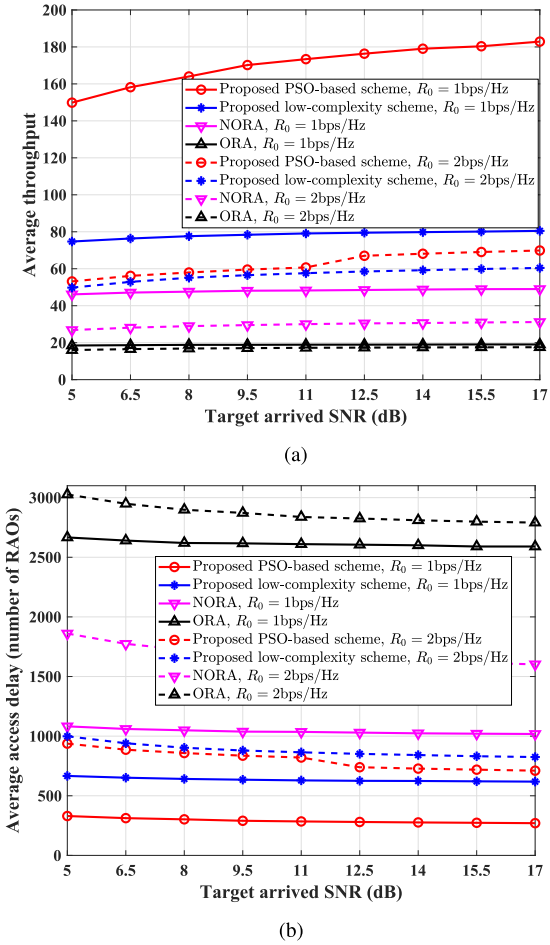


Fig. 7. Performance evaluations of our proposed PSO-based and low-complexity NORA schemes as well as the existing NORA and ORA schemes as the function of the target arrived SNR $\frac{P_y}{\sigma^2}$, where 10^5 MTCD are activated simultaneously. (a) The average throughput. (b) The average access delay.

i.e., $I^\star$ and I_{sub}, become the same when the target data rate is increased, which verifies the correctness of the conducted analysis under high data rate requirement.

Figure 5 shows the average throughput, total service time, and average access delay versus the number of MTCDs for the proposed PSO-based and low-complexity schemes, the NORA scheme, and the ORA scheme with different values of the target data rate R_0. We can observe from Fig. 5(a) that the average throughputs achieved by all the schemes increase with the number of MTCDs and become saturated if the number of MTCDs keeps increasing. However, the proposed PSO-based and low-complexity NORA schemes significantly outperform the existing NORA and ORA schemes when the number of MTCDs is large, especially under the low target data rate requirement, i.e., $R_0 = 1$ bps/Hz. As illustrated in Fig. 5(b), although the required total service time for all the schemes is an increasing function of the number of MTCDs, the proposed PSO-based and low-complexity schemes can also efficiently reduce the consumed number of RAOs for the given number of MTCDs. From Fig. 5(c), we can find that the proposed PSO-based and low-complexity schemes can also efficiently reduce the access delay compared with the existing NORA and ORA schemes. Moreover, as indicated by the solid red curve with circle in Fig. 5(a), the throughput achieved by the proposed PSO-based scheme under $R_0 = 1$ bps/Hz does not become saturated even when 2×10^5 MTCDs are activated simultaneously. Thus, the total service time and average access delay of our proposed PSO-based scheme remain unchanged, which are demonstrated by the solid red curve with circle in Figs. 5(b) and 5(c), respectively. In summary, Fig. 5 shows that both the proposed PSO-based and low-complexity schemes can efficiently improve the network performance as compared with the existing NORA and ORA schemes. The performance superiority mainly comes from that the proposed PSO-based and low-complexity schemes not only allow multiple MTCDs to share the same PA for access, but also can efficiently utilize the power-domain resources. Moreover, we also optimize the power back-off factor and the number of MTCDs sharing the same PUSCH such that the achievable average throughput can be improved. Furthermore, we can also observe from Fig. 5 that, although the proposed PSO-based scheme outperforms the low-complexity scheme, the low-complexity scheme can achieve near-PSO-based performance under the high target data rate requirement, i.e., $R_0 = 2$ bps/Hz, which is consistent

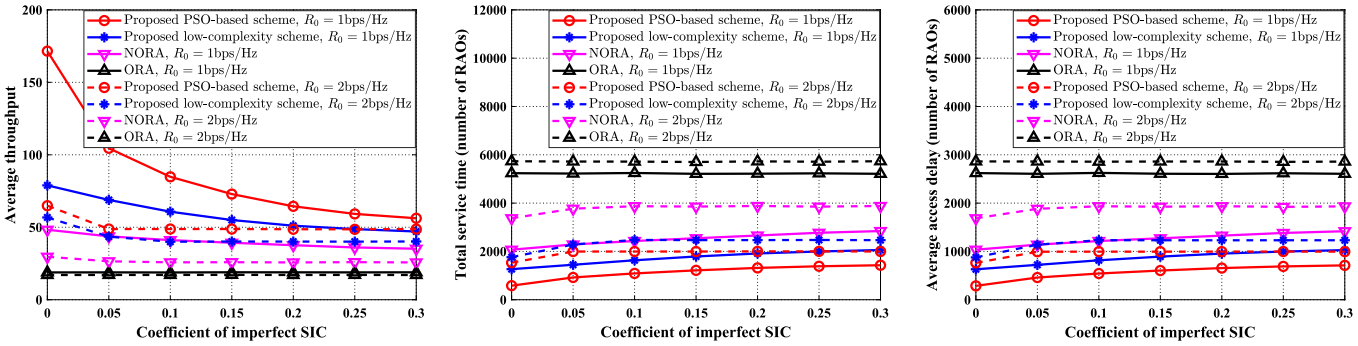


Fig. 8. Performance evaluations and comparisons versus the coefficient of imperfect SIC in our proposed PSO-based and low-complexity schemes, NORA scheme and ORA scheme, where 10^5 MTCDs are activated simultaneously. (a) The average throughput. (b) The total service time. (c) The average access delay.

