Priority Random Access and User Barring Design for NOMA-ALOHA in Heterogeneous mMTC

Wenbo Fan[1], Pingzhi Fan[1], and Zilong Liu[2]

[1]*Information Coding & Transmission Key Laboratory of Sichuan Province, CSNMT Int. Coop. Res. Centre (MoST), Southwest Jiaotong University, Chengdu, China*

[2]*School of Computer Science and Electronics Engineering, University of Essex, Colchester, U.K.*

***Abstract*—This paper presents a priority random access (PRA) with non-orthogonal multiple access (NOMA)-ALOHA (PRA-NA) scheme to enable access priority control for machine-type devices (MTDs) with delay-sensitive and delay-tolerant requirements. To enhance the energy efficiency of random access, we also introduce a channel quality-based user barring algorithm (CQ-UBA) that can alleviate the user overload problem and reduce the average transmit power of MTDs. By extending CQ-UBA to the proposed PRA-NA scheme, simulation results demonstrate that the random access performance in terms of access delay and energy efficiency for delay-sensitive MTDs can be significantly improved compared to the conventional NOMA-ALOHA.**

***Index Terms*—6G, random access, NOMA-ALOHA, priority, energy efficiency, overload traffic.**

I. Introduction

With the rapid growth of machine-type devices (MTDs) in the Internet of Things, massive machine-type communications (mMTC) have been identified as a critical use case for the next-generation mobile systems. Typically, mMTC is envisioned to support a large set of uncoordinated devices concurrently communicating over a shared wireless medium, whereby the data services are characterized by sparse and sporadic transmission with short packets [1], [2]. A challenge of such transmission is that the coordination of these randomly emitted signals may consume significant signalling overhead when using grant-based random access. To address this problem, grant-free random access has received significant research attention in recent years [3].

MTDs can transmit data to the evolved node base station (eNB) in an "arrive and go" manner in grant-free transmissions. There are several proposals for grant-free transmissions [4]–[7]. For example, in compressed sensing-based grant-free transmissions, the sparsity and sporadic nature of MTD data transmissions are leveraged to reformulate the multi-user detection problem as a sparse signal recovery problem, thereby allowing efficient multi-user transmissions [4], [5]. Besides, spreading-based non-orthogonal multiple access (NOMA) has also been studied for grant-free transmission [6], [7], where MTDs are required to transmit data according to the transmission pattern associated with the signature, allowing the receiver to successfully perform multi-user detection. However, the aforementioned grant-free transmission schemes need to construct a preamble superset to support massive connectivity from the physical layer perspective.

From the medium access control (MAC) layer aspect, a recent research direction for grant-free transmission is the integration of slotted ALOHA with power-domain NOMA, yielding NOMA-ALOHA. The NOMA-ALOHA scheme can enhance system throughput by using successive interference cancellation (SIC) to decode multiple data packets with specific received power levels within a single slot. Firstly, [8] modeled the received power levels and analyzed the throughput of multichannel NOMA-ALOHA scheme. [9] considered two different packet arrival models (i.e., binomial and Poisson arrivals), and analyzed the throughput of NOMA-ALOHA under the impact of power collisions. Furthermore, the throughput bounds of NOMA-ALOHA were studied under the assumption of perfect SIC [10], [11] and imperfect SIC [12].

There are also several studies focused on the energy efficiency of NOMA-ALOHA systems. To reduce the average transmit power, [8] proposed a channel-dependent (CD) power level selection scheme by setting corresponding channel gain thresholds for different received power levels. In [13], a new received power model was designed by ignoring the impact of multi-user interference in the SIC process, thereby reducing the transmit power. However, these schemes may not be feasible in the case of user overload. To address this problem, [9] and [14] attempted to mitigate user overload by controlling network load based on the optimal load derived from the throughput bound expression. Recently, priority random access (PRA) schemes have been extensively studied in grant-based random access for heterogeneous mMTC [15]–[17]. These PRA schemes are typically built by assigning different access resource groups to different categories. In [15], an online control algorithm was introduced to adaptively adjust the sizes of resource groups. [16] and [17] supported access priority by adaptively adjusting the access class barring factors for MTDs with different priorities. Furthermore, [18] designed a PRA scheme from the perspective of access resource quality, where orthogonal preambles for high-priority MTDs achieve a higher detection probability compared to non-orthogonal preambles used by low-priority MTDs.

