IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received July 11, 2019, accepted July 22, 2019, date of publication July 29, 2019, date of current version August 15, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2931657

Energy-Efficient Resource Allocation With Hybrid TDMA–NOMA for Cellular-Enabled Machine-to-Machine Communications

ZEMING LI AND JINSONG GUI
School of Information Science and Engineering, Central South University, Changsha 410083, China
Corresponding author: Jinsong Gui (jsgui06@163.com)

ABSTRACT Machine-to-machine (M2M) communications, which are considered as an integral part of the Internet of Things (IoT), are being more and more ubiquitous. Meanwhile, due to the advantages of cellular networks (e.g., excellent coverage and mobility/roaming support), cellular-enabled M2M communication is a promising solution for M2M-based applications. However, there are significant challenges in cellular-enabled M2M communications due to the special features of M2M-based applications [e.g., massive concurrent uplink transmissions, small bursty traffic, and the high requirements of energy efficiency (EE)]. On the other hand, non-orthogonal multiple access (NOMA) can simultaneously serve multiple users at the same frequency and time by splitting different users in the power domain and thus increase the number of concurrent connections. In this paper, we propose an energy-efficient resource allocation scheme with hybrid time division multiple access (TDMA)–non-orthogonal multiple access (NOMA) for cellular-enabled M2M networks. First, we configure the user equipments (UEs) as the machine type communication gateways (MTCGs). Then, we propose our time sharing scheme. Next, we formulate the resource allocation problems as a noncooperative game. Finally, in order to obtain the optimal EE of machine type communication devices (MTCDs) and UEs in the game to confirm the power allocation, we transform each non-convex optimization problem into the convex form by using nonlinear fractional programming and solve the transformed problem by Dinkelbach's method and Lagrangian duality theory. The simulation results show that our scheme can dramatically shorten the total transmission time at the cost of a little more total energy when compared to the existing works.

INDEX TERMS Cellular networks, M2M, NOMA, resource allocation, energy effciency.

I. INTRODUCTION

Machine-to-machine (M2M) communications, which are also known as the machine-type communications (MTC), refer to the intelligent communications among MTC devices (MTCDs) without or with very little human interventions. Cisco Visual Networking Index report predicts that the share of M2M connections will increase from 34 percent in 2017 to 51 percent by 2022, which will reach 14.6 billion M2M connections by 2022 [1]. In other words, M2M communications are being more and more ubiquitous. Typical M2M-based applications include smart grids (SG), smart cities, environment monitoring, health monitoring, industry automation, home automation, and so on [2]–[7]. As various machines or devices in a wide range of application area are involved, M2M communications are also considered as an integral part of the Internet of Things (IoT) [8]–[11].

Generally, there are two communication scenarios in M2M communications [12]. One scenario considers communications between MTCDs and MTC servers (MTCSs) via network infrastructures (e.g., health monitoring systems). Another scenario considers direct communications among MTCDs without MTCSs (e.g., home M2M networks [13]).

The MTCDs can be connected to network infrastructures using either wired (e.g., cable, optical fiber) or wireless access methods [12]. The wireless access methods can be based on either capillary links (e.g., Wi-Fi, Bluetooth, ZigBee) or cellular links (i.e., GSM, GPRS, 3G, 4G, 5G). On the one hand, the wired solutions have low cost, but lack of the support of scalability and mobility. On the other hand, the wireless capillary solutions have small coverage, low rate, weak security, severe interference, but lack of

The associate editor coordinating the review of this manuscript and approving it for publication was Parul Garg.

universal infrastructure. Furthermore, cellular solutions offer excellent coverage, mobility/roaming support, good security, and ready-to-use infrastructure. As a result, cellular-enabled M2M communication is a more promising solution for M2M-based applications.

Hence, the communications between MTCDs and MTCSs are usually via the cellular network, while the inter-M2M device communications can be either via the cellular network or in ad-hoc mode. The M2M communications via the cellular network are also known as the 3GPP M2M communications, while those in ad-hoc mode or via other networks are also known as the General M2M communications [14].

However, there are significant challenges in cellular-enabled M2M communications due to the special features of M2M-based applications (e.g., massive concurrent uplink transmissions, small bursty traffic, and the high requirements of energy efficiency (EE) [2]). Since MTCDs may be deployed in dangerous or non-reachable places, the batteries are hardly charged or replaced, and thus saving energy for MTCDs is very important. More crucially, when a large quantity of MTCDs send transmission requests to a cellular base station (BS) directly, the intense competition for radio resource may cause severe network congestion. An effective way to tackle this problem is to deploy MTC gateways (MTCGs) in cellular-enabled M2M networks [15].

Nevertheless, it is still an open issue to choose what devices to act as the MTCGs. The deployment of specialized infrastructures will incur new network cost and complexity in cellular-enabled M2M networks [16]. Alternatively, electing MTCDs to become the MTCGs will bring more cost on air interfaces and more complexity on device design, since every MTCD should have the same configuration to face the possibility of being the MTCG. Moreover, beside an infrastructure or a MTCD, an MTCG could also be a mobile device (i.e., user equipment, UE) [17]. In home and office environments, there exists a certain number of UEs. Since UEs may have more computational power and storage space than MTCDs, it is recommended that configuring the UEs as the MTCGs in this environment [18], [19]. Furthermore, there are a lot of existing researches for improving the performance of UEs (e.g., device-to-device (D2D) communications [20]–[22]), which is also an advantage of UEs for being MTCGs.

On the other hand, non-orthogonal multiple access (NOMA) can simultaneously serve multiple users at the same frequency and time resource by splitting different users in the power domain [23]. Consequently, the NOMA-based access scheme yields a significant gain in spectrum efficiency (SE) over conventional orthogonal access method [24]–[27]. This favorable character makes NOMA to be a promising access solution for supporting the massive MTCDs in cellular-enabled M2M networks. In conjunction with NOMA transmission, the joint use of superposition code at the transmitter and successive interference cancellation (SIC) at the receiver have been also studied [28]–[32].

However, NOMA also has its limitation. To reduce the receiver complexity and error propagation due to SIC, it is reasonable for the same resource to be multiplexed by a small number (usually two to four) of devices [33]. Hence, it is necessary to combine the NOMA with the existing access technology (e.g., time division multiple access, TDMA). In this paper, we study and propose an energy-efficient resource allocation scheme with hybrid TDMA-NOMA for cellular-enabled M2M networks.

The main contributions are as follows.

1) We propose a time sharing scheme with hybrid TDMA-NOMA in the resource allocation, where we let MTCDs reuse the time slots of UEs to exploit the character of the low power of MTCDs. Due to the NOMA and the reusing of time slots, the SE of MTCDs can be improved, where the problem of low SE brought from the massive deployment of MTCDs can be alleviated.
2) The proposed time sharing scheme is based on the orthogonal frequency division multiplexing (OFDM) in the long-term evolution-advanced (LTE-A) system, where it uses one subchannel in the frequency domain and is flexible in the time domain. Hence, the proposed scheme facilitates the scale expansion of cellular-enabled M2M networks either in the frequency domain or in the time domain. In other words, the proposed scheme contributes to the massive deployment of MTCDs.
3) We model the resource allocation problem among MTCDs and UEs as an upward cellular resource allocation non-cooperative game. Every UE and its serving MTCD group form a team. Each UE optimizes the EE of every member in its managing team, where the strong computational power of UEs can be exploited and the computational burden of MTCDs can be relieved. Since the optimization objective is EE and the UEs take charge of the computation of EE optimizations, the energy saving problems of MTCDs can be improved.
4) The proposed game-theoretic scheme also has a hybrid structure. The game is under the unified scheduling of the evolved Node B (eNB), individual UEs perform the EE optimization of every member in their teams under the maximum transmission power constraints and the quality of service (QoS) requirements in every round of the game. That is, the computational burden of the BS can be also relieved.
5) We also set upper limits of SE in the EE optimization of MTCDs to solve the SE fairness problem brought from different levels of intra-group interference due to the SIC.

The remainder of this paper is organized as follows: In Section II, we give a brief overview of the state of the art. In Section III and IV, we introduce the system model of LTE-A-enabled M2M network architecture, the time sharing scheme and the corresponding problem formulation respectively. In Section V and VI, we discuss our energy-efficient

resource allocation scheme in detail. We introduce the simulation parameters, results, and analyses in Section VII. Finally, we summarize our results and give the conclusions in Section VIII.

II. RELATED WORK

Since cellular networks offer excellent coverage, mobility/roaming support, good security, and ready-to-use infrastructure for M2M communications, many researches of resource allocation in cellular-enabled M2M networks have been attracted from industry and academia in recent years. For example, a joint scheduling and power allocation issue for M2M communications in uplink single-carrier frequency-division multiple access (SC-FDMA)-based LTE-A networks was studied in [34]. A QoS-driven energy-efficient resource allocation for the uplink LTE networks in M2M/H2H(human-to-human) co-existence scenarios was investigated in [35]. A resource allocation scheme for M2M LTE networks, where LTE users, MTCDs, and relay nodes co-exist, was proposed in [36].

On the other hand, some works devoted to exploiting NOMA to support the massive communication of MTCDs in the M2M network. In [37], an uncoordinated non-orthogonal random access protocol, which is an enhancement to the recently introduced Slotted Aloha-NOMA (SAN) protocol, was presented to improve the network throughput. In [38], a reconfigurable medium access control with the ability to choose a proper access scheme with the optimal configuration for devices based on the network status was proposed, and the network throughput was also improved.

As we can see the, most of the existing works considered either the cellular-enabled M2M networks or the M2M networks with NOMA. However, there are also some works considering combining the NOMA and the cellular-enabled M2M networks. In [39], a new millimeter-wave NOMA (mmWave-NOMA) downlink transmission scheme based on a novel MTC pairing scheme for cellular M2M systems was designed and the outage probability was improved. In [40], a novel random access and resource allocation scheme for the coexistence of NOMA-based and orthogonal multiple access (OMA)-based M2M communications was proposed to improve the number of successful data packet transmissions and guarantee the QoS.

However, none of the above works related to NOMA has taken the limited energy of MTCDs into consideration. The work in [41] proposed a novel architecture for M2M communications over cellular networks, where the K-mean clustering for machines as well as cluster head (CH) reselection method was applied to balance the power consumption within the machines to increase their battery life. The work in [42] studied and compared two energy efficient resource allocation schemes with nonlinear energy harvesting and two different multiple access strategies (i.e., NOMA and TDMA) for the cellular-enabled M2M network, where the energy consumption is reduced.

In the above works, only the work in [42] considered the energy efficient resource allocation in the cellular-enabled M2M networks. However, the resource allocation in [42] was given as a centralized algorithm, which only can be run on the BS side. And the UEs were not configured as MTCGs in [42]. On the one hand, the time cost of the algorithm is high when compared with the non-centralized algorithm. On the other hand, the computational burden of the BS was aggravated and the strong computational power of UEs was not exploited.

Besides, there are also some works of resource allocation considering reusing the spectrum resource of UEs for MTCDs in the cellular-enabled M2M networks. The work in [18] configured the UEs as the full-duplex (FD) relaying based gateways and reused the channels of UEs for MTCDs. The work in [43] studied the power control for cognitive M2M communications underlaying cellular network, where MTCDs reuse the licensed spectrum of UEs in an opportunistic and fair manner.