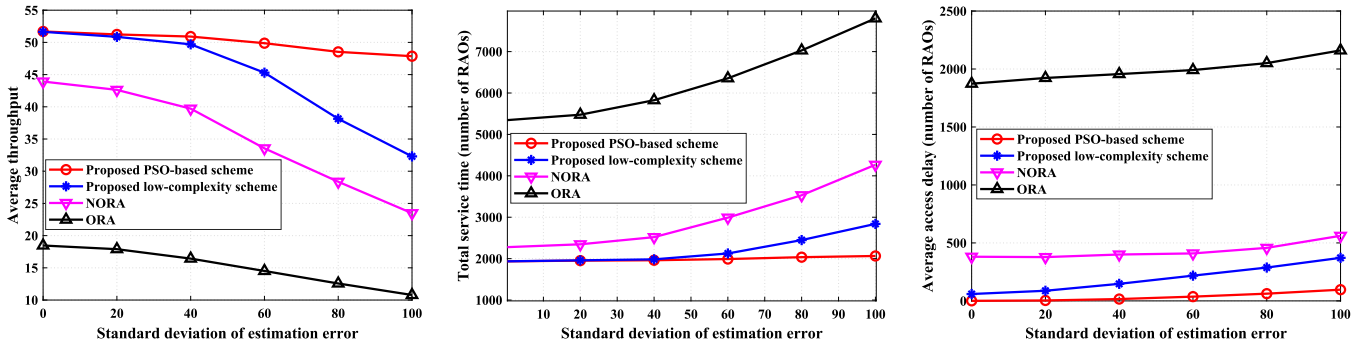


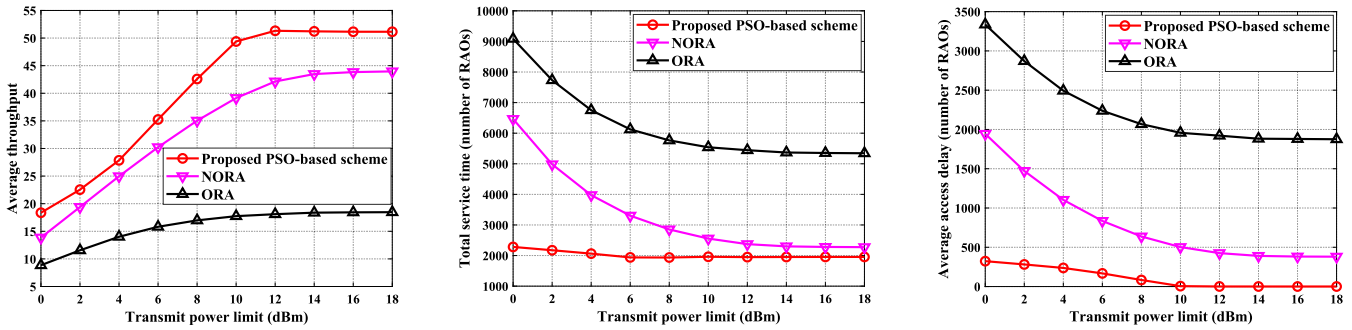
Fig. 9. Performance evaluations and comparisons versus the standard deviation of the estimation error σ_E in our proposed PSO-based scheme, proposed low-complexity scheme, NORA scheme and ORA scheme, where 10^5 MTCDs are activated following the Beta distribution with $T = 10$ s and the target data rate $R_0 = 1$ bit/s/Hz. (a) The average throughput. (b) The total service time. (c) The average access delay.

with our analysis in the previous sections and demonstrates that the proposed PSO-based scheme can achieve near-optimal performance.

Figure 6 shows the average throughput and the average access delay versus the target data rate for the proposed PSO-based and low-complexity schemes as well as the existing NORA and ORA schemes. We can observe from Fig. 6 that, although all schemes can only achieve lower throughput and incur larger access delay when the target data rate requirement becomes more stringent, i.e., the value of R_0 becomes larger, our proposed PSO-based and low-complexity schemes also outperform the NORA and ORA schemes. Moreover, the low-complexity scheme can achieve nearly the same performance compared with the PSO-based scheme under the high target data rate requirement.

Figure 7 illustrates the average throughput and the average access delay as the function of the target arrived SNR under our proposed PSO-based and low-complexity schemes as well as the existing NORA and ORA schemes. We can observe from Fig. 7 that, because higher target arrived SNR can increase the successful decoding probability of the SIC process, the proposed PSO-based and low-complexity schemes as well as the existing NORA scheme can achieve better performance when the target arrived SNR increases. Fig. 7 also shows that the proposed schemes can efficiently improve the network performance compared with the existing schemes and the developed low-complexity scheme can achieve near-PSO-based performance under the high target data rate requirement.