The aforementioned NOMA-ALOHA schemes primarily focus on the throughput under the assumption of a single type of MTD in the network. However, in practical systems, different types of coexisting MTDs may simultaneously transmit data to the eNB, rendering current NOMA-ALOHA schemes [8]–[12] less effective. Therefore, it is essential to design priority-based random access for the NOMA-

ALOHA scheme to accommodate heterogeneous mMTC with different delay requirements. Additionally, NOMA-ALOHA schemes are affected by user overload, which can significantly deteriorate throughput performance. Although random access control can alleviate this issue, it may result in high transmit power. For example, when MTDs with poor channel quality select a higher received power level, they may require a high transmit power. Against the above backdrop, we aim to achieve priority-based random access for NOMA-ALOHA in heterogeneous mMTC, while effectively alleviating the user overload problem and reducing energy consumption. The main contributions are summarized as follows:

- We propose a novel PRA-based NOMA-ALOHA (PRA-NA) scheme to provide access priority for heterogeneous MTDs associated with different delay requirements. By utilizing the error propagation in SIC, the core idea of the proposed PRA-NA scheme is to provide access priority by assigning different power levels with different reliability transmission qualities to different types of MTDs.
- We design a channel quality-based user barring algorithm (CQ-UBA) to alleviate the user overload problem and reduce the average transmit power, where the channel quality threshold is used as the barring factor to control the network load. To this end, we analyze the optimal channel quality threshold to stabilize the system throughput and improve energy efficiency. Then, the CQ-UBA is extended to the proposed PRA-NA scheme to provide access priority.
- Simulation results are presented using the 3GPP traffic model based on the Beta distribution (3GPP TR 37.868) [19]. It is demonstrated that the proposed schemes are capable of providing access priority, mitigating the user overload problem, and reducing average transmit power.

The remainder of this paper is organized as follows. The system model and traffic model of NOMA-ALOHA are introduced in Section II. The PRA-NA scheme with CQ-UBA is proposed and analyzed in Section III. Simulation results are presented in Section IV. Finally, this paper is concluded in Section V.

II. System Model

Let us consider a slotted-ALOHA uplink transmission network, which comprises M MTDs that communicate with a single eNB. It is assumed that each MAC frame has a duration of τ seconds and comprises S time slots. Based on the channel reciprocity in the time division duplexing (TDD) mode, let us assume that the channel coefficient can be estimated at the active MTDs using the well-known pilot-based channel estimation technique.

A. Uplink NOMA-ALOHA Transmission

By performing SIC techniques at the receiver, the uplink NOMA-ALOHA scheme can support data transmission from multiple MTDs within a time slot. We assume perfect synchronization, let λ denote the number of active MTDs in each MAC frame, and the active MTDs randomly select a time slot for transmission at the beginning of each MAC frame. In this case, the received signal at the eNB can be expressed as

$$y = \sum_{n=1}^{N} h_n \sqrt{\rho_n} s_n + n_0, \tag{1}$$

where N denotes the number of active MTDs in a time slot, s_n and ρ_n represent the signal and transmit power of MTD n, respectively. h_n represents the channel coefficient between the n-th MTD and eNB. $n_0 \sim \mathcal{CN}(0, \sigma^2)$ denotes additive white Gaussian noise where σ^2 is the noise power. In this paper, the noise power is assumed to be normalized without loss of generality.

Assume that there are Q preset power levels in a time slot for NOMA-ALOHA, yielding

$$r_1 > r_2 > \cdots > r_q > \cdots > r_Q > 0, \tag{2}$$

where r_q represents the received power value at the eNB for the MTD n that selects the q-th power level. Define the target signal-to-interference-plus-noise ratio (SINR) for all MTDs as Γ to avoid the computational and signalling overhead associated with dynamically adjusting parameters for each MTD in mMTC scenarios. In order to successfully decode the signal at all power levels, there should have only one packet at each power level. In this case, the receive power at the q-th power level can be written as [9], [10]

$$r_q = \Gamma(\Gamma+1)^{Q-q}, \tag{3}$$

where the received power values of all power levels can be stored in MTDs. When an active MTD attempts to transmit data, it will randomly select a slot and a power level. The corresponding received power value can be mapped based on the selected power level, and the transmit power can be adjusted by $\rho_n = r_q / |h_n|^2$ to achieve the received power of r_q at the eNB. To apply this power model, it is assumed that the signal can be perfectly removed with ideal SIC.