Thus, we propose an energy efficient resource allocation scheme with the algorithm in the hybrid architecture (i.e., a combined centralized and distributed algorithm) in this paper. Furthermore, to further exploit the character of low power of MTCDs, we consider reusing the time slots of UEs for MTCDs in the proposed scheme. That is, on the premise of guaranteeing the EE, we aim at not only lowering the time cost of the algorithm but also shortening the transmission time.

III. SYSTEM MODEL

For a general case, we consider a total of N MTCD groups[1] in an LTE-A system with an eNB, as shown in Fig. 1. In MTCD group $n(n = 1, 2, \ldots, N)$, the set of MTCDs is denoted as G_n. And $G = \{G_n | n = 1, 2, \ldots, N\}$ denotes the set of all MTCD groups. For every MTCD group n, there is a corresponding UE n to serve it. Also, $U = \{n | n = 1, 2, \ldots, N\}$ denotes the set of all UEs.

The LTE-A system allocates resource by resource blocks (RBs). The structure of an RB is illustrated in Fig. 2. An RB consists of 84 resource elements (REs). An RE consists of an OFDM symbol in the time domain and a subcarrier in the frequency domain. Every OFDM symbol has a duration of 1/14 ms, and every subcarrier has a bandwidth of 15kHz. Hence, an RB consists of 0.5ms (i.e., the duration of 7 OFDM symbols) in the time domain and 180kHz (i.e., the bandwidth of 12 subcarriers) in the frequency domain.

Furthermore, two RBs form a scheduling block (SB) in the LTE-A system, which consists of a transmission time interval (TTI) in the time domain (i.e., 1 ms) and a subchannel with 12 subcarriers in the frequency domain.

In our resource allocation, the time sharing scheme is based on a subchannel in the frequency domain and a certain amount of TTIs in the time domain, which depends on the

[1]Our works focus on the resource allocation of MTCD groups in the cellular-enabled M2M networks. The relatively simple grouping method will be used, such as, the K-means clustering in the simulations. We plan to study the better grouping method in the future works.

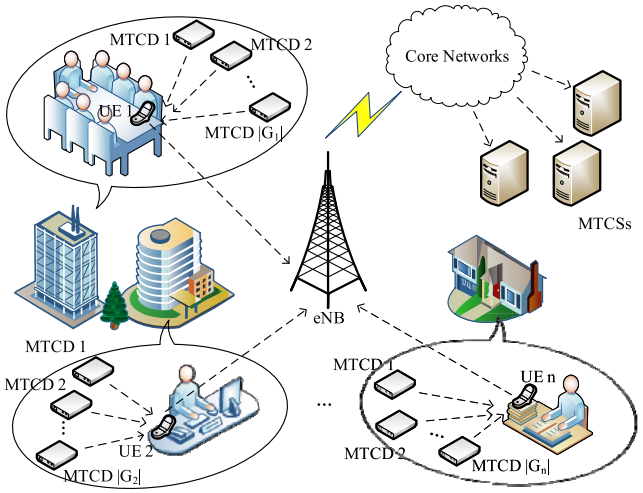

FIGURE 1. LTE-A-enabled M2M network architecture.

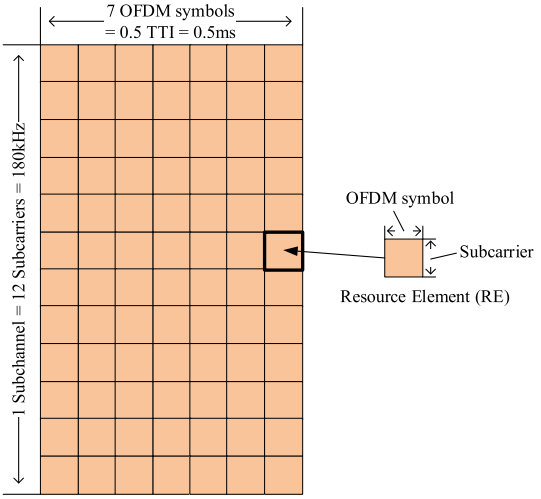

FIGURE 2. An RB structure in the LTE-A system.

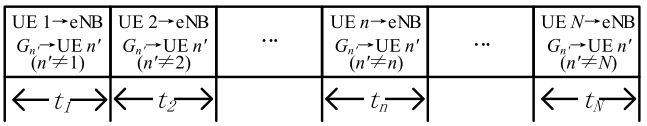

FIGURE 3. Time sharing scheme in one uplink transmission period.

transmission period. As depicted in Fig. 3, one uplink transmission period T is divided into N equal slots which are denoted as $t_1, t_2, \ldots, t_N$, i.e., $T = \sum_{n=1}^{N} t_n$. In every uplink transmission period, UE n transmits the decoded data to the eNB in t_n, where the encoded data is simultaneously transmitted from the served MTCD group n following the NOMA principle in reused t_m of the last uplink transmission period, and is decoded by applying SIC in UE n. Considering the LTE-A air interface of UEs is half duplex, we have $m \neq n$. To further exploit the benefit of reusing spectrum resource, we build our scheme with the design that the MTCD groups communicate with UEs by reusing the time slots of UEs instead of the independent time slots. Since MTCDs communicate with their corresponding UE in low power and short distance, it can be predicted that MTCDs would hardly cause or suffer severe interference. And for further improving the SE of MTCD groups, we allow them to reuse more than one time slot.

The achievable SE (defined as bits/s/Hz) of MTCD i in group n (i.e., $i \in G_n$) in one uplink transmission period is given by

$$C_{i,n}^{D} = \sum_{\substack{1 \le m \le N \\ m \neq n}} \frac{s_{n,m}^{D} t_m}{T} log_2 \left(1 + \frac{p_{i,n,m}^{D} g_{i,n}^{n}}{I_{i,n,m}^{D,A} + I_{i,n,m}^{D,B} + I_{i,n,m}^{U} + N_0}\right) \tag{1}$$

where $p_{i,n,m}^{D}$ represents the transmission power allocated to MTCD i in group n in time t_m. g_a^b represents subchannel attenuation from a to b in this paper. Accordingly, $g_{i,n}^{n}$ represents the subchannel attenuation of the link from MTCD i in group n to UE n. $s_{n,m}^{D}$ is a binary indicator, where $s_{n,m}^{D} = 1$ represents that MTCD group n reuse the time t_m, and otherwise, $s_{n,m}^{D} = 0$. Note that MTCDs in the same group will reuse the same time slots for the feasibility of NOMA protocol. N_0 is the additive zero-mean Gaussian noise.

In (1), $I_{i,n,m}^{D,A}$ and $I_{i,n,m}^{D,B}$ are the intra-group interference and the inter-group interference caused by other MTCDs respectively, and are given by

$$I_{i,n,m}^{D,A} = \sum_{j=i+1}^{|G_n|} p_{j,n,m}^{D} g_{j,n}^{n} \tag{2}$$

$$I_{i,n,m}^{D,B} = \sum_{G_{n'} \in G \setminus \{G_n\}} \sum_{j \in G_{n'}} s_{n',m}^{D} p_{j,n',m}^{D} g_{j,n'}^{n} \tag{3}$$

where $p_{j,n,m}^{D} g_{j,n}^{D}$ represents the intra-group interference from MTCD j in group n to UE n in time t_m. We assume that the MTCDs in group n (i.e., the set G_n) have been sorted as $g_{1,n}^{n} \ge g_{2,n}^{n} \ge \cdots \ge g_{i,n}^{n} \ge \cdots \ge g_{|G_n|,n}^{n}$. The signals of the MTCDs which are farer from the eNB in the same group will be decoded first. Thus, by applying SIC, the link from MTCD i in group n to UE n only suffer the intra-group interference from MTCDs with bigger indexes in the same group. $s_{n',m}^{D} p_{j,n',m}^{D} g_{j,n'}^{n}$ represents the inter-group interference from MTCD j in group n' to UE n in time t_m, i.e., $n' \neq n$.

$I_{i,n,m}^{U}$ is the UE interference, which is given by

$$I_{i,n,m}^{U} = p_m^{U} g_m^{n} \tag{4}$$

where $p_m^{G} g_m^{n}$ represents the UE interference from UE m to UE n in time t_m.

The achievable SE of UE n (i.e., $n \in U$) in one uplink transmission period is given by

$$C_n^{U} = \frac{t_n}{T} log_2 \left(1 + \frac{p_n^{U} g_n^{e}}{I_n^{D} + N_0}\right) \tag{5}$$

where p_n^{U} represents the transmission power of UE n. g_n^{e} represents the subchannel attenuation of the link from UE n to eNB.

In (5), I_n^D is the MTCD interference, and is given by

$$I_n^D = \sum_{G_{n'} \in G \setminus \{G_n\}} \sum_{i \in G_{n'}} s_{n',n}^D p_{i,n',n}^D g_{i,n'}^n \tag{6}$$

where $s_{n',n}^D p_{i,n',n}^D g_{i,n'}^n$ represents the MTCD interference from MTCD i in MTCD group n' to UE n in time t_n.

The total power consumptions of MTCD i in group n and UE n are respectively given by

$$p_{i,n}^{D,t} = \frac{1}{\eta} \sum_{\substack{1 \le m \le N \\ m \ne n}} \frac{s_{n,m}^D t_m}{T} p_{i,n,m}^D + p_{i,n}^{D,C} \tag{7}$$

$$p_n^{U,t} = \frac{1}{\eta} \cdot \frac{t_n}{T} \cdot p_n^U + p_n^{U,C} \tag{8}$$

where $p_{i,n}^{D,t}$ is constituted by the average transmission power of MTCD i in group n on all time slots, i.e., $\frac{1}{\eta} \sum_{\substack{1 \le m \le N \\ m \ne n}} \frac{s_{n,m}^D t_m}{T} p_{i,n,m}^D$, and the circuit power of it, i.e., $p_{i,n}^{D,C}$. On the other hand, $p_n^{U,t}$ is constituted by the transmission power of UE n (in time t_n), i.e., $\frac{1}{\eta} \cdot \frac{T_n}{T} \cdot p_n^U$, and the circuit power of it, i.e., $p_n^{U,C}$.

IV. PROBLEM FORMULATION

Since each device (i.e., one of MTCDs and UEs) is only interested in maximizing its individual benefit rationally and selfishly, we modeled a noncooperative game G to address the resource allocation problem, which can be denoted as the triplet $\mathbb{G} = \langle \mathbb{T}, \mathbb{S}, \mathbb{E} \rangle$. $\mathbb{T}$ is the set of gamers (i.e., MTCDs and UEs) participating in the game. $\mathbb{S}$ is the set of possible strategies that gamers can take in the game. $\mathbb{E}$ is the set of gamers' utilities when they take any strategy in $\mathbb{S}$.

To make the best of the computational power of UEs and save the energy of MTCDs, we let UEs participate in the game on behalf of their corresponding MTCD groups. Specifically, UEs take charge of computing the optimal strategies of them and their MTCD groups, while MTCDs just wait for receiving the game results which will be sent from UEs after the game. Hence, UE n and its MTCD group G_n can be considered as Team n taking actions together in the game, i.e., $\mathbb{T} = \{(n, G_n) | n \in U, G_n \in G, n = 1, 2, \ldots, N\}$.