In the previous discussions, we assume that perfect SIC can be achieved, which usually cannot be realized in practice. Consequently, we need to analyze the network performance under the imperfect SIC characterized by the *coefficient of imperfect SIC* denoted by $\eta \in [0, 1]$ [46]. The coefficient of imperfect SIC implies that only $1 - \eta$ portion of interference can be removed if one MTCD successfully receives the data. Thus, the case when $\eta = 0$ refers to perfect SIC and $\eta = 1$ corresponds to the case that no interference can be removed, i.e., the fully imperfect SIC. Define event $A_i \triangleq \{\text{the signal of the } i\text{th MTCD is correctly decoded}\}$ $(i = 1, \cdots, I)$. Let $\chi_i = 1$ if event A_i occurs; otherwise, $\chi_i = 0$. Then, the SINR at the eNB of the ith MTCD becomes $\text{SINR}_i = \frac{p_{r,i}}{\sum_{m=i+1}^{I} p_{r,m} + \sum_{m=1}^{i-1} \eta p_{r,m} \chi_m + \sum_{m=1}^{i-1} p_{r,m}(1-\chi_m) + \sigma^2}$. Fig. 8 shows the average throughput, total service time, and average access delay versus the coefficient of imperfect SIC η for the proposed PSO-based scheme, proposed low-complexity scheme, NORA scheme, and ORA scheme. We can clearly observe that when the imperfect SIC is taken into consid-



(a) (b) (c)

Fig. 10. Performance evaluations and comparisons versus the maximum power limit $p_{\max}$ in our proposed PSO-based NORA scheme, the conventional NORA scheme, and ORA scheme, where 10^5 MTCDs are activated following the Beta distribution with $T = 10$ s, the target arrived SNR is 10 dB, and the target data rate $R_0 = 1$ bps/Hz. (a) The average throughput. (b) The total service time. (c) The average access delay.

eration, the performances of the proposed PSO-based and low complexity schemes as well as the conventional NORA scheme are degraded. However, the proposed schemes still outperform the conventional NORA and ORA schemes, which further verifies the superiority of the proposed schemes.

To evaluate the impact of imperfect load estimation on the network performance, we model the estimation error $E_{\text{est}} = \hat{N}_a - N_a$ as a Gaussian random variable having zero mean and variance σ_E^2, where $\hat{N}_a$ is the estimation of N_a. Based on the stochastic process theory, the square root σ_E is defined as the standard deviation, which provides a measure of dispersion of the random variable around the mean. Thus, σ_E can be viewed as the typical difference between N_a and $\hat{N}_a$, i.e., the estimation error E_{est}. Fig. 9 shows the average throughput, total service time, and average access delay versus the standard deviation σ_E for the proposed PSO-based and low-complexity schemes, the NORA scheme, and the ORA scheme. We can observe from Fig. 9 that the network performance degrades when the standard deviation σ_E increases, which implies that the estimation error will degrade the network performance. However, the proposed schemes still significantly outperforms the conventional NORA and ORA schemes.

In this paper, the average throughput of the proposed scheme is analyzed under the assumption of $p_{t,i} < p_{\max}$. However, the maximum allowed transmit power will definitely affect the network performance. Thus, we evaluate the impact of the transmit power limit through simulations. Specifically, we assume that MTCDs are uniformly distributed in the cell with radius 500 m. The noise power spectral density is -138.8 dBm/Hz and the bandwidth of each RB is set to 200 KHz. Fig. 10 shows various performances versus the maximum power limit $p_{\max}$ in the proposed PSO-based NORA scheme, the conventional NORA scheme, and ORA schemes, where 10^5 MTCDs are activated following the Beta distribution with $T = 10$ s, the target arrived $\frac{P_u}{\sigma^2} = 10$ dB, and the target data rate $R_0 = 1$ bps/Hz. We can clearly observe from Fig. 10 that larger average throughput, shorter total service time, and smaller average access delay can be achieved under large transmit power limit and thus small value of $p_{\max}$ or large path-loss will degrade the network performance. Moreover, the proposed NORA scheme still shows the performance superiority compared to the conventional NORA and ORA schemes.

VI. Conclusion

We proposed the throughput-oriented NORA scheme for mMTC networks. By employing the Sukhatme's classic theory and the characteristic function, we formulated the optimization problem which aims at maximizing the throughput subject to the constraints on the power back-off factor, the number of MTCDs in a NOMA group, and the successful transmission probability. Then, we efficiently solved the throughput maximization problem based on the PSO algorithm. Moreover, the low-complexity suboptimal scheme, which can achieve near-PSO performance under high data rate requirement, was also developed. Simulation results demonstrate the performance superiority of the proposed schemes.