B. MTDs Traffic Model

In this paper, a typical Beta traffic model proposed by 3GPP is considered within the duration of T MAC frames. As the duration of each MAC frame is τ, the number of newly activated MTDs at frame i is given by

$$v_i = M \int_{(i-1)\tau}^{i\tau} f(t)\, dt, i = 1, 2, \cdots, T/\tau, \tag{4}$$

where M is the total number of MTDs in the network, and $f(t)$ is the random access intensity described by the probability density function [19]:

$$f(t) = \frac{t^{\alpha-1}(T-t)^{\beta-1}}{T^{\alpha+\beta-1}\boldsymbol{B}(\alpha,\beta)}, t \in (0, T), \tag{5}$$

where $\boldsymbol{B}(\cdot)$ represents the Beta function and ($\alpha = 3$, $\beta = 4$) are considered in this paper. An MTD that failed to transmit in the previous frame and requires retransmission in the current frame is referred to as a backlogged active MTD. The number of active MTDs in a frame (i.e., λ) is the sum of the number of newly active MTDs (i.e., v_i) and the number of backlogged active MTDs.

Similar to [16], [18], the MTDs in heterogeneous mMTC can be classified into two types based on their delay requirements: delay-sensitive MTDs and delay-tolerant MTDs. For example, delay-sensitive MTDs (e.g., vehicular sensors and public safety devices) are classified as high-priority due to their requirement for a high transmission success rate and low access delay. For delay-tolerant MTDs (e.g., smart meters, etc.), the MTDs report data periodically to the eNB with less stringent access delay constraints (e.g., half an hour). Thus, delay-tolerant MTDs can be regarded as low-priority MTDs. Let μ denote the proportion of high-priority MTDs, then the total number of high-priority and low-priority MTDs in the network can be represented as μM and $(1-\mu)M$, respectively. Generally, the proportion of high-priority devices is small.

III. PRA-NA scheme with CQ-UBA for Uplink Transmission

In NOMA-ALOHA, if two or more active MTDs select the same power level in a slot, power collision occurs, resulting in transmission failure for all active MTDs in this power level. Moreover, due to the error propagation in SIC technology used in NOMA-ALOHA, when power collision occurs at power level q, the signals at power levels $q' > q$ will also experience decoding failure. However, when power collisions occur at the power levels with lower received power (e.g., r_3 and r_4), it leads to two different consequences on the decoding at the power levels with higher received power. For example, let system parameters $Q = 4$ and $\Gamma = 2$. From (3), the received power values of all power levels can be obtained as $\{r_1, r_2, r_3, r_4\} = \{54, 18, 6, 2\}$. When $\{N_1, N_2, N_3, N_4\} = \{1, 1, 2, 0\}$, where N_q is defined as the number of active MTDs choosing the q-th power level. The SINRs of active MTDs that select r_1 and r_2 are given by $\frac{r_1}{r_2+2r_3+1} = \frac{54}{31} \approx 1.74$ and $\frac{r_2}{r_1+2r_3+1} = \frac{18}{67} \approx 0.26$, both of which are lower than the target SINR $\Gamma = 2$. Consequently, decoding failure may occur at power levels with higher received power. However, when $\{N_1, N_2, N_3, N_4\} = \{1, 1, 0, 2\}$, the SINRs of active MTDs selecting r_1 and r_2 become $\frac{r_1}{r_2+2r_4+1} = \frac{54}{23} \approx 2.34$ and $\frac{r_2}{2r_4+1} = \frac{18}{5} \approx 3.6$, both exceeding the target SINR Γ. In this scenario, signals at power levels with higher received power can be successfully decoded despite power collisions occurring at lower received power levels. Since all active MTDs independently select power levels according to an identical probability distribution (i.e., uniform distribution), the probability of power collisions occurring at each power level remains the same. Consequently, transmissions at power levels with higher received power might be more reliable than transmissions at power levels with lower received power.

Based on the above analysis, we propose a novel PRA-NA scheme to provide access priority, as illustrated in Fig. 1. Firstly, in heterogeneous mMTC, MTDs are categorized into high-priority MTDs and low-priority MTDs according to their delay requirements. Subsequently, the power levels are divided into two groups, including a high-power group (e.g., r_1 and r_2) and a low-power group (e.g., r_3 and r_4). High-priority MTDs select power levels from the high-power group, while power levels from the low-power group are utilized to serve low-priority MTDs. Given the fact that the transmissions on power levels in the high-power group are more reliable than those in the low-power group, the retransmission probability for high-priority MTDs can be reduced. Consequently, the proposed PRA-NA scheme is capable of enhancing the access delay performance for high-priority MTDs.