Accordingly, $\mathbb{S} = \{(S_n^D, P_n^D, P_n^U) | n = 1, 2, \ldots, N\}$, where (S_n^D, P_n^D, P_n^U) is the set of possible strategies of Team n. And $S_n^D = \{s_{n,m}^D | s_{n,m}^D = \{0, 1\}, m = 1, 2, \ldots, N, m \ne n\}$ specifies the possible strategies for MTCD group n to reuse different time slots. $P_n^D = \{P_{i,n}^D | i \in G_n\}$ and $P_{i,n}^D = \{p_{i,n,m}^D | 0 \le p_{i,n,m}^D \le p_{i,n}^{D,max}, m = 1, 2, \ldots, N, m \ne n\}$ specifies the possible strategies of transmission power of MTCD group n. Similarly, $P_n^U = \{p_n^U | 0 \le p_n^U \le p_n^{U,max}\}$ specifies the possible strategies of transmission power of UE n.

The utility function is defined as the EE (bits/J/Hz), which is the ratio of the SE to the total power consumption [44]. When we aim at maximizing the EE of every MTCD and every UE, $\mathbb{E} = \{E_{i,n}^D | i \in G_n, n = 1, 2, \ldots, N\} \cup \{E_n^U | n = 1, 2, \ldots, N\}$, where $E_{i,n}^D$ and E_n^U are the EE of MTCD i in group n and UE n respectively and are given by

$$E_{i,n}^D \left(S_n^D, P_{i,n}^D\right) = \frac{C_{i,n}^D \left(S_n^D, P_{i,n}^D\right)}{p_{i,n}^{D,t} \left(S_n^D, P_{i,n}^D\right)} \tag{9}$$

$$E_n^U \left(P_n^U\right) = \frac{C_n^U \left(P_n^U\right)}{p_n^{U,t} \left(P_n^U\right)} \tag{10}$$

The corresponding EE optimization problems are formulated as

$$\begin{cases} \max\limits_{\left(s_n^D, P_{i,n}^D\right)} & E_{i,n}^D \left(S_n^D, P_{i,n}^D\right) \\ s.t. & C_1^m : s_{n,m}^D = \{0, 1\} \\ & C_2^m : 0 \le p_{i,n,m}^D \le p_{i,n}^{D,max} \\ & C_3 : C_{i,n}^D \left(S_n^D, P_{i,n}^D\right) \ge C_{i,n}^{D,min} \end{cases} \tag{11}$$

$$\begin{cases} \max\limits_{\left(P_n^U\right)} & E_n^U \left(P_n^U\right) \\ s.t. & C_4 : 0 \le p_n^U \le p_n^{U,max} \\ & C_5 : C_n^U \left(P_n^U\right) \ge \sum\limits_{i \in G_n} C_{i,n}^D \end{cases} \tag{12}$$

where $C_1 = \{C_1^m | m = 1, 2, \ldots, N, m \ne n\}$ and $C_2 = \{C_2^m | m = 1, 2, \ldots, N, m \ne n\}$ are power constraint set and time slot selection constraint set for MTCD i in group n. C_4 are the power constraint. All of them are mentioned as strategy sets hereinbefore. C_3 and C_5 indicate the QoS constraints.

However, since all MTCDs in group n use the same strategy S_n^D, the serial of optimal S_n^D obtained by maximizing $E_{i,n}^D$ (i.e., solving problem (11)) for $i \in G_n$ are conflicting. Thus, aiming at maximizing the average EE of MTCD group n would be a better approach. Accordingly, $\{E_{i,n}^D | i \in G_n, n = 1, 2, \ldots, N\}$ in $\mathbb{E}$ can be rewritten as $\{E_n^D | n = 1, 2, \ldots, N\}$, where E_n^D is the average EE of MTCD group n given by (13). The corresponding rewritten EE optimization problem is given by (14).

$$E_n^D \left(S_n^D, P_n^D\right) = \frac{\sum_{i \in G_n} C_{i,n}^D \left(S_n^D, P_{i,n}^D\right)}{\sum_{i \in G_n} p_{i,n}^{D,t} \left(S_n^D, P_{i,n}^D\right)} \tag{13}$$

$$\begin{cases} \max\limits_{\left(S_n^D, P_n^D\right)} & E_n^D \left(S_n^D, P_n^D\right) \\ s.t. & C_1^m : \quad s_{n,m}^D = \{0, 1\} \\ & C_6^{i,m} : \quad 0 \le p_{i,n,m}^D \le p_{i,n}^{D,max} \\ & C_7^i : \quad C_{i,n}^D \left(S_n^D, P_{i,n}^D\right) \ge C_{i,n}^{D,min} \end{cases} \tag{14}$$

where $C_6 = \{C_6^{i,m} | i \in G_n, m = 1, 2, \ldots, N, m \ne n\}$ and $C_7 = \{C_7^i | i \in G_n\}$ are power constraint set and QoS constraint set respectively.

There are challenges when optimizing the EE of MTCD i in group n and UE n by solving the above problems (12) and (14) respectively.

(a) C_1 is a constraint set of Boolean variables.

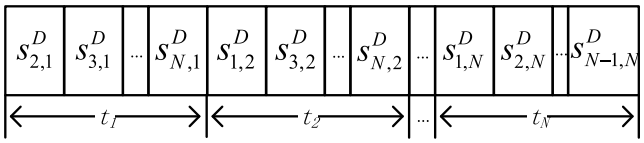

FIGURE 4. Revised time sharing scheme in one uplink transmission period.

(b) The SE fairness among MTCDs in group n cannot be guaranteed by maximizing E_n^D (i.e., solving problem (14)).
(c) The objective functions are not convex.
(d) There are a large number of constraints in problem (14).

Clearly, challenges (a), (c) and (d) means solving problems (12) and (14) by convex optimization theory is difficult, while challenge (b) indicates a potential problem in game G. In the next section, we will present our recourse allocation scheme to cope with the above challenges.

V. THE ENERGY-EFFICIENT RESOURCE ALLOCATION SCHEME

To cope with the above challenges, we introduce our energy-efficient resource allocation scheme in this section. First, we revise the time sharing scheme by introducing constraint relaxation. Second, we guarantee the SE fairness among MTCDs in one group by estimating and distributing the maximum forwarding capability of UEs. Third, we transform the non-convex optimization problems into the convex form by nonlinear fractional programming, and solve the transformed problems by Dinkelbach's method and Lagrangian duality theory. Finally, we describe the architecture and analyze the Nash equilibrium of the noncooperative game.

A. REVISED TIME SHARING SCHEME

We first deal with challenge (a). We relax $s^D_{n,m}$ in constraint C_1 from a Boolean value to a real number between 0 and 1 (i.e., $0 \le s^D_{n,m} \le 1$) to handle the problem like [45]. In this case, $s^D_{n,m}$ can be explained as a time-sharing factor for N-1 MTCD groups (except MTCD group m) to utilize time slot t_m. In other words, MTCD group n uses a part (i.e., $s^D_{n,m}$) of every time slot t_m except t_n. A graphical representation for this approach is shown in Fig. 4. Note that with this approach, $I^{D,B}_{i,n,m}$ estimated by (3) is 0.

Accordingly, problem (14) is rewritten as

$$\begin{cases} \max\limits_{(S^D_n, P^D_n)} & E^D_n\left(S^D_n, P^D_n\right) \\ s.t. & C^m_1: \quad 0 \le s^D_{n,m} \le 1 \\ & C^{i,m}_6: \quad 0 \le p^D_{i,n,m} \le p^{D,max}_{i,n} \\ & C^i_7: \quad C^D_{i,n}\left(S^D_n, P^D_{i,n}\right) \ge C^{D,min}_{i,n} \end{cases} \tag{15}$$

From (15), we can find $\frac{\partial E^D_n}{\partial s^D_{n,m}} > 0$. That is to say, reusing slots as much part as possible will always benefit the EE of MTCDs. However, $s^D_{n,m}$ should also meet $\sum_{\substack{1\le n\le N \\ n\ne m}}$ $s^D_{n,m} = 1$ for $m = 1, 2, \ldots, N$[46]. Thus, it is not suggested to consider strategies S^D_n in game $\mathbb{G}$. To tackle this problem, we can find out the appropriate S^D_n according to some evaluation strategies before the start of game $\mathbb{G}$. Some evaluation methods will be given in the simulations. In other words, S^D_n is regarded as a set of constants in game $\mathbb{G}$. Accordingly, (15) is rewritten as

$$\begin{cases} \max\limits_{(p^D_n)} & E^D_n\left(P^D_n\right) \\ s.t. & C^{i,m}_6: \quad 0 \le p^D_{i,n,m} \le p^{D,max}_{i,n} \\ & C^i_7: \quad C^D_{i,n}\left(P^D_{i,n}\right) \ge C^{D,min}_{i,n} \end{cases} \tag{16}$$

where (16) is the optimization problem about the average EE of MTCD group n. Since S^D_n is considered as a constant set, we can further study the EE of every MTCD i in group n, which is rewritten from (9) and is given as (17). And the corresponding EE optimization problem is rewritten from (11) and given as (18)

$$E^D_{i,n}\left(P^D_{i,n}\right) = \frac{C^D_{i,n}\left(P^D_{i,n}\right)}{p^{D,t}_{i,n}\left(P^D_{i,n}\right)} \tag{17}$$

$$\begin{cases} \max\limits_{(p^D_{i,n})} & E^D_{i,n}\left(P^D_{i,n}\right) \\ s.t. & C^m_2: \quad 0 \le p^D_{i,n,m} \le p^{D,max}_{i,n} \\ & C_3: \quad C^D_{i,n}\left(P^D_{i,n}\right) \ge C^{D,min}_{i,n} \end{cases} \tag{18}$$

On the other hand, to keep off mathematically complex operation, we let UE n transmit with different power in every part of t_n (see Appendix A for detail). Accordingly, (5) and (8) are rewritten as

$$C^U_n = \sum_{\substack{1\le n'\le N \\ n'\ne n}} \frac{s^D_{n',n}t_n}{T} log_2\left(1 + \frac{p^U_{n',n}g^e_n}{I^D_{n',n} + N_0}\right) \tag{19}$$

$$p^{U,t}_n = \frac{1}{\eta}\sum_{\substack{1\le n'\le N \\ n'\ne n}} \frac{s^D_{n',n}t_n}{T} p^U_{n',n} + p^{U,C}_n \tag{20}$$

where $p^U_{n,n'}$ is the transmission power of UE n in the part $s^D_{n',n}$ of t_n. The corresponding strategy set is rewritten as $P^U_n = \{p^U_{n',n} | 0 \le p^U_{n',n} \le p^{U,max}_n, n' = 1, 2, \ldots, N, n' \ne n\}$. And $I^D_{n',n}$ is the MTCD interference in the part $s^D_{n',n}$ of t_n., which is rewritten from (6) and is given as (21).

$$I^D_{n',n} = \sum_{i\in G_{n'}} p^D_{i,n',n} g^n_{i,n'} \tag{21}$$

Accordingly, the EE optimization problem (12) should be rewritten as

$$\begin{cases} \max\limits_{(P^U_n)} & E^U_n\left(P^U_n\right) \\ s.t. & C^{n'}_8: \quad 0 \le p^U_{n',n} \le p^{U,max}_n \\ & C_5: \quad C^U_n\left(P^U_n\right) \ge \sum\limits_{i\in G_n} C^D_{i,n} \end{cases} \tag{22}$$

where $C_8 = \{C_8^{n'} | n' = 1, 2, \ldots, N, n' \neq n\}$ is the transmission power constrain set. So far, we have solved challenge (a) and transformed the EE optimization problems of MTCD i in group n and UE n (i.e., problems (18) and (22)) into the similar form by adjusting the architecture of game $\mathbb{G}$. However, the architecture still needs further revising to meet challenge (b), which will be introduced in the next subsection.