Appendix A
Proof of Theorem 2

Proof: We denote the characteristic function of the exponential random variable Y_n by $\phi_{Y_n}(t)$. Then, we have $\phi_{Y_n}(t) = (1 - j2\ u^2\ t)^{-1}$. According to the characteristic function of the exponential random variable, the characteristic function of $a_{i,n}Y_n$, denoted by $\phi_{a_{i,n}Y_n}(t)$, can be expressed as $\phi_{a_{i,n}Y_n}(t) = (1 - j2\ u^2\ a_{i,n}t)^{-1}$. Since the characteristic function of a sum of independent random variables is equal to the product of individual characteristic functions, the characteristic function of $Z_{I,q,i}$, denoted by $\phi_{Z_{I,q,i}}(t)$, can be written as

$$\phi_{Z_{I,q,i}}(t) = \prod_{n=i}^{I} \phi_{a_{i,n}Y_n}(t) = \sum_{n=i}^{I} \alpha_{I,i,n}\phi_{a_{i,n}Y_n}(t) \quad (55)$$

where the coefficients $\alpha_{I,i,n}$ $(n = i, \cdots, I)$ are determined by (25) and the corresponding derivation can be followed from [43].

As the characteristic function is the Fourier transformation of the corresponding PDF, by performing the inverse Fourier transformation on $\phi_{a_{i,n}Y_n}(t)$, we have

$$\begin{aligned} & f_{a_{i,n}Y_n}(x) \\ &= \mathcal{F}^{-1}\{\phi_{a_{i,n}Y_n}(t)\} \\ &= \begin{cases} -\dfrac{1}{2u^2 a_{i,n}} \exp\left(-\dfrac{x}{2u^2 a_{i,n}}\right) \mathbb{I}\{a_{i,n} < 0\}, & x \le 0 \\ \dfrac{1}{2u^2 a_{i,n}} \exp\left(-\dfrac{x}{2u^2 a_{i,n}}\right) \mathbb{I}\{a_{i,n} > 0\}, & x > 0 \end{cases} \end{aligned} \quad (56)$$

where $f_{a_{i,n}Y_n}(x)$ denotes the PDF of $a_{i,n}Y_n$, $\mathcal{F}^{-1}\{\cdot\}$ represents the inverse Fourier transformation function, and $\mathbb{I}\{\cdot\}$ denotes the indicator function. Consequently, based on (55) and (56), we can obtain the PDF of $Z_{I,q,i}$, which is determined by (24). ■

Appendix B
Proof of Theorem 3

Proof: We use the mathematical induction to prove *Theorem 3*. Specifically, from the first equality of (35), we have $T_{PA}(1,q) = Q_1(1,q)$ and

$$\begin{aligned} T_{PA}(2,q) &= Q_1(2,q)\big(1 - Q_2(2,q)\big) + 2Q_1(2,q)Q_2(2,q) \\ &= Q_1(2,q) + Q_1(2,q)\cdot Q_2(2,q). \end{aligned} \tag{57}$$

Thus, we obtain that (35) holds when $I = 1$ and 2. Assume (35) holds when $I = n$. Then, we have

$$\begin{aligned} T_{PA}(n,q) &= \sum_{i=1}^{n}\left\{ i\cdot\prod_{t=1}^{i} Q_t(n,q)\Big[1 - Q_{i+1}(n,q)\Big]\right\} \\ &= \sum_{i=1}^{n}\left\{\prod_{t=1}^{i} Q_t(n,q)\right\}. \end{aligned} \tag{58}$$

Note that $T_{PA}(n+1,q)$ can be written as

$$\begin{aligned} &T_{PA}(n+1,q) \\ &= \sum_{i=1}^{n+1}\left\{ i\cdot\prod_{t=1}^{i} Q_t(n+1,q)\Big[1 - Q_{i+1}(n+1,q)\Big]\right\} \\ &= \sum_{i=1}^{n-1}\left\{ i\cdot\prod_{t=1}^{i} Q_t(n+1,q)\Big[1 - Q_{i+1}(n+1,q)\Big]\right\} \\ &\quad + n\prod_{t=1}^{n} Q_t(n+1,q) - n\prod_{t=1}^{n} Q_t(n+1,q) \\ &\quad + n\prod_{t=1}^{n} Q_t(n+1,q)\Big[1 - Q_{n+1}(n+1,q)\Big] \\ &\quad + (n+1)\prod_{t=1}^{n+1} Q_t(n+1,q). \end{aligned} \tag{59}$$

Then, according to (58), eq. (59) can be further expressed as

$$\begin{aligned} T_{PA}(n+1,q) &= \sum_{i=1}^{n}\left\{\prod_{t=1}^{i} Q_t(n+1,q)\right\} + \prod_{t=1}^{n+1} Q_t(n+1,q) \\ &= \sum_{i=1}^{n+1}\left\{\prod_{t=1}^{i} Q_t(n+1,q)\right\}. \end{aligned} \tag{60}$$

Based on the above analysis, we can conclude that if (35) is true when $I = n$, it also holds when $I = n+1$. Therefore, by the principle of mathematical induction, *Theorem 3* is proved. ■

Appendix C
Proof of Theorem 4

Proof: The probability $Q_i(I,q)$ given by (12) can be expressed as

$$Q_i(I,q) = \Pr\left\{{g_i}^2 \geq H_{PA}(i,I,q)\right\} \tag{61}$$

where $H_{PA}(i,I,q) = \beta\left[\Phi + \sum_{m=i+1}^{I}\left({g_m}^2 q^{m-1}\right)\right]/q^{i-1}$. Since $H_{PA}(i,I,q)$ increases when I increases, $Q_i(I,q)$ is a decreasing function of I. Then, based on (35), we have

$$\begin{aligned} T_{PA}(I,q) &\geq \sum_{i=1}^{I}\left\{\prod_{t=1}^{i} Q_t(I+1,q)\right\} \\ &\geq \sum_{i=1}^{I}\left\{\prod_{t=1}^{I+1} Q_t(I+1,q)\right\} = I\prod_{t=1}^{I+1} Q_t(I+1,q) \end{aligned} \tag{62}$$

and

$$\begin{aligned} T_{PA}(I+1,q) &= \sum_{i=1}^{I}\left\{\prod_{t=1}^{i} Q_t(I+1,q)\right\} + \prod_{t=1}^{I+1} Q_t(I+1,q) \\ &\leq \sum_{i=1}^{I}\left\{\prod_{t=1}^{i} Q_t(I,q)\right\} + \prod_{t=1}^{I+1} Q_t(I+1,q) \\ &= T_{PA}(I,q) + \prod_{t=1}^{I+1} Q_t(I+1,q). \end{aligned} \tag{63}$$