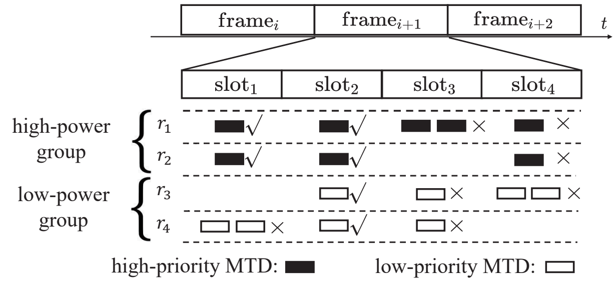

Fig. 1. The proposed PRA-NA scheme.

A. Throughput on PRA-NA Scheme

In this paper, throughput is defined as the average number of MTDs successful transmission per slot (i.e., packets that can be successfully decoded at the eNB). As previously analysis, when power collisions occur at lower received power levels, the decoding outcomes at higher received power levels may exhibit two distinct scenarios (i.e., successful decoding or decoding failure). As N and Q increase, these two different outcomes complicate the analysis of the probability of successfully decoded packets at each power level, making it challenging to derive an exact throughput expression. To address this issue, similar to [9], [10], [12], we assume that power collisions at lower received power levels do not affect the decoding of signals at higher received power levels. Under this assumption, the derived throughput can serve as an upper bound for the system throughput. Although this assumption may be difficult to achieve in practice, the optimal load derived based on this throughput upper bound has been validated as effective in existing studies. In this case, we consider a single time slot with N active MTDs and Q power levels, where the number of power levels in the high-power group and the low-power group is equal. Let η denote the throughput of the PRA-NA scheme. we have

$$\eta = \eta_h + \eta_l, \tag{6}$$

where η_h and η_l represent the average number of MTDs with successful transmission in the high-power group and low-power group, respectively. Denoting the successful access probability of MTDs in the high-power group as $P_a(N_q; q \leq Q/2)$, hence η_h can be expressed as

$$\eta_h = \sum_{q=1}^{Q/2} \sum_{N_q=1}^{\mu N} N_q P_a(N_q; q \leq Q/2), \tag{7}$$

where N_q is the number of MTDs selecting the q-th power level. Let a and b be the probabilities of $N_q = 0$ and $N_q = 1$ in the high-power group, we can obtain

$$P_a(N_q; q \leq Q/2) = \begin{cases} 0 & if\ N_q > 1 \\ b(a+b)^{q-1} & if\ N_q = 1 \end{cases}, \quad (8)$$

where N_q follows the binomial distribution, and we have

$$a = P(N_q = 0) = \left(1 - \frac{2}{Q}\right)^{\mu N}$$
$$b = P(N_q = 1) = \frac{2\mu N}{Q}\left(1 - \frac{2}{Q}\right)^{\mu N - 1}. \quad (9)$$

By substituting (9) and (8) into (7), the average number of MTDs with successful transmission in the high-power group of the proposed FPRA-NA scheme can be obtained as

$$\begin{aligned} \eta_h =& \sum_{q=1}^{Q/2} b(a+b)^{q-1} \\ =& b\frac{1-(a+b)^{Q/2}}{1-(a+b)}. \end{aligned} \quad (10)$$

Similarly, denote by $P_a(N_q; q > Q/2)$ the successful access probability of MTDs in the low-power group. The average number of MTDs with successful transmission in the low-power group can be expressed as

$$\eta_l = \sum_{q=1+Q/2}^{Q} \sum_{N_q=1}^{(1-\mu)N} N_q P_a(N_q; q > Q/2). \quad (11)$$

Let c and d be the probabilities that $N_q = 0$ and $N_q = 1$ in the low-power group. Then, $P_a(N_q; q > Q/2)$ can be expressed as

$$P_a(N_q; q > \frac{Q}{2}) = \begin{cases} 0 & if\ N_q > 1 \\ d(a+b)^{\frac{Q}{2}}(c+d)^{q-1-\frac{Q}{2}} & if\ N_q = 1 \end{cases}. \quad (12)$$