B. FAIRNESS AMONG MTCDS IN ONE GROUP

In this subsection, we will address challenge (b). It is not difficult to see that MTCDs in the same group will suffer different levels of intra-group interference (i.e., (2)) due to the SIC in NOMA. Hence, the SE of MTCDs in the same group obtained by (18) will be widely divergent. To guarantee the SE fairness, we plan to set an upper limit to the SE in problem (18). First, UE n should estimate its current maximum forwarding capability by $C_n^U\left(P_n^{U,max}\right)$, which is obtained by $P_n^U = \{P_n^{U,max} = p_{n',n}^U | p_{n',n}^U = p_n^{U,max}, n' = 1, 2, \ldots, N, n' \neq n\}$ in (19). Next, UE n distributes a part of $C_n^U\left(P_n^{U,max}\right)$ to every MTCD i in group n as the upper limit of SE, which is denoted as $C_{i,n}^{D,max}$ and is given by

$$C_{i,n}^{D,max} = \frac{C_{i,n}^{D,min}}{\sum_{i \in G_n} C_{i,n}^{D,min}} C_n^U\left(P_n^{U,max}\right) \tag{23}$$

which indicates UE n allocates the forwarding capability according to the QoS requirement ratio. Then, (18) is rewritten as

$$\begin{cases} \max\limits_{(P_{i,n}^D)} & E_{i,n}^D\left(P_{i,n}^D\right) \\ s.t. & C_2^m: \quad 0 \le p_{i,n,m}^D \le p_{i,n}^{D,max} \\ & C_3: \quad C_{i,n}^D\left(P_{i,n}^D\right) \ge C_{i,n}^{D,min} \\ & C_9: \quad C_{i,n}^D\left(P_{i,n}^D\right) \le C_{i,n}^{D,max} \end{cases} \tag{24}$$

Finally, since the maximum forwarding capability $C_n^U\left(P_n^{U,max}\right)$ will not be totally run out of in general., (22) is rewritten as

$$\begin{cases} \max\limits_{(P_n^U)} & E_n^U\left(P_n^U\right) \\ s.t. & C_8^{n'}: \quad 0 \le p_{n',n}^U \le p_n^{U,max} \\ & C_{10}: \quad C_n^U\left(P_n^U\right) = \sum\limits_{i \in G_n} C_{i,n}^D\left(P_{i,n}^{D,*}\right) + C_n^{U,0} \end{cases} \tag{25}$$

where $P_{i,n}^{D,*}$ is the optimal $P_{i,n}^D$ obtained by solving (24). $C_n^{U,0}$ is the required SE for UE n to transmit its own data (i.e., not the forwarding data). Note that for reducing interference to MTCDs, constraint C_{10} is an equality constraint.

So far, we have addressed challenge (b). In the next subsection, we will meet challenges (c) and (d) to solve problems (24) and (25).

C. CONVEXIFICATION AND SOLUTION OF EE OPTIMIZATION PROBLEMS

To address challenge (c), we consider transforming the objective functions of (24) and (25) into the convex form. Taking (24) for example, we first let $e_{i,n}^{D,*}$ be the maximum EE of MTCD i in group n, which is given by

$$e_{i,n}^{D,*} = \max_{(P_{i,n}^D)} E_{i,n}^D\left(P_{i,n}^D\right) = \frac{C_{i,n}^D\left(P_{i,n}^{D,*}\right)}{p_{i,n}^{D,t}\left(P_{i,n}^{D,*}\right)} \tag{26}$$

Next, there is the following theorem which can be easily proved in a similar way to [47].

Theorem 1: $e_{i,n}^{D,*}$ is achieved if and only if

$$\begin{aligned} &\max_{(P_{i,n}^D)} C_{i,n}^D\left(P_{i,n}^D\right) - e_{i,n}^{D,*} p_{i,n}^{D,t}\left(P_{i,n}^D\right) \\ &\qquad = C_{i,n}^D\left(P_{i,n}^{D,*}\right) - e_{i,n}^{D,*} p_{i,n}^{D,t}\left(P_{i,n}^{D,*}\right) = 0 \end{aligned} \tag{27}$$

Theorem 1 shows that if we can find a $e_{i,n}^D$ satisfies $\max\limits_{(P_{i,n}^D)} C_{i,n}^D\left(P_{i,n}^D\right) - e_{i,n}^D p_{i,n}^{D,t}\left(P_{i,n}^D\right) = 0$, the $e_{i,n}^D$ is exactly the maximum EE, i.e., $e_{i,n}^{D,*}$. Thus, Theorem 1 indicates that an original objective function in the fractional form can be transformed into a new objective function in the subtractive form. Then, from $\frac{\partial^2 C_{i,n}^D\left(P_{i,n}^D\right) - e_{i,n}^D p_{i,n}^{D,t}\left(P_{i,n}^D\right)}{\partial\left(P_{i,n}^D\right)^2} \le 0$ we can know, the new objective function is concave. To obtain a convex EE optimization problem, we just turn to study minimizing the new objective function in its minus form. Accordingly, (24) is rewritten as

$$\begin{cases} \min\limits_{(P_{i,n}^D)} & e_{i,n}^D p_{i,n}^{D,t}\left(P_{i,n}^D\right) - C_{i,n}^D\left(P_{i,n}^D\right) \\ s.t. & C_2^m, C_3, C_9 \end{cases} \tag{28}$$

Similarly, let $e_n^{U,*}$ be the maximum EE of UE n, (25) is rewritten as

$$\begin{cases} \max\limits_{(P_n^U)} & e_n^U p_n^{U,t}\left(P_n^U\right) - C_n^U\left(P_n^U\right) \\ s.t. & C_8^{n'}, C_{10} \end{cases} \tag{29}$$

Now, the convexification of EE optimization problems is finished and challenge (c) is solved. Note that C_2, C_3, C_8 and C_{10} are all convex sets. As for C_9, even though it is not a convex set, we can still obtain the global optimal solution $P_{i,n}^{D,*}$ using the convex optimization which will be introduced later.

As mentioned before, we want to find out the $e_{i,n}^D$ and e_n^U with appropriate values to make the minimum values of (28) and (29) be 0. We can adopt an iterative algorithm to find them, which is known as Dinkelbach's method [47]. Taking (28) for example, $e_{i,n}^D$ is set to be a very small positive number (e.g., $e_{i,n}^D = 10^{-4}$) in the beginning. Then, in every iteration, we solve (28) and check whether the minimum value is close to 0 enough (i.e., satisfies the precision). If the precision is not satisfied, we update $e_{i,n}^D$ by (17) with the current optimal

strategy set $P_{i,n}^{D}$ for the next iteration. For clarity, the above process is summarized as Algorithm 1, where Δ_D is the precision.

Algorithm 1 Dinkelbach's Method

Run at: UE n

Input: Δ_D

Output: $\left(P_{i,n}^{D,*}, e_{i,n}^{D,*}\right)$ or $\left(P_{n}^{U,*}, e_{n}^{U,*}\right)$

1. $e_{i,n}^{D} \leftarrow 10^{-4}$ or $e_{n}^{U} \leftarrow 10^{-4}$
2. **while**true **do**
3. Solve (28) or (29) to obtain the minimum and the corresponding power strategy set $P_{i,n}^{D,*}$ or $P_{n}^{U,*}$.
4. **if** minimum $\leq \Delta_D$ **then**
5. $\left(P_{i,n}^{D,*}, e_{i,n}^{D,*}\right) \leftarrow \left(P_{i,n}^{D}, e_{i,n}^{D}\right)$ or $\left(P_{n}^{U,*}, e_{n}^{U,*}\right) \leftarrow \left(P_{n}^{U}, e_{n}^{U}\right)$
6. **return** $\left(P_{i,n}^{D,*}, e_{i,n}^{D,*}\right)$ or $\left(P_{n}^{U,*}, e_{n}^{U,*}\right)$
7. **else** update $e_{i,n}^{D}$ by (17) or update e_{n}^{U} by (10)
8. **end if**
9. **end while**

However, we still need a method to solve problems (28) and (29). Since the optimal solution $P_{i,n}^{D,*}$ and $P_{n}^{U,*}$ can be obtained by using convex optimization, we can adopt Lagrange duality theory to find them [48]. Still taking (28) for example, the associated Lagrangian function is given as follows.

$$\begin{aligned} L_{i,n}^{D}\left(P_{i,n}^{D}, \beta_{i,n}^{D}, \gamma_{i,n}^{D}\right) = {} & e_{i,n}^{D} p_{i,n}^{D,t}\left(P_{i,n}^{D}\right) - C_{i,n}^{D}\left(P_{i,n}^{D}\right) \\ & + \beta_{i,n}^{D}\left(C_{i,n}^{D,min} - C_{i,n}^{D}\left(P_{i,n}^{D}\right)\right) \\ & + \gamma_{i,n}^{D}\left(C_{i,n}^{D}\left(P_{i,n}^{D}\right) - C_{i,n}^{D,max}\right) \end{aligned} \tag{30}$$

$\beta_{i,n}^{D}$ and $\gamma_{i,n}^{D}$ are the Lagrange multipliers associated with C_3 and C_9 respectively. Note that C_2 is not considered in (30), which will be introduced later. According to the Lagrange duality theory, solving original problem (28) is equal to solving the dual problem (31) with constraint set C_2.

$$\max_{\beta_{i,n'}^{D}, \gamma_{i,n}^{D}} \max_{\beta_{i,n}^{D}} L_{i,n}^{D}\left(P_{i,n}^{D}, \beta_{i,n}^{D}, \gamma_{i,n}^{D}\right) \tag{31}$$

To find the $P_{i,n}^{D,*}$, we need the Karush–Kuhn–Tucker (KKT) conditions, which associated with (31) is given by

$$\begin{cases} \nabla_{P_{i,n}^{D}} L_{i,n}^{D}\left(P_{i,n}^{D}, \beta_{i,n}^{D}, \gamma_{i,n}^{D}\right) = 0 & (32-1) \\ C_{i,n}^{D,min} - C_{i,n}^{D}\left(P_{i,n}^{D}\right) \leq = 0 & (32-2) \\ C_{i,n}^{D} - \left(P_{i,n}^{D}\right) - C_{i,n}^{D,max} \leq = 0 & (32-3) \\ \beta_{i,n}^{D}\left(C_{i,n}^{D,min} - C_{i,n}^{D}(P_{i,n}^{D})\right) = 0 & (32-4) \\ \gamma_{i,n}^{D}\left(C_{i,n}^{D}(P_{i,n}^{D}) - C_{i,n}^{D,max}\right) = 0 & (32-5) \\ \beta_{i,n}^{D} \geq, \gamma_{i,n}^{D} \geq & (32-6) \end{cases} \tag{32}$$