The second inequality of (62) comes from that the value of probability $Q_t(I+1,q)$ $(t = 1,\cdots,I+1)$ belongs to the interval $[0,1]$. Based on (62), we can obtain

$$\frac{T_{PA}(I,q)}{I} \geq \prod_{t=1}^{I+1} Q_t(I+1,q). \tag{64}$$

Then, according to (63) and (64), we can obtain $T_{PA}(I+1,q) - T_{PA}(I,q) \leq \prod_{t=1}^{I+1} Q_t(I+1,q) \leq \frac{T_{PA}(I,q)}{I}$. Thus, we have $I\cdot T_{PA}(I+1,q) - (I+1)\cdot T_{PA}(I,q) \leq 0$. Consequently, we can obtain $\frac{T_{PA}(I+1,q)}{I+1} - \frac{T_{PA}(I,q)}{I} \leq 0$. Based on the above analysis, we can conclude that $T_{PA}(I,q)/I$ is a decreasing function of I. ■

Appendix D
Proof of Theorem 5

Proof: Before proving *Theorem 5*, we define $A(I,q)$, $B(I,q)$, $\gamma(I,q)$, $V(I,q)$ and $\chi(I,q)$ as follows:

$$A(I,q) = \prod_{t=1}^{I}(I+1-t+t\beta q), \tag{65}$$

$$B(I,q) = 1 + \exp\left[-\psi\beta(I+1)/q^I\right], \tag{66}$$

$$\gamma(I,q) = B(I,q)/A(I,q), \tag{67}$$

$$V = \psi\beta(I+1)/q^I > 0, \tag{68}$$

$$\chi(I,q) = \prod_{t=1}^{I}\frac{I+1}{1 + \sum_{e=2}^{I+1-t}\left(\frac{I+1-t}{t\beta q^e} + \frac{1}{q^{e-1}}\right)^{-1}}. \tag{69}$$

As $q \geq 1/\beta$, we have $\beta q \geq 1$, then $\psi(I\beta q+1)/q^{I+1} = \psi\beta\left(I + \frac{1}{\beta q}\right)/q^I \leq \psi\beta(I+1)/q^I$. Since $V(I,q) > 0$, then we can get

$$\frac{2\psi(I\beta q+1)}{q^{I+1}} - e^{\psi\beta\frac{I+1}{q^I}} \leq 2V(I,q) - e^{V(I,q)} \leq 1. \tag{70}$$

Since the inequality $\frac{i}{I+1-i+i\beta q}=\frac{i}{I+1+(\beta q-1)\,i} \geq \frac{i}{I+1+(\beta q-1)\,I}=\frac{i}{1+I\beta q}$ $(i=1,\cdots,I)$ holds for $q \geq 1/\beta$, we can obtain

$$\begin{aligned}&\frac{\psi\,(I+1)\,I}{q^{I+1}\sum\limits_{i=1}^{I}\frac{i}{I+1-i+i\beta q}}-e^{\psi\beta\frac{I+1}{q^I}}\\&<\frac{\psi\,(I+1)\,I}{q^{I+1}\sum\limits_{i=1}^{I}\frac{i}{1+I\beta q}}-e^{\psi\beta\frac{I+1}{q^I}}\\&=\frac{2\psi\,(I\beta q+1)}{q^{I+1}}-e^{\psi\beta\frac{I+1}{q^I}}\leq 1.\end{aligned}\tag{71}$$

Then, we have

$$\psi\,(I+1)\,I\leq q^{I+1}\left(\mathrm{e}^{\psi\beta\frac{I+1}{q^I}}+1\right)\sum_{i=1}^{I}\frac{i}{I+1-i+i\beta q}\tag{72}$$

and thus

$$\frac{\partial B(I,q)}{\partial q}-B(I,q)\sum_{i=1}^{I}\frac{i\beta}{I+1-i+i\beta q}\leq 0.\tag{73}$$

Since the value of each multiplied factor in (65) is positive, we have

$$\frac{\partial A(I,q)}{\partial q}=A(I,q)\sum_{i=1}^{I}\frac{i\beta}{I+1-i+i\beta q}.\tag{74}$$

Therefore, based on (73) and (74), we can get $\frac{\partial\gamma(I,q)}{\partial q}=\left[\frac{\partial B(I,q)}{\partial q}A(I,q)-\frac{\partial A(I,q)}{\partial q}B(I,q)\right]/A^2(I,q) \leq 0$, which implies that $\gamma(I,q)$ decreases monotonously with q within $[1/\beta,1)$. Moreover, since $\gamma\,(I,q)$ and $\chi\,(I,q)$ are positive, and $\chi\,(I,q)$ is also a decreasing function of q within $[1/\beta,1)$, we can conclude that $F(I,q)$, which can be written as $F(I,q)=\gamma\,(I,q)\,\chi\,(I,q)$, and it is a decreasing function of q in $[1/\beta,1)$ for any given I. ■