Similar to (9), we have

$$c = P\left(N_q = 0\right) = \left(1 - \frac{2}{Q}\right)^{(1-\mu)N}$$
$$d = P(N_q = 1) = \frac{2(1-\mu)N}{Q}\left(1 - \frac{2}{Q}\right)^{(1-\mu)N-1}. \quad (13)$$

By substituting (13) and (12) into (11), the average number of MTDs with successful transmission in the low-power group of the proposed FPRA-NA scheme can be obtained as

$$\begin{aligned} \eta_l =& \sum_{q=1+Q/2}^{Q} d(a+b)^{Q/2}(c+d)^{q-1-Q/2} \\ =& d\left(a+b\right)^{Q/2}\frac{1-(c+d)^{Q/2}}{1-(c+d)}. \end{aligned} \quad (14)$$

B. CQ-UBA for PRA-NA Scheme

In the previous sections, we assumed that MTDs have perfect knowledge of the channel coefficients between themselves and the eNB, which allows them to adjust their transmit power based on the channel quality extracted from these coefficients.

Consider a scenario where MTDs are uniformly distributed within a cell of radius R. The channel quality with large-scale fading for MTD n is given by [8]

$$\mathbb{E}[|h_n|^2] = A_0 d_n^{-\kappa}, \quad 0 < d_n \leq R \quad (15)$$

where A_0 is a constant, d_n represents the distance between the MTD n and eNB, and κ is the path loss exponent. Let P^* denote the optimal channel quality threshold, we have

$$P^* = A_0 d_*^{-\kappa}, \quad (16)$$

where d_* represents the optimal distance. Active MTDs with channel quality higher than the optimal channel quality threshold are allowed to transmit data, and these MTDs can be grouped as follows:

$$\mathcal{A} = \{n \mid d_n \leq d_*\}. \quad (17)$$

Here, $|\mathcal{A}|$ represents the number of the MTDs permitted to transmit data in a frame. Since MTDs are uniformly distributed within a cell, the optimal distance can be written as

$$d_* = R\sqrt{\frac{\lambda_*}{\lambda}}, \quad (18)$$

where λ_* is the optimal number of MTDs allowed to transmit data in a frame, and λ is the number of active MTDs in a frame within the cell of radius R. The optimal load based on the upper-bound throughput may unstabilize the NOMA-ALOHA system, because the load may be increased to meet the throughput based on an upper-bound throughput, which could lead to system overload. In this sense, the optimal load based on lower-bound throughput could be useful. Thus, the optimal load derived from the lower-bound throughput is used to analyze the optimal channel quality threshold, where the optimal load is equal to $\sqrt{Q}$ for a single slot with Q power levels [10]. As there are S slots in a frame in this paper, the optimal load λ_* can be expressed as $S\sqrt{Q}$ in a frame.

To calculate the optimal distance, we adopt the method from [9], assuming that idle power levels can be observed during the decoding process and estimating the number of active MTDs in the next frame based on the current load $\tilde{\lambda}$. Let $\tilde{N}_s$ denote the number of MTDs in set $\mathcal{A}$ during the the s-th slot. Denote $\tilde{N}_s$ as the number of MTDs in the group $\mathcal{A}$ for the s-th slot, The estimated total number of MTDs allowed to transmit data in the current frame (i.e., $\tilde{\lambda_*}$) could be obtained by $\tilde{\lambda_*} = |\mathcal{A}| = \sum_{s=1}^{S} \tilde{N}_s$. According to (9), the expected number of idle power levels in the s-th slot Q_{idle}^s can be written as

$$\mathbb{E}[Q_{idle}^s] = Q\left(1 - \frac{1}{Q}\right)^{\tilde{N}_s}. \quad (19)$$