The $P_{i,n}^{D}$ which satisfies C_2 and (32) and minimizes (30) is the $P_{i,n}^{D,*}$. First, the $P_{i,n}^{D}$ satisfies (32-1) is given by (33).

$$\hat{p}_{i,n,m}^{D} = \frac{\left(1 + \beta_{i,n}^{D} - \gamma_{i,n}^{D}\right) \eta log_2 e}{e_{i,n}^{D}} - \frac{I_{i,n,m}^{D,A} + I_{i,n,m}^{G} + N_0}{g_{i,n}^{n}} \tag{33}$$

Next, from the following Theorem 2, (33) can be rewritten as

$$\hat{p}_{i,n,m}^{D} = \left[\hat{p}_{i,n,m}^{D}\right]_{0}^{p_{i,n}^{D,max}} \tag{34}$$

where C_2 will be satisfied and the meaning of $[x]_{\mathrm{a}}^{\mathrm{b}}$ is given by (35).

$$[x]_{\mathrm{a}}^{\mathrm{b}} = \begin{cases} a & x < a \\ x & a \leq x \leq b \\ b & x > b \end{cases} \tag{35}$$

Theorem 2: Direct constraint sets C_2 for $P_{i,n}^{D}$ can be absorbed as the upper limits and the lower limits of the $P_{i,n}^{D}$ obtained by the KKT condition (32-1) instead of considered as inequality constraints in the Lagrangian function (30).

Proof: Please refer to Appendix B.

Theorem 2 also explains why C_2 is not considered in (30). Accordingly, using Theorem 2 can avoid modeling the Lagrangian function associated with (28) with a large amount of inequation constraints (i.e. C_2) and can also avoid the corresponding complicated KKT conditions. In other words, challenge (d) is solved.

Then, let $\widehat{P}_{i,n}^{D} = \{\hat{p}_{i,n,m}^{D} | m = 1, 2, \ldots, N, m \neq n\}$ denote the set of $\hat{p}_{i,n,m}^{D}$. And we discuss about different cases in (32-4) and (32-5).

(a) If $\beta_{i,n}^{D} = \gamma_{i,n}^{D} = 0$, $C_{i,n}^{D,min} < C_{i,n}^{D}\left(P_{i,n}^{D,*}\right) < C_{i,n}^{D,max}$ should hold.

In this case, from $\beta_{i,n}^{D} = \gamma_{i,n}^{D} = 0$, (34) is rewritten as

$$\hat{p}_{i,n,m}^{D} = \left[\frac{\eta log_2 e}{e_{i,n}^{D}} - \frac{I_{i,n,m}^{D,A} + I_{i,n,m}^{G} + N_0}{g_{i,n}^{n}}\right]_{0}^{p_{i,n}^{D,max}} \tag{36}$$

And we plug (36) into (1) to check whether $C_{i,n}^{D,min} \leq C_{i,n}^{D}\left(\hat{P}_{i,n}^{D}\right) \leq C_{i,n}^{D,max}$ holds. If it does, $P_{i,n}^{D,*} = \hat{P}_{i,n}^{D}$.

(b) If $\beta_{i,n}^{D} > 0$, $\gamma_{i,n}^{D} = 0$, $C_{i,n}^{D}\left(P_{i,n}^{D,*}\right) = C_{i,n}^{D,min} < C_{i,n}^{D,max}$ should hold.

In this case, from $\beta_{i,n}^{D} > 0$, $\gamma_{i,n}^{D} = 0$, (34) is rewritten as

$$\hat{p}_{i,n,m}^{D} = \left[\frac{\left(1 + \beta_{i,n}^{D}\right) \eta log_2 e}{e_{i,n}^{D}} - \frac{I_{i,n,m}^{D,A} + I_{i,n,m}^{G} + N_0}{g_{i,n}^{n}}\right]_{0}^{p_{i,n}^{D,max}} \tag{37}$$

We plug (37) into (1), and confirm whether there is a $\beta_{i,n}^{D} > 0$ making $C_{i,n}^{D}\left(\hat{P}_{i,n}^{D}\right) = C_{i,n}^{D,min}$ hold ($C_{i,n}^{D,min} < C_{i,n}^{D,max}$ certainly holds). If it does, $P_{i,n}^{D,*} = \hat{P}_{i,n}^{D}$.

To find the appropriate $\beta_{i,n}^D$, we can use the gradient method [49]. The initial value of $\beta_{i,n}^D$ is set to be 0, and is updated by (38) iteratively.

$$\beta_{i,n}^D = \left[\beta_{i,n}^D + \mu_{i,n}^{D,\beta} \frac{\partial L_{i,n}^D\left(P_{i,n}^D, \beta_{i,n}^D, \gamma_{i,n}^D\right)}{\partial \beta_{i,n}^D}\right]_0^{\infty} \tag{38}$$

where $\mu_{i,n}^{D,\beta}$ is the positive step associated with $\beta_{i,n}^D$. The $\frac{\partial L_{i,n}^D\left(P_{i,n}^D,\beta_{i,n}^D,\gamma_{i,n}^D\right)}{\partial \beta_{i,n}^D}$ indicates the positive gradient direction to find the maximum.

(c) If $\beta_{i,n}^D = 0, \gamma_{i,n}^D > 0, C_{i,n}^D\left(P_{i,n}^{D,*}\right) = C_{i,n}^{D,max} > C_{i,n}^{D,min}$ should hold.

In this case, from $\beta_{i,n}^D = 0, \gamma_{i,n}^D > 0$, (34) is rewritten as

$$\hat{p}_{i,n,m}^D = \left[\frac{\left(1-\gamma_{i,n}^D\right)\eta log_2 e}{e_{i,n}^D} - \frac{I_{i,n,m}^{D,A} + I_{i,n,m}^G + N_0}{g_{i,n}^n}\right]_0^{p_{i,n}^{D,max}} \tag{39}$$

Similarly , we plug (39) into (1), and confirm whether there is a $\gamma_{i,n}^D > 0$ making $C_{i,n}^D\left(\widehat{P}_{i,n}^D\right) = C_{i,n}^{D,min}$ hold ($C_{i,n}^{D,max} > C_{i,n}^{D,min}$ certainly holds). If it does, $P_{i,n}^{D,*} = \hat{P}_{i,n}^D$. To find the appropriate value, $\gamma_{i,n}^D$ is set to be 0 initially, and is updated by (40) iteratively.

$$\gamma_{i,n}^D = \left[\gamma_{i,n}^D + \mu_{i,n}^{D,\gamma} \frac{\partial L_{i,n}^D\left(P_{i,n}^D, \beta_{i,n}^D, \gamma_{i,n}^D\right)}{\partial \gamma_{i,n}^D}\right]_0^{\infty} \tag{40}$$

where $\mu_{i,n}^{D,\beta}$ is the positive step associated with $\beta_{i,n}^D$.

(d) If $\beta_{i,n}^D > 0, \gamma_{i,n}^D > 0, C_{i,n}^D\left(P_{i,n}^{D,*}\right) = C_{i,n}^{D,min}$ and $C_{i,n}^D\left(P_{i,n}^{D,*}\right) = C_{i,n}^{D,max}$ should hold. Clearly, this is a contradictory case unless $C_{i,n}^{D,min} = C_{i,n}^{D,max}$. However, even if $C_{i,n}^{D,min} = C_{i,n}^{D,max}$, the problem is still can be solved by case (b) or (c). Thus, case (d) can be ignored.

Note that all KKT conditions are involved and satisfied in the above cases. Finally, we study how to comprehensively consider case (a) to (c). First, we check whether case (a) holds. If $C_{i,n}^D\left(\hat{P}_{i,n}^D\right) < C_{i,n}^{D,min}$, we turn to case (b). Otherwise, i.e., $C_{i,n}^D\left(\hat{P}_{i,n}^D\right) > C_{i,n}^{D,max}$, we turn to case (c). Actually, that the initial values of $\beta_{i,n}^D$ and $\gamma_{i,n}^D$ in iterations are 0 reflects the internal relation between case (a), (b) and (c).

Moreover, we can see that when C_9 works, the corresponding case is case (a) or (b). Thus, we can learn that all $\hat{p}_{i,n,m}^D$ given by (36) or (37) constitute a convex constraint set $\hat{P}_{i,n}^D = \{\hat{p}_{i,n,m}^D | m = 1, 2, \ldots, N, m \neq n\}$ in problem (28). Actually, $\hat{P}_{i,n}^D$ are subsets of $P_{i,n}^D$, which means $p_{i,n,m}^D$ in C_9 are not totally freedom. In other words, C_9 can be narrowed as a convex set by $\hat{p}_{i,n,m}^D$.

Problem (29) can be solved in the similar way. To save space, we directly give the related formulas as (41) ∼ (43).

$$\begin{aligned} L_n^U\left(P_n^U, \alpha_n^D\right) &= e_n^U p_n^{U,t}\left(P_n^U\right) - C_n^U\left(P_n^U\right) \\ &\quad + \alpha_n^U\left(C_n^U\left(P_n^U\right) - \sum_{i\in G_n} C_{i,n}^D\left(P_{i,n}^{D,*}\right) - C_n^{U,0}\right) \end{aligned} \tag{41}$$

$$\hat{p}_{n',n}^U = \left[\frac{\left(1+\alpha_n^D\right)\eta log_2 e}{e_n^U} - \frac{I_{n',n}^D + N_0}{g_n^e}\right]_0^{p_n^{U,max}} \tag{42}$$

$$\alpha_n^U = \alpha_n^U + \mu_n^{U,\alpha} \frac{\partial L_n^U\left(P_n^U, \alpha_n^U\right)}{\partial \alpha_n^U} \tag{43}$$

For clarity, the above process is also summarized as algorithm 2-a and 2-b, where Δ_L is the precision.

Algorithm 2-a Lagrange Duality for UEs

Run at: UE n
Input: $\Delta_L, e_n^U, C_{i,n}^D\left(P_{i,n}^{D,*}\right)$ for $i \in G_n$
Output: minimum in (29)

1. $\alpha_n^D \leftarrow 0$
2. **while** true **do**
3. Calculate $p_{n',n}^{U,*}$ for $n' = 1, 2, \ldots, N, n' \neq n$
4. Calculate $C_n^U\left(P_n^{U,*}\right)$ by (19)
5. **if** $\left|C_n^U\left(P_n^{U,*}\right) - \left(\sum_{i\in G_n} C_{i,n}^D\left(P_{i,n}^{D,*}\right) - C_n^{U,0}\right)\right| \leq \Delta_L$ **then**
6. Update α_n^D by (43)
7. **else return** the minimum of (29) using $P_n^{U,*}$
8. **end if**
9. **end while**

D. GAME ARCHITECTURE AND NASH EQUILIBRIUM ANALYSIS

The noncooperative game $\mathbb{G}$ has a hybrid architecture. In the prepare phase, UEs collect the information (e.g. the location and the QoS) of their corresponding MTCD group first. Then, the eNB collects the location information of every team. Next, every UE receives the location information of other teams from the eNB and reports that it's prepared for the game. Last, the eNB declares the start of the formal stage after receiving all preparation reports from every UE.