APPENDIX E
PROOF OF THEOREM 6

Proof: Rewrite $T_{PA}\,(I+1,q)$ given by (50) as $T_{PA}(I+1,q)=\sum_{i=1}^{I}H_i(I,q)M_i(I,q)$, where $H_i(I,q)=X_i(I,q)$ $(i=1,\cdots,I-1)$ and $H_I(I,q)=F(I,q)$. Based on *Theorem 5*, we know that $H_i(I,q)$ $(i=1,\cdots,I)$ are decreasing functions of q. Define two power back-off factors q_1 and q_2, and assume that $q_1<q_2$, then we have

$$\begin{aligned}&T_{PA}\,(I+1,q_1)-T_{PA}\,(I+1,q_2)\\&=\sum_{i=1}^{I}\left[H_i(I,q_1)M_i(I,q_1)-H_i(I,q_2)M_i(I,q_2)\right]\\&\geq\sum_{i=1}^{I}H_i(I,q_2)\left[M_i(I,q_1)-M_i(I,q_2)\right]\\&\geq H^*(I,q)\left[T_{PA}(I,q_1)-T_{PA}(I,q_2)\right]\end{aligned}\tag{75}$$

where $H^*(I,q)=\mathrm{argmin}_{i=1,\ldots,I}\,H_i(I,q_2)$. As $H^*(I,q)>0$, we can prove that if $T_{PA}\,(I,q)$ is a decreasing function of q, then (75) is nonnegative. Thus $T_{PA}\,(I+1,q)$ is also a decreasing function of q. Moreover, since $T_{PA}(1,q)=e^{-\psi\beta}$ is unrelated to q and $T_{PA}(2,q)=2\cdot e^{-\psi\beta}/\,(1+\beta q)+2\cdot e^{-\psi\beta(1+2/q)}/\,(1+\beta q)$ is a decreasing function of q within $[1/\beta,1)$, we can conclude that $T_{PA}\,(I,q)$ decreases with q within the interval $[1/\beta,1)$. ■

REFERENCES

[1] *RAN Improvements for Machine-Type Communications*, document TR 37.868 V11.0.0, 3GPP, Oct. 2011.

[2] H. Shariatmadari *et al.*, "Machine-type communications: Current status and future perspectives toward 5G systems," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 10–17, Sep. 2015.

[3] T. Taleb and A. Kunz, "Machine type communications in 3GPP networks: Potential, challenges, and solutions," *IEEE Commun. Mag.*, vol. 50, no. 3, pp. 178–184, Mar. 2012.

[4] Q. Du *et al.*, "Massive access control aided by knowledge-extraction for co-existing periodic and random services over wireless clinical networks," *J. Med. Syst.*, vol. 40, no. 7, pp. 1–8, Jul. 2016.

[5] 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.

[6] H. M. Gürsu, M. Vilgelm, W. Kellerer, and M. Reisslein, "Hybrid collision avoidance-tree resolution for M2M random access," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 53, no. 4, pp. 1974–1987, Aug. 2017.

[7] M. S. Ali, E. Hossain, and D. I. Kim, "LTE/LTE-A random access for massive machine-type communications in smart cities," *IEEE Commun. Mag.*, vol. 55, no. 1, pp. 76–83, Jan. 2017.

[8] C. Bockelmann *et al.*, "Massive machine-type communications in 5G: Physical and MAC-layer solutions," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 59–65, Sep. 2016.

[9] L. Tello-Oquendo *et al.*, "Performance analysis and optimal access class barring parameter configuration in LTE-A networks with massive M2M traffic," *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 3505–3520, Apr. 2018.

[10] Z. Wang and V. W. S. Wong, "Optimal access class barring for stationary machine type communication devices with timing advance information," *IEEE Trans. Wireless Commun.*, vol. 14, no. 10, pp. 5374–5387, Oct. 2015.

[11] S. Duan, V. Shah-Mansouri, Z. Wang, and V. W. S. Wong, "D-ACB: Adaptive congestion control algorithm for bursty M2M traffic in LTE networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 12, pp. 9847–9861, Dec. 2016.

[12] Z. Wen and D. Lin, "Throughput optimization for massive random access of M2M communications in LTE networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Paris, France, May 2017, pp. 1–6.

[13] H. Wu, C. Zhu, R. J. La, X. Liu, and Y. Zhang, "FASA: Accelerated S-ALOHA using access history for event-driven M2M communications," *IEEE/ACM Trans. Netw.*, vol. 21, no. 6, pp. 1904–1917, Dec. 2013.

[14] H. He, Q. Du, H. Song, W. Li, Y. Wang, and P. Ren, "Traffic-aware ACB scheme for massive access in machine-to-machine networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, London, U.K., Jun. 2015, pp. 617–622.

[15] M. Tavana, V. Shah-Mansouri, and V. W. S. Wong, "Congestion control for bursty M2M traffic in LTE networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, London, U.K., Jun. 2015, pp. 5815–5820.