Algorithm 1: PRA-NA with CQ-UBA

Input: S, T, Q, R.
1 Initialize the optimal distance $(d_{*,h},\ d_{*,l})$ and optimal channel quality threshold $(P_h^*,\ P_l^*)$ for different priority MTDs.
2 **for** $t = 1, 2, ..., T$ **do**
3 Active MTDs with different priorities compare its channel quality with the corresponding channel quality threshold $(P_h^*,\ P_l^*)$ and perform CQ-UBA to form $\mathcal{A}$;
4 Active MTDs with different priorities in $\mathcal{A}$ randomly choose a slot and select a power level to access from the corresponding power level group;
5 **for** $s = 1, 2, ..., S$ **do**
6 Perform SIC at the receiver
7 **if** *power level* $< Q/2$ **then**
8 Count the number of idle power levels in the s-th slot for high-power group, i.e., $Q_{idle}^{s,h}$;
9 **else**
10 Count the number of idle power levels in the s-th slot for low-power group, i.e., $Q_{idle}^{s,l}$;
11 **end**
12 Calculate the number of different priority MTDs in $\mathcal{A}$ according to (19), i.e., $\tilde{N}_{s,h}, \tilde{N}_{s,l}$;
13 **end**
14 Calculate the current load according to (20), i.e., $\tilde{\lambda}_h, \tilde{\lambda}_l$;
15 Update the optimal distance according to (18), i.e., $d_{*,h}, d_{*,l}$;
16 Update the optimal channel quality threshold according to (16), i.e., P_h^*, P_l^*;
17 **end**
18 Return the optimal distance $(d_{*,h},\ d_{*,l})$ and optimal channel quality threshold $(P_h^*,\ P_l^*)$ for next traffic arrival.

As $\tilde{N}_s$ increases, Q_{idle}^s monotonically decreases. Therefore, $\tilde{N}_s$ can be estimated based on the observed Q_{idle}^s. Using this approach, the number of active MTDs in the next frame could be approximated as

$$\lambda \approx \tilde{\lambda} = \left(\frac{R}{\tilde{d}_*}\right)^2 \sum_{s=1}^{S} \tilde{N}_s, \tag{20}$$

where $\tilde{d}_*$ represents the current optimal distance. According to the estimated load λ and optimal load $\lambda_* = S\sqrt{Q}$, the optimal distance and the optimal channel quality threshold for the next frame could be calculated based on (18) and (16), respectively. Prior to uplink transmission, active MTDs will receive system information containing the channel quality threshold to perform the CQ-UBA operation. MTDs with channel quality below this channel quality threshold are prohibited from transmitting data to the eNB. This CQ-UBA mechanism effectively mitigates system overload by restricting data transmission from MTDs that require high transmission power.

Finally, the PRA-NA scheme is integrated with CQ-UBA as shown in **Algorithm 1**. Steps 2-4 of **Algorithm 1** show that MTDs perform the PRA-NA scheme in conjunction with CQ-UBA. Steps 5-18 detail the update process of the optimal channel quality threshold and optimal distance for the next frame.

IV. Simulation Results and Discussions

In this section, we present simulation results to compare the performance of the proposed scheme with the existing

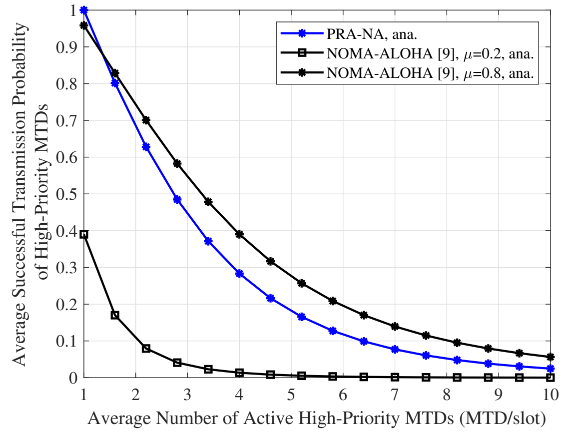

Fig. 2. The successful access probability of high-priority MTDs versus the number of active high-priority MTDs with $Q = 8$.

NOMA-ALOHA scheme in terms of access priority, average transmit power, and average access delay. To compare fairly with existing schemes, we assume that $R = 1$ and $A_0 = 1$ in (15) for normalization purposes [8].

A. Average Successful Transmission Performance

Fig. 2 plots the average successful transmission probability of high-priority MTDs for the proposed PRA-NA scheme and the existing NOMA-ALOHA scheme. The average probability of successful transmission is defined as the ratio of the number of MTDs with successful transmission to the number of active MTDs. It can be observed that when the number of high-priority MTDs in the network is small (i.e., $\mu = 0.2$), the proposed PRA-NA scheme outperforms the conventional NOMA-ALOHA scheme. Since the PRA-NA scheme provides more reliable transmission for high-priority MTDs. However, when the network experiences high traffic from high-priority MTDs (i.e., $\mu = 0.8$), a relatively rare scenario, the average successful transmission probability of high-priority MTDs in the PRA-NA scheme is slightly lower than that in the NOMA-ALOHA scheme. This is due to the fixed number of power levels in each power group. If the number of power levels in each power group could be dynamically adjusted, the average successful access probability of high-priority MTDs in the PRA-NA scheme would be further improved.