In the formal stage, the game will repeat for many rounds. In every round, every UE receives the new game information from eNB first. Next, every UE estimates its maximum forwarding ability, and optimizes the EE of its MTCDs by Algorithm 1 and 2-a. Then, every UE optimizes its EE according to the optimal SE of its MTCDs by Algorithm 1 and 2-b. Last, every UE reports the new transmission power strategies (i.e., the current optimal transmission power strategies) of its team to the eNB.

Meanwhile, in the beginning of every round, the eNB will broadcasts the new collected game information to every UE, until the average transmission power of all teams is stable

Algorithm 2-b Lagrange Duality for MTCDs

Run at: UE n
Input: Δ_L, $e_{i,n}^D$, $C_{i,n}^{D,max}$
Output: the minimum in (28)

$\beta_{i,n}^D \leftarrow 0$, $\gamma_{i,n}^D \leftarrow 0$
Calculate $p_{i,n,m}^{D,*}$ for $m = 1, 2, \ldots, N, m \neq n$
Calculate $C_{i,n}^D\left(P_{i,n}^{D,*}\right)$ by (1)
If $C_{i,n}^{D,min} < C_{i,n}^D\left(P_{i,n}^{D,*}\right) < C_{i,n}^{D,max}$ **then**
  return the minimum of (28) using $P_{i,n}^{D,*}$
else if $C_{i,n}^D\left(P_{i,n}^{D,*}\right) < C_{i,n}^{D,min}$ **then**
  while true **do**
    Update $\beta_{i,n}^D$ by (38)
    Calculate $p_{i,n,m}^{D,*}$ for $m = 1, 2, \ldots, N, m \neq n$
    Calculate $C_{i,n}^D\left(P_{i,n}^{D,*}\right)$ by (1)
    If $\left|C_{i,n}^D\left(P_{i,n}^{D,*}\right) - C_{i,n}^{D,min}\right| \leq \Delta_L$ **then**
      return the minimum of (28) using $P_{i,n}^{D,*}$.
    end if
  end while
else
  while true **do**
    Update $\gamma_{i,n}^D$ by (39)
    Calculate $p_{i,n,m}^{D,*}$ for $m = 1, 2, \ldots, N, m \neq n$
    Calculate $C_{i,n}^D\left(P_{i,n}^{D,*}\right)$ by (1)
    If $\left|C_{i,n}^D\left(P_{i,n}^{D,*}\right) - C_{i,n}^{D,max}\right| \leq \Delta_L$ **then**
      return the minimum of (28) using $P_{i,n}^{D,*}$.
    end if
  end while
end if

(i.e., changes very a little). Then, the game is over and a Nash equilibrium has been reached.

Definition 1: A Nash equilibrium in game $\mathbb{G}$ is a set of strategies that none of devices (neither MTCDs nor UEs) can unilaterally improve its EE performance by choosing a different strategy set.

Theorem 3: The optimal strategy set $\left\{P_n^{D,*}, P_n^{U,*} | n = 1, 2, \ldots, N\right\}$ obtained by the proposed resource allocation scheme constitutes a Nash equilibrium in game $\mathbb{G}$.

Proof: According to [50], a Nash equilibrium exists if the utility function is continuous and quasi-concave, and the set of strategies is a nonempty compact convex subset of a Euclidean space. In our proposed noncooperative game $\mathbb{G}$, the utility functions of MTCDs and UEs are given by the objective functions in (28) and (29), where left terms are affine functions and right terms are convex functions. Furthermore, the corresponding strategy sets C_2, C_3, C_8 and C_{10} are all nonempty compact convex subsets of Euclidean spaces. And C_9 can be narrowed to be convex. Hence, it is easy to prove that a Nash equilibrium exists in our proposed noncooperative game.

TABLE 1. Simulation parameters.

Parameter	Value
Cell radius R	300m
Transmission period T	any
Number of UEs N	10
Number of MTCDs in each group	4
Payload of each MTCD	10kbits
System bandwidth B	180KHz
PA efficiency η	90%
Noise power N_0	−104 dBm
Convergence precision Δ_L and Δ_D	10^{-6}
Circuit power of MTCDs $p_{i,n}^{D,C}$	0.5 mW for all
Circuit power of UEs $p_n^{U,C}$	0.5W for all
Maximal transmission power of MTCDs $p_{i,n}^{D,max}$	5 mW for all
Maximal transmission power of UEs $p_n^{U,max}$	0.2W for all
QoS of MTCDs $C_{i,n}^{D,min}$	0.01 bits/s/Hz for all
Required SE for UEs' own data $C_n^{U,0}$	0 bits/s/Hz for all

VI. SIMULATION RESULTS

In this section, the proposed energy-efficient resource allocation scheme with hybrid TDMA-NOMA, which is labeled as "HTN", is compared with the NOMA scheme in the existing work [42], which is labeled as "IPCTA-NOMA". To this end, first, we summarized the simulation parameters as Table 1, where the parameters values are based on [42] unless otherwise specified. The path loss model also refers to [42] and is given by $128.1 + 37.6\log_{10} d_a^b$ (d_a^b is in km), where d_a^b denotes the distance from a to b. Second, we give three kinds of evaluation methods to set the values of S_n^D ($n = 1, 2, \ldots, N$), which are given by

$$s_{n',n}^D = \frac{1}{N-1} t_n \tag{44}$$

$$s_{n',n}^D = \frac{\frac{1}{d_n^{n'}}}{\sum_{\substack{1 \leq m \leq N \\ m \neq n}} \frac{1}{d_n^m}} t_n \tag{45}$$

$$s_{n',n}^D = \frac{\sum_{i \in G_{n'}} d_{i,n'}^e}{\sum_{\substack{1 \leq m \leq N \\ m \neq n}} \sum_{i \in G_m} d_{i,m}^e} t_n \tag{46}$$

where $n' \neq n$ and the corresponding schemes are labeled as "HTN-1", "HTN-2" and "HTN-3" respectively. Clearly, (44) indicates an equal distribution of time t_n for each MTCD group n'. Since there is mutual interference between UE n and every MTCD group n' in time t_n, (45) consider the situation that the interference from the former to the latter is severer than that from the latter to the former, and (46) is on the contrary. To be specific, (45) indicates that the shorter distance between UE n and UE n' (UE n' is the signal receiver of MTCD group n') leads to the severer interference from UE n to MTCD group n', hence a larger part of time t_n will be allocated to MTCD group n' (i.e., $s_{n',n}^D$). Similarly, (46) indicates that the shorter total distance between MTCD group n' and the eNB (the eNB is the signal receiver of UE n) leads to the severer interference from MTCD group n' to UEn, hence a smaller $s_{n',n}^D$ will be allocated.

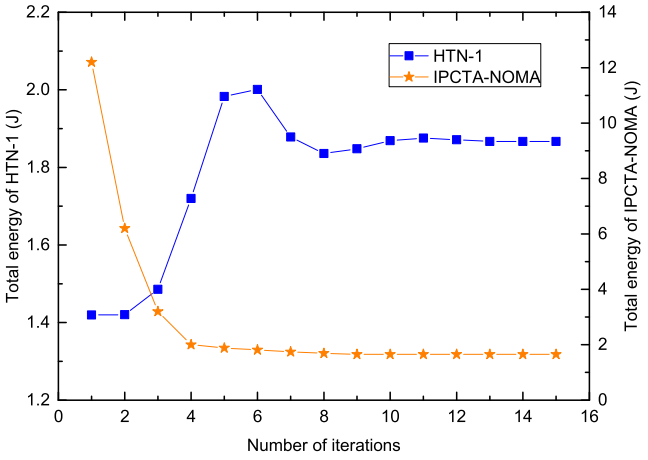

FIGURE 5. Convergence behaviors of HTNs and IPCTA-NOMA.

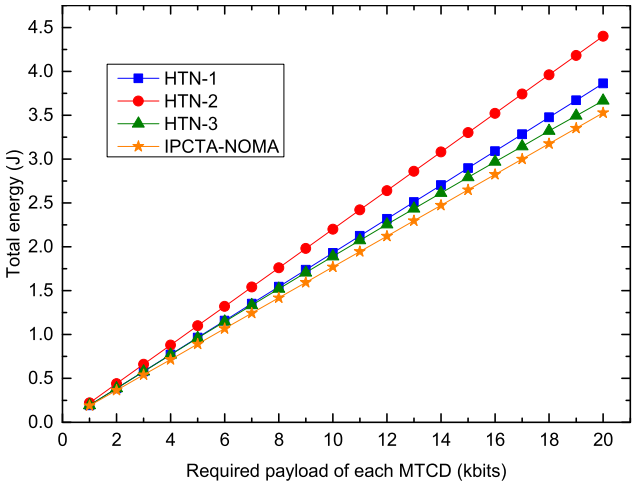

FIGURE 6. Total energy versus required payload of each MTCD.

Note that (44) ∼ (46) are based on d_a^b but not g_a^b, because g_a^b usually have different orders of magnitude.

Last, to intuitively compare the simulation results, we consider the indicator "total energy" just like [42], which can be converted from the EE and is given by

$$TE = \sum_{n\in U}\left(\frac{\sum_{i\in G_n} D_{i,n}^D}{BE_n^U} + \sum_{i\in G_n}\frac{D_{i,n}^D}{BE_{i,n}^D}\right) \tag{47}$$

where $D_{i,n}^D$ is the data of MTCD i in group n to be sent ($D_{i,n}^D$ is in bits), which is known as the payload in [42]. B is the subchannel bandwidth. Similarly, we also compare the total time for transmitting all data, which is given by

$$TT = \frac{1}{N}\sum_{n\in U}\frac{\sum_{i\in G_n} D_{i,n}^D}{BC_n^U} \tag{48}$$

where the total time is an average value since UEs and MTCDs transmit data simultaneously in a transmission period T.

Note that in some simulations with extreme conditions, the QoS of some MTCDs will be unsatisfied, where we just mark those MTCDs and let them transmit with maximal power to maintain game $\mathbb{G}$ in HTNs.

A. CONVERGENCE BEHAVIORS

Fig. 5 compares the convergence behaviors of HTN-1 and IPCTA-NOMA, where the iterations refer to the game rounds in HTN-1. The convergence behaviors of HTN-2 and HTN-3 are similar to that of HTN-1, which are omitted here.

On the one hand, we can find that the HTN-1 scheme converges with fluctuation. This phenomenon is reasonable. From (34) and (42) we can know that, the severer interference the MTCDs and UEs suffer, the smaller the optimal transmission powers are. When the transmission powers reduce, the mutual interference also reduce, and the new optimal transmission powers will increase and lead to the large interference again. However, the magnitude of the power change and the interference change will decrease with the game rounds (i.e., iterations). The transmission power and the EE will converge at last, and so do the total energy.