[16] S.-Y. Lien, T.-H. Liau, C.-Y. Kao, and K.-C. Chen, "Cooperative access class barring for machine-to-machine communications," *IEEE Trans. Wireless Commun.*, vol. 11, no. 1, pp. 27–32, Jan. 2012.

[17] M. Koseoglu, "Pricing-based load control of M2M traffic for the LTE-A random access channel," *IEEE Trans. Commun.*, vol. 65, no. 3, pp. 1353–1365, Mar. 2017.

[18] Y. Wu, N. Zhang, and G. Kang, "A new hybrid protocol for random access and data transmission based on two-phase ACB mechanisms for M2M communications," *Mobile Inf. Syst.*, vol. 2017, pp. 1–17, Apr. 2017.

[19] N. Zhang, G. Kang, J. Wang, Y. Guo, and F. Labeau, "Resource allocation in a new random access for M2M communications," *IEEE Commun. Lett.*, vol. 19, no. 5, pp. 843–846, May 2015.

[20] F. Ghavimi, Y. Lu, and H. Chen, "Uplink scheduling and power allocation for m2m communications in SC-FDMA-based LTE-A networks with QoS guarantees," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 6160–6170, Jul. 2017.

[21] W. Li, Q. Du, L. Liu, P. Ren, Y. Wang, and L. Sun, "Dynamic allocation of RACH resource for clustered M2M communications in LTE networks," in *Proc. IEEE Int. Conf. Identificat., Inf., Knowl. Internet Things*, Oct. 2015, pp. 140–145.

[22] Q. Du, W. Li, L. Liu, P. Ren, Y. Wang, and L. Sun, "Dynamic RACH partition for massive access of differentiated M2M services," *Sensors*, vol. 16, no. 4, pp. 1–19, Apr. 2016.

[23] F. Morvari and A. Ghasemi, "Two-stage resource allocation for random access M2M communications in LTE network," *IEEE Commun. Lett.*, vol. 20, no. 5, pp. 982–985, May 2016.

[24] 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, Sep. 2017.

[25] Z. Feng, Z. Feng, W. Li, and T. A. Gulliver, "An optimal service strategy for grouped machine-type communications in cellular networks," *IEEE Commun. Lett.*, vol. 21, no. 1, pp. 140–143, Jan. 2017.

[26] D. M. Amitu, R. N. Akol, and P. Nakeba, "QoS-aware splitting and radio resource allocation for machine type communications," in *Proc. IEEE Comput. Commun. Workshop Conf.*, Jan. 2018, pp. 941–947.

[27] T.-H. Chen, J.-W. Chang, and H.-Y. Wei, "Dynamic inter-channel resource allocation for massive M2M control signaling storm mitigation," in *Proc. IEEE Veh. Technol. Conf. (VTC-Fall)*, Montreal, QC, Canada, Sep. 2016, pp. 1–5.

[28] Z. Alavikia and A. Ghasemi, "Collision-aware resource access scheme for LTE-based machine-to-machine communications," *IEEE Trans. Veh. Technol.*, vol. 67, no. 5, pp. 4683–4688, May 2018.

[29] G. C. Madueño, S. Stefanović, and P. Popovski, "Efficient LTE access with collision resolution for massive M2M communications," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Austin, TX, USA, Dec. 2014, pp. 1433–1438.

[30] J. Yuan, H. Shan, A. Huang, T. Q. S. Quek, and Y.-D. Yao, "Massive machine-to-machine communications in cellular network: Distributed queueing random access meets MIMO," *IEEE Access*, vol. 5, pp. 2981–2993, 2017.

[31] H. S. Jang, S. M. Kim, H.-S. Park, and D. K. Sung, "An early preamble collision detection scheme based on tagged preambles for cellular M2M random access," *IEEE Trans. Veh. Technol.*, vol. 66, no. 7, pp. 5974–5984, Jul. 2017.

[32] Y. Zhang, K. Peng, Z. Chen, and J. Song, "SIC vs. JD: Uplink NOMA techniques for M2M random access," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Paris, France, May 2017, pp. 1–6.

[33] Z. Ding, Z. Yang, P. Fan, and H. V. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," *IEEE Signal Process. Lett.*, vol. 21, no. 12, pp. 1501–1505, Dec. 2014.

[34] Z. Ding, M. Peng, and H. V. Poor, "Cooperative non-orthogonal multiple access in 5G systems," *IEEE Commun. Lett.*, vol. 19, no. 8, pp. 1462–1465, Aug. 2015.

[35] L. Xiao, Y. Li, C. Dai, H. Dai, and H. V. Poor, "Reinforcement learning-based NOMA power allocation in the presence of smart jamming," *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 3377–3389, Apr. 2018.

[36] M. Akkouchi, "On the convolution of exponential distributions," *J. Chungcheong Math. Soc.*, vol. 21, no. 4, pp. 501–510, Dec. 2008.

[37] N. Zhang, J. Wang, G. Kang, and Y. Liu, "Uplink nonorthogonal multiple access in 5G systems," *IEEE Commun. Lett.*, vol. 20, no. 3, pp. 458–461, Mar. 2016.

[38] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[39] *Evolved Universal Terrestrial Radio Access (E-UTRA): Physical Layer Procedures*, document TS36.213 V15.6.0, 3GPP, Jun. 2019.

[40] *Evolved Universal Terrestrial Radio Access (E-UTRA); Medium Access Control (MAC) Protocol Specification*, document TS36.321 V15.6.0, 3GPP, Jun. 2019.