B. Average Transmit Power Performance

Fig. 3 illustrates the average transmit power of the proposed PRA-NA scheme with CQ-UBA, compared to the NOMA-ALOHA scheme with conventional UBA and CD. It can be observed that the average transmit power increases as the target SINR Γ increases, since higher target SINR requires higher received power according to (3). Moreover, the average transmit power of the proposed PRA-NA scheme with CQ-UBA is lower than that of the NOMA-ALOHA schemes with UBA and CD, highlighting the significant advantage of CQ-UBA in improving energy efficiency.

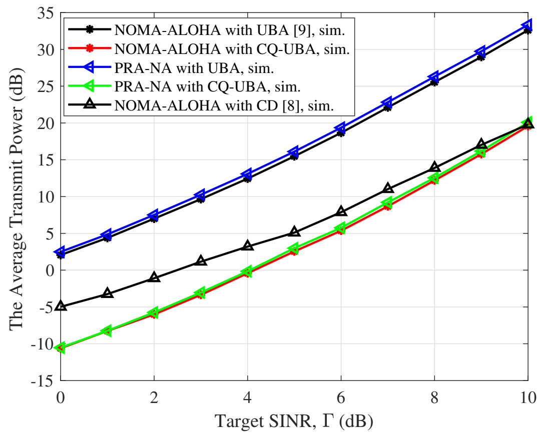

Fig. 3. The average transmit power versus the target SINR Γ, $Q = 4$, $S = 8$, $\mu = 0.5$, and $M = 4000$.

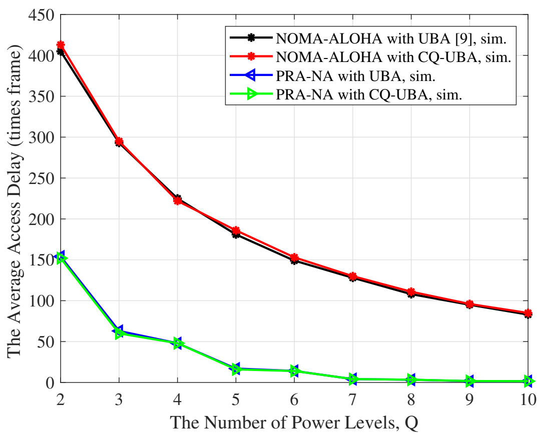

Fig. 4. The average access delay of high-priority MTDs versus the number of power level Q, $\mu = 0.2$, $S = 8$, $\Gamma = 3$ dB, and $M = 4000$.

C. Average Access Delay Performance

Fig. 4 shows that as Q increases, the access delay of high-priority MTDs gradually decreases in both the proposed PRA-NA scheme and the NOMA-ALOHA scheme. Meanwhile, the proposed PRA-NA scheme consistently outperforms the NOMA-ALOHA scheme in terms of the average access delay for high-priority MTDs. Since high-priority MTDs are assigned more reliable transmission resources in the PRA-NA scheme, the access delay of high-priority MTDs is effectively reduced. Furthermore, the performance of the schemes using CQ-UBA and UBA is very close, demonstrating that the proposed CQ-UBA is as effective as UBA in alleviating user overload problem.

V. CONCLUSION

In this paper, we proposed a novel PRA-NA scheme to provide access priority for heterogeneous mMTC. Moreover, CQ-UBA was developed to improve energy efficiency and alleviate the user overload problem. By integrating CQ-UBA with the proposed PRA-NA scheme, simulation results based on the 3GPP Beta traffic demonstrate the superiority of the proposed scheme over the conventional NOMA-ALOHA.

ACKNOWLEDGEMENTS

This work was supported by NSFC under Project 62020106001. The work of Z. Liu was supported in part by the UK Engineering and Physical Sciences Research Council under Grant EP/Y000986/1 ('SORT'). The work of W. Fan was supported in part by the CSC scholarship under No. 202407000042.