On the other hand, as depicted in Fig.5, HTN-1 takes about twice the number of iterations as much as IPCTA-NOMA. However, in each iteration, our scheme can run on the distributed architecture. The EE optimization problems can be hand out to every UE, or can be considered as a series of subtasks run in the cloud computing architecture (e.g., C-RAN [51]). Accordingly, twice the number of iterations doesn't mean that our scheme has the higher time cost. The more MTCD groups there are, the more beneficial our scheme is.

B. REQUIRED PAYLOAD OF EACH MTCD

Fig. 6 and Fig.7 show the total energy and the average transmission time of HTNs and IPCTA-NOMA corresponding to various required payload of each MTCD respectively. It is observed that with the growing required payload of each MTCD, both the total energy and the average transmission time of HTNs and IPCTA-NOMA increase. Moreover, IPCTA-NOMA outperforms HTNs in terms of the total energy. This is due to the fact that, compared to IPCTA-NOMA, there exists interference between MTCDs to UEs in HTNs, which lower the EE of UEs and MTCDs. However, on the other hand, HTNs outperforms IPCTA-NOMA in terms of the total transmission time. This is because we let MTCDs reuse the time slots of UEs in HTNs, which leads to a shorter total transmission time of HTNs when compared to that of IPCTA-NOMA.

It is also found that the larger required payload of each MTCD, the more outperformance of HTNs in terms of the total transmission time and the more outperformance of IPCTA-NOMA in terms of the total energy. With the default setting of simulation parameters, the total energy of IPCTA-NOMA is 8.35%, 19.56% and 6.34% lower than that of HTN-1, HTN-2 and HTN-3 respectively, while the total transmission time of HTN-1, HTN-2 and HTN-3 is 87.14%, 61.51% and 78.53% lower than that

Fig. 6. Total energy versus required payload of each MTCD

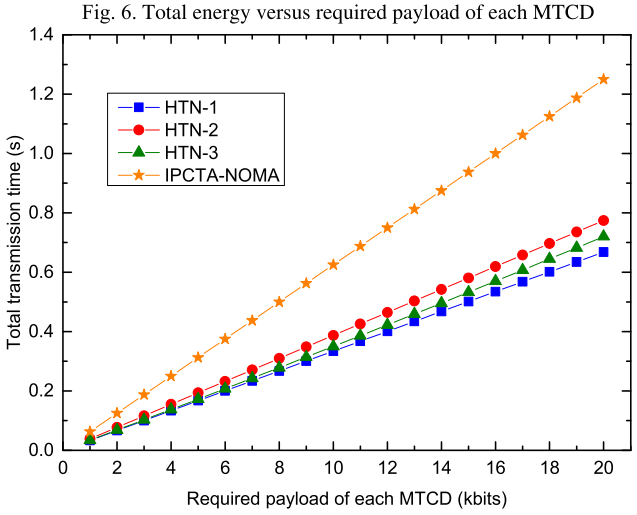

FIGURE 7. Total transmission time versus required payload of each MTCD.

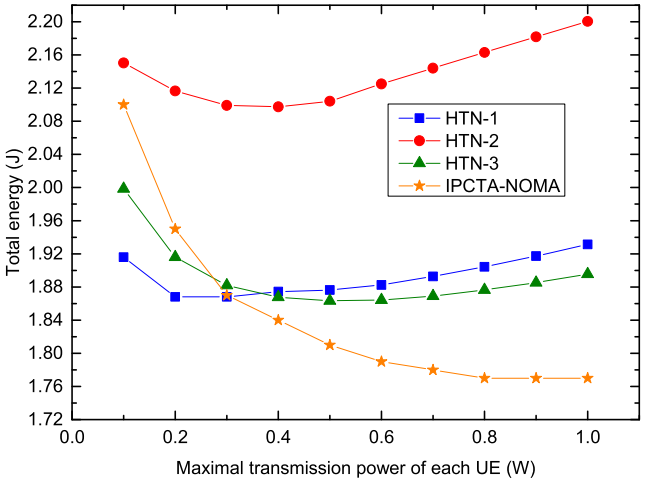

FIGURE 8. Total energy versus the maximal transmission power of each UE ($p_{i,n}^{D,C}$ = 0.5mW).

of IPCTA-NOMA respectively. Thus, with the tradeoff between the total energy consumption and the total transmission time, we consider HTNs have the better performance. Furthermore, HTN-2 has the worst performance among HTNs. And compared with HTN-3, HTN-1 performs shorter total transmission time with the price of higher total energy consumption. Besides, IPCTA-NOMA will not always outperform HTNs in terms of the total energy, which will be shown in the next simulation results.

C. MAXIMAL TRANSMISSION POWER OF EACH UE

Fig. 8 and Fig.9 show the total energy of HTNs and IPCTA-NOMA corresponding to various maximal transmission power of each UE when $p_{i,n}^{D,C} = 0.5$mW and $p_{i,n}^{D,C} = 5$mW respectively.

From Fig. 8 and Fig.9, we can see, IPCTA-NOMA outperforms HTNs in terms of the total energy in most cases when the circuit power of each MTCD is low

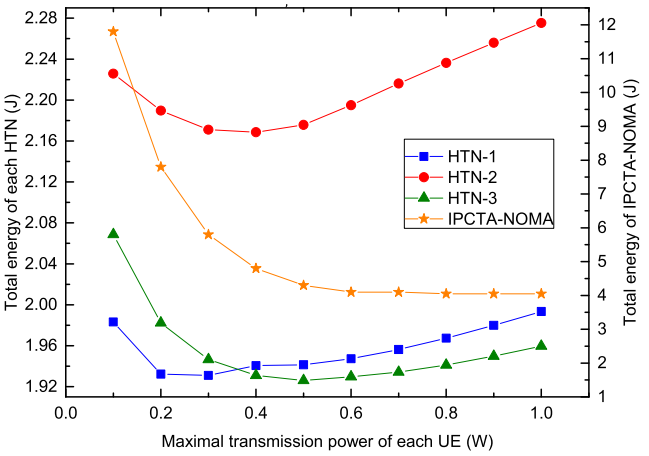

FIGURE 9. Total energy versus the maximal transmission power of each UE ($p_{i,n}^{D,C}$ = 5mW).

i.e., $p_{i,n}^{D,C} = 0.5$mW, while HTNs outperforms IPCTA-NOMA in terms of the total energy when the circuit power of each UE is high, i.e., $p_{i,n}^{D,C} = 5$mW. According to the analysis in [42], the total energy consumption of IPCTA-NOMA lies in the transmission power of MTCDs and the energy consumed by MTCGs to charge the MTCDs at low circuit power regime, while it lies in the circuit power of MTCDs and the charging consumption of MTCGs at high circuit power regime. In other words, the low transmission power of MTCDs lead to the low total energy consumption at low circuit power regime, while the high circuit power of MTCDs lead to the high total energy consumption at high circuit power regime. Thus, we consider that the reason why IPCTA-NOMA outperforms HTNs in terms of the total energy in most cases at low circuit power regime is also because we let MTCDs reuse the time slots of UEs in HTNs, and that causes the higher transmission power of MTCDs and UEs to resist the larger interference. Nevertheless, we consider that the reason why every HTN outperforms IPCTA-NOMA in terms of the total energy at high circuit power regime is because the HTNs have the shorter total transmission time, which has been shown in subsection B

Furthermore, it is found that with the growing maximal transmission power of each MTCD, the total energy of every HTN decreases first and then increases, while the total energy of IPCTA-NOMA decreases all the time. According to the analysis in [42], the monotone decreasing of the total energy in IPCTA-NOMA is due to the fact that a larger maximal transmission power of each UE ensures UEs can transmit with more power, and the required payload can be uploaded in a shorter time, which results in low total energy consumption. This analysis also applies to HTNs, where the decrease of the total energy in HTNs at low maximal transmission power validates the analysis. However, there exists interference from UEs to MTCDs in HTNs. When the transmission power of each UE is too large, the interference from UEs to MTCDs will be severe and results in the lower SE of MTCDs. As a result, the transmission time will be longer and results

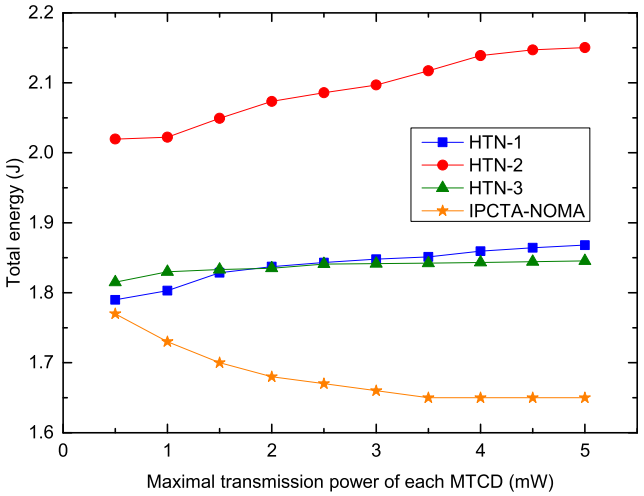

FIGURE 10. Total energy versus the maximal transmission power of each MTCD ($p_{i,n}^{D,C}$ = 0.5mW).

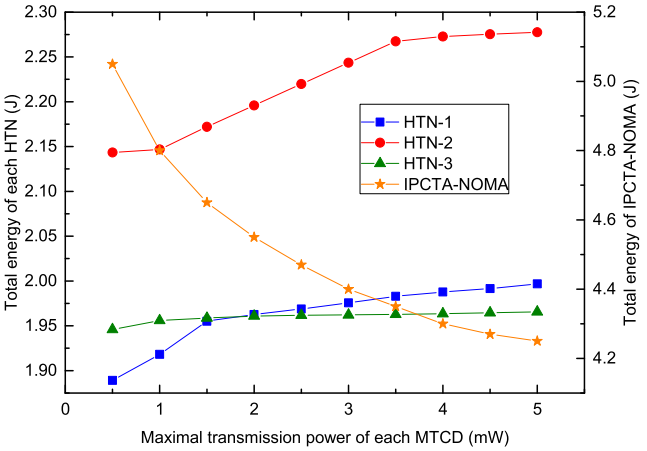

FIGURE 11. Total energy versus the maximal transmission power of each MTCD ($p_{i,n}^{D,C}$ = 5mW).

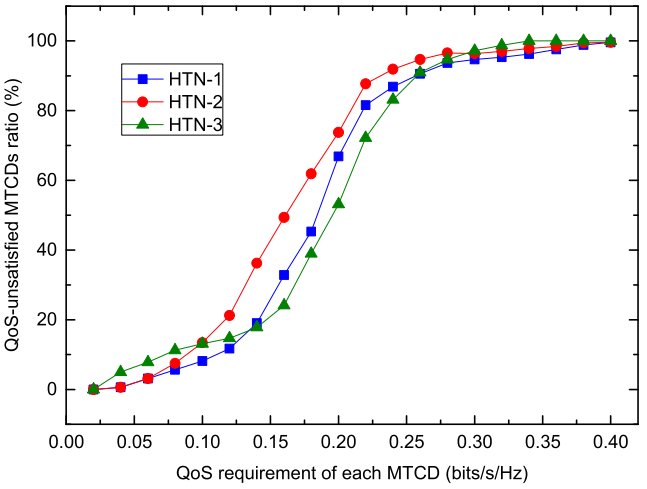

FIGURE 12. QoS-unsatisfied MTCDs ratio versus QoS requirement of each MTCD.

in high total energy consumption, which can be validated by the increase of the total energy in HTNs at high maximal transmission power.

It is also illustrated that HTN-2 always performs the highest in terms of total energy consumption among HTNs. HTN-1 outperforms HTN-3 in terms of The total energy when the maximal transmission power of each UE is low, i.e., $p_n^{U,max} \leq 0.3$W, while HTN-3 outperforms HTN-1 in terms of the total energy when the maximal transmission power of each UE is high, i.e., $p_n^{U,max} \geq 0.5$W.

D. MAXIMAL TRANSMISSION POWER OF EACH MTCD

Fig. 10 and Fig.11 show the total energy of HTNs and IPCTA-NOMA corresponding to various maximal transmission power of each MTCD when $p_{i,n}^{D,C} = 0.5$mW and $p_{i,n}^{D,C} = 5$mW respectively. Similar to Fig. 8 and Fig. 9, IPCTA-NOMA outperforms all HTNs in terms of the total energy when the circuit power of each MTCD is low (i.e., $p_{i,n}^{D,C} = 0.5$mW), while every HTN outperforms IPCTA-NOMA in terms of the total energy when the circuit power of each MTCD is high, i.e., $p_{i,n}^{D,C} = 5$mW.

Furthermore, it is found that the total energy of every HTN increases with the growing maximal transmission power of each MTCD, while the total energy of IPCTA-NOMA is on the contrary. This is also due to the fact that there exists interference from MTCDs to UEs in HTNs, but there is no such interference in IPCTA-NOMA. And the interference from MTCDs to UEs is severer relatively, such that the effect of large transmission power of MTCDs to improve their SE and shorten the transmission time is not significant. Accordingly, the larger transmission power always leads to the higher interference from MTCDs to UEs and the lower SE of UEs. As a result, the transmission time will be longer, which is on the contrary of IPCTA-NOMA.

It's also illustrated that HTN-2 always performs the highest total energy consumption among HTNs. In fact, according to the different trends of the total power of HTNs corresponding the various maximal transmission power of MTCDs and UEs, it can be concluded that the interference from MTCDs to UEs is usually severer than the interference from UEs to MTCDs. Thus, it is no wonder that HTN-2, which considers the situation that the interference from the UEs to the MTCDs is severer as described in the beginning of this section, always have the worst performance among HTNs in simulation results.

Moreover, HTN-1 outperforms HTN-3 in terms of the total energy when the maximal transmission power of each MTCD is low, i.e., $p_{i,n}^{D,max} \leq 1.5$mW, while HTN-3 outperforms HTN-1 in terms of the total energy when the maximal transmission power of each MTCD is high, i.e., $p_{i,n}^{D,max} \geq 2$mW. A similar result has been shown in the last subsection. In fact, since the interference from MTCDs to UEs is dominant, the HTN-3, which aims at mitigating this interference, will certainly perform the best effects.

E. QOS-UNSATISFIED MTCDS RATIO

Fig. 12 shows the QoS-unsatisfied MTCDs ratio corresponding to various QoS requirement of each MTCD. It can be found that the QoS-unsatisfied MTCDs ratio of every HTN increases rapidly and then slowly with the growing QoS requirement of each MTCD. All HTNs can run with the 0% QoS-unsatisfied MTCDs when $C_{i,n}^{D,min} \leq 0.02$bits/s/Hz. When $C_{i,n}^{D,min} \leq 0.12$bits/s/Hz and $C_{i,n}^{D,min} \geq 0.26$bits/s/Hz, HTN-1 shows the lowest ratio, while HTN-3 shows the lowest ratio when 0.12bits/s/Hz$< C_{i,n}^{D,min} <$ 0.26bits/s/Hz. In general, both HTN-1 and HTN-3 have their merits, as shown in all simulation results.

VII. CONCLUSION

In this paper, we propose an energy-efficient resource allocation scheme with Hybrid TDMA-NOMA in LTE-A-enabled M2M networks. First, we configure the UEs as the MTCGs. Then, we propose the improved time sharing scheme. Next, we formulate the resource allocation problems as a noncooperative game. Finally, in order to obtain the optimal EE of MTCDs and UEs in the game to confirm the power allocation, we transform each non-convex optimization problem into the convex form by using nonlinear fractional programming, and solve the transformed problem by Dinkelbach's method and Lagrangian duality theory.

The simulation results show that the total energy consumption of our scheme outperforms the existing works when the circuit power of each MTCD is low, or when the circuit power of each MTCD is high and the maximal transmission power of each UE is low. More importantly, our scheme outperforms the existing works in terms of the total transmission time. That is, even in the worst case, our scheme can dramatically shorten the total transmission time at the cost of a little more total energy when compared to the existing work, which is more suitable for supporting the numerous MTCDs in the time domain.

APPENDIX

A. DISADVANTAGES OF CONSTANT POWER IN T_N FOR UE N

In this situation, the EE optimization problem for UE n is given by (12), where C_n^U and $p_n^{U,t}$ are given by (49) and (50), which are rewritten by (19) and (20) with $p_{n',n}^U = p_n^U$.

$$C_n^U = \sum_{\substack{1\leq n'\leq N \\ n'\neq n}} \frac{s_{n',n}^D t_n}{T} log_2\left(1+\frac{p_n^U g_n^e}{I_{n',n}^D + N_0}\right) \tag{49}$$

$$p_n^{U,t} = \frac{1}{\eta}\sum_{\substack{1\leq n'\leq N \\ n'\neq n}} \frac{s_{n',n}^D t_n}{T} p_n^U + p_n^{U,C} \tag{50}$$

Following the discussion in subsection C of section V. The corresponding Lagrangian function is given by

$$L_n^U\left(P_n^U, \alpha_n^D\right) = e_n^U p_n^{U,t}\left(P_n^U\right) - C_n^U\left(P_n^U\right) + \alpha_n^U\left(C_n^U\left(P_n^U\right) - \sum_{i\in G_n} C_{i,n}^D\right) \tag{51}$$

where $P_n^U = \{p_n^U | 0 \leq p_n^U \leq p_n^{U,max}\}$. From $\nabla_{P_n^U} L_n^U\left(P_n^U, \alpha_n^D\right)$ we have

$$\sum_{\substack{1\leq n'\leq N \\ n'\neq n}} \frac{s_{n',n}^D t_n}{T}\left(\frac{e_n^U}{\eta} - \frac{\left(1+\alpha_n^U\right) log_2 e}{\frac{\sum_{i\in G_{n'}} p_{i,n',n}^D g_{i,n'}^n + N_0}{g_n^e} + p_n^U}\right) = 0 \tag{52}$$

Here we can see, finding a $\left[p_n^U\right]_0^{p_n^{U,max}}$ satisfying (52) is difficult. On the other hand, when $p_n^U = p_{n',n}^U$, equation (52) at least has a solution set consists of $\frac{e_n^U}{\eta} - \frac{\left(1+\alpha_n^U\right) log_2 e}{\frac{\sum_{i\in G_{n'}} p_{i,n',n}^D g_{i,n'}^n + N_0}{g_n^e} + p_{n',n}^U} = 0$ for $n' = 1, 2, \ldots, N, n' \neq n$, which is exactly (42) in our scheme.

B. PROOF OF THEOREM 2

If consider C_2, the Lagrangian function associated (28) is given by

$$\begin{aligned} L_{i,n}^D\left(P_{i,n}^D, \beta_{i,n}^D, \gamma_{i,n}^D, \delta_{i,n}^D, \varepsilon_{i,n}^D\right) &= e_{i,n}^D p_{i,n}^{D,t}\left(P_{i,n}^D\right) - C_{i,n}^D\left(P_{i,n}^D\right) \\ &+ \beta_{i,n}^D\left(C_{i,n}^{D,min} - C_{i,n}^D\left(P_{i,n}^D\right)\right) \\ &+ \gamma_{i,n}^D\left(C_{i,n}^D\left(P_{i,n}^D\right) - C_{i,n}^{D,max}\right) \\ &+ \sum_{\substack{1\leq m\leq N \\ m\neq n}} \delta_{i,n,m}^D\left(p_{i,n,m}^D - p_{i,n}^{D,max}\right) \\ &+ \sum_{\substack{1\leq m\leq N \\ m\neq n}} \varepsilon_{i,n,m}^D\left(-p_{i,n,m}^D\right) \end{aligned} \tag{53}$$

where $\delta_{i,n}^D = \{\delta_{i,n,m}^D | m = 1, 2, \ldots, N, m \neq n\}$ and $\varepsilon_{i,n,m}^D \{\varepsilon_{i,n,m}^D | m = 1, 2, \ldots, N, m \neq n\}$ are the Lagrangian multiplier sets associated with C_2. The associated KKT conditions is given by

$$\begin{cases} \nabla_{P_{i,n}}^D L_{i,n}^D\left(P_{i,n}^D, \beta_{i,n}^D, \gamma_{i,n}^D, \delta_{i,n}^D, \varepsilon_{i,n}^D\right) = 0 & (54-1) \\ C_{i,n}^{D,min} - D_{i,n}^D\left(P_{i,n}^D\right) \leq 0 & (54-2) \\ C_{i,n}^D\left(P_{i,n}^D\right) - C_{i,n}^{D,max} \leq 0 & (54-3) \\ p_{i,n,m}^D - p_{i,n}^{D,max} \leq 0, 1 \leq m \leq N, m \neq n & (54-4) \\ -p_{i,n,m}^D \leq 0, 1 \leq m \leq N, m \neq n & (54-5) \\ \beta_{i,n}^D\left(C_{i,n}^{D,min} - C_{i,n}^D(P_{i,n}^D)\right) = 0 & (54-6) \\ \gamma_{i,n}^D\left(C_{i,n}^D(P_{i,n}^D) - C_{i,n}^{D,max}\right) = 0 & (54-7) \\ \delta_{i,n}^D\left(p_{i,n,m}^D - p_{i,n}^{D,ma}\right) = 0 & (54-8) \\ \varepsilon_{i,n}^D\left(-p_{i,n,m}^D\right) = 0 & (54-9) \\ \beta_{i,n}^D \geq 0, \gamma_{i,n}^D \geq 0, \delta_{i,n}^D \geq 0, \varepsilon_{i,n}^D \geq 0 & (54-10) \end{cases} \tag{54}$$

Here we can see the Lagrangian function (53) and the KKT conditions (54) are complex when considering C_2. The $P_{i,n}^D$ which satisfies (54) and minimizes (53) is the $P_{i,n}^{D,*}$.

First, the $P^D_{i,n}$ satisfies (54-1) is given by (55).

$$\hat{p}^D_{i,n,m} = \frac{\left(1+\beta^D_{i,n}-\gamma^D_{i,n}\right)\eta log_2 e}{e^D_{i,n}+\eta\delta^D_{i,n}-\eta\varepsilon