[41] Y. Liang, X. Li, J. Zhang, and Z. Ding, "Non-orthogonal random access for 5G networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4817–4831, Jul. 2017.

[42] H.-C. Yang and M.-S. Alouini, *Order Statistics in Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2011.

[43] J. W. Nilsson and S. A. Ridel, *Electric Circuits*. Upper Saddle River, NJ, USA: Prentice-Hall, 2000.

[44] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. IEEE Int. Conf. Neural Netw.*, vol. 4, Nov. 1995, pp. 1942–1948.

[45] "Smart grid systems and other radio systems suitable for utility operations and their long-term spectrum requirements," ETSI, Sophia Antipolis, France, Tech. Rep. TR 103 401 V1.1.1, Nov. 2016.

[46] *Smart Grid Systems and Other Radio Systems Suitable for Utility Operations and Their Long-Term Spectrum Requirements*, document TR 103 401 V1.1.1, ETSI, Nov. 2016.

[47] G. R. Gallager, *Stochastic Processes: Theory for Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2013.

Yichen Wang (S'13–M'14) received the B.S. degree in information engineering and the Ph.D. degree in information and communications engineering from Xi'an Jiaotong University, China, in 2007 and 2013, respectively. From August 2014 to August 2015, he worked as a Visiting Scholar with the Signal and Information Group, Department of Electrical and Computer Engineering, University of Maryland at College Park, College Park, USA. He is currently an Assistant Professor with the Information and Communications Engineering Department, Xi'an Jiaotong University. He has published more than 90 technical articles on international journals and conferences. His research interests include mobile wireless communications and networks with emphasis on cognitive radio techniques, ad hoc networks, MAC protocol design, statistical QoS provisioning technology for wireless networks, resource allocation over wireless networks, massive machine-type communications, ultra-reliable low-latency communications, device-to-device communications, and security-related technologies for wireless networks.

Dr. Wang is a member of the IEEE Communications Society and the IEEE Vehicular Technology Society. He received the Best Letter Award from IEICE Communications Society in 2010 and the Exemplary Reviewers Award from IEEE COMMUNICATIONS LETTERS in 2014. He has served as the TPC Co-Chair for IEEE VTC'16-Spring Workshop on User-Centric Networking for 5G and Beyond and the Track Co-Chair for Cloud Communications and Networking of CHINACOM'17. He also serves and has served as the Technical Program Committee Members for many world-renowned conferences including IEEE GLOBECOM, ICC, WCNC, VTC, and so on. He is currently serving as an Editor of *KSII Transactions on Internet and Information Systems*.

Tao Wang received the B.S. degree in information engineering from Xi'an Jiaotong University, China, in 2014. He is currently pursuing the M.S. degree with the Information and Communications Engineering Department, Xi'an Jiaotong University. From July 2014 to March 2017, he worked in Xi'an Thinkjoy Information Technology Co., Ltd., China. His current research interests include cognitive radio techniques, statistical QoS provisioning, and resource allocation in wireless communications.

Zihuan Yang received the B.S. degree in information engineering and the M.S. degree in information and communications engineering from Xi'an Jiaotong University, China, in 2016 and 2019, respectively. She is currently with ZTE Corporation, China. Her research interests include massive random access and resource allocation in machine-type communications. She received the Outstanding Graduate Cadres Award and Graduate Union's Advanced Individual Award from Xi'an Jiaotong University in 2018.

Dawei Wang (S'14–M'18) received the B.S degree from the University of Jinan, China, in 2011, and the Ph.D. degree from Xi'an Jiaotong University, China, in 2018. From 2016 to 2017, he was a Visiting Student with the School of Engineering, The University of British Columbia. He is currently an Associate Professor with the School of Electronics and Information, Northwestern Polythechnical University, Xi'an, China. He has served as a technical program committee (TPC) member for many International conferences, such as IEEE GLOBECOM, IEEE ICC, and so on. His research interests include physical-layer security, cognitive radio networks, cooperative communication, energy harvesting, and resource allocation.

Julian Cheng (S'96–M'04–SM'13) received the B.Eng. degree (Hons.) in electrical engineering from the University of Victoria, Victoria, BC, Canada, in 1995, the M.Sc.(Eng.) degree in mathematics and engineering from Queens University, Kingston, ON, Canada, in 1997, and the Ph.D. degree in electrical engineering from the University of Alberta, Edmonton, AB, Canada, in 2003. He was with Bell Northern Research and NORTEL Networks. He is currently a Full Professor with the School of Engineering, Faculty of Applied Science, The University of British Columbia, Kelowna, BC, Canada. His current research interests include digital communications over fading channels, statistical signal processing for wireless applications, optical wireless communications, and 5G wireless networks. He was the Co-Chair of the 12th Canadian Workshop on Information Theory in 2011, the 28th Biennial Symposium on Communications in 2016, and the 6th EAI International Conference on Game Theory for Networks (GameNets 216). He currently serves as an Area Editor for IEEE TRANSACTIONS ON COMMUNICATIONS. He was a past Associate Editor of IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE COMMUNICATIONS LETTERS, and IEEE ACCESS. He served as a Guest Editor for a Special Issue of IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS on Optical Wireless Communications. He is also a Registered Professional Engineer with the Province of British Columbia, Canada. He currently serves as the President for the Canadian Society of Information Theory and the Secretary for the Radio Communications Technical Committee of the IEEE Communications Society.