REFERENCES

[1] Z. Xiang, W. Yang, Y. Cai, Z. Ding, Y. Song, and Y. Zou, "NOMA-assisted secure short-packet communications in IoT," *IEEE Wireless Commun.,*, vol. 27, no. 4, pp. 8–15, Aug. 2020.

[2] L. Yang and P. Fan, "Improved sparse vector code based on optimized spreading matrix for short-packet in URLLC," *IEEE Wireless Commun. Lett.*, vol. 12, no. 4, pp. 728–732, Apr. 2023.

[3] J. Ding, M. Nemati, C. Ranaweera, and J. Choi, "IoT connectivity technologies and applications: A survey," *IEEE Access*, vol. 8, pp. 67 646–67 673, Apr. 2020.

[4] J. Choi, "NOMA-based compressive random access using gaussian spreading," *IEEE Trans. Commun.*, vol. 67, no. 7, pp. 5167–5177, Jul. 2019.

[5] P. Gao, Z. Liu, P. Xiao, C. H. Foh, and J. Zhang, "Low-complexity block coordinate descend based multiuser detection for uplink grant-free NOMA," *IEEE Trans. Veh. Technol.*, vol. 71, no. 9, pp. 9532–9543, Sep. 2022.

[6] L. Yang, P. Fan, L. Li, Z. Ding, and L. Hao, "Two-stage active user detection with false alarm correction for GF-NOMA system," *IEEE Trans. Wireless Commun.*, early access, Jan. 2024, doi: 10.1109/TWC.2024.3354956.

[7] K. Lai, Z. Liu, J. Lei, G. Chen, P. Xiao, and L. Wen, "Sparse code multiple access with enhanced K-repetition scheme: Analysis and design," *IEEE Trans. Wireless Commun.*, early access, Oct. 2023, doi: 10.1109/TWC.2023.3325397.

[8] J. Choi, "NOMA-based random access with multichannel ALOHA," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 12, pp. 2736–2743, Dec. 2017.

[9] W. Yu, C. H. Foh, A. U. Quddus, Y. Liu, and R. Tafazolli, "Throughput analysis and user barring design for uplink NOMA-enabled random access," *IEEE Trans. Wireless Commun.*, vol. 20, no. 10, pp. 6298–6314, Oct. 2021.

[10] J. Choi, "On throughput bounds of NOMA-ALOHA," *IEEE Wireless Commun. Lett.*, vol. 11, no. 1, pp. 165–168, Jan. 2022.

[11] Y. Jin and T.-J. Lee, "Throughput analysis of NOMA-ALOHA," *IEEE Trans. Mobile Comput.*, vol. 21, no. 4, pp. 1463–1475, Apr. 2022.

[12] W. Fan, P. Fan, and Z. Ding, "On the throughput of NOMA-ALOHA in massive IoT with sparse active users," *IEEE Wireless Commun. Lett.*, vol. 13, no. 3, pp. 582–586, Mar. 2024.

[13] Z. Ding, R. Schober, and H. V. Poor, "NOMA-assisted grant-free transmission: How to design pre-configured SNR levels?" *IEEE Wireless Commun. Lett.*, vol. 13, no. 2, pp. 412–416, Feb. 2024.

[14] I. N. A. Ramatryana and S. Y. Shin, "Priority access in NOMA-based slotted ALOHA for overload 6G massive IoT," *IEEE Commun. Lett.*, vol. 26, no. 12, pp. 3064–3068, Dec. 2022.

[15] J. Liu, M. Agiwal, M. Qu, and H. Jin, "Online control of preamble groups with priority in massive IoT networks," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 3, pp. 700–713, Mar. 2021.

[16] Z. Chen and D. B. Smith, "Heterogeneous machine-type communications in cellular networks: Random access optimization by deep reinforcement learning," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2018, pp. 1–6.

[17] W. Fan, P. Fan, and Y. Long, "Joint delay-energy optimization for multi-priority random access in machine-type communications," *IEEE Trans. Wireless Commun.*, vol. 23, no. 2, pp. 1416–1431, Feb. 2024.

[18] J. Choi, "Random access with layered preambles based on NOMA for two different types of devices in MTC," *IEEE Trans. Wireless Commun.*, vol. 20, no. 2, pp. 871–881, Feb. 2021.

[19] *Study on RAN Improvements for Machine-Type Communications*, document TR 37.868, version 11.0.0, 3GPP, Sep. 2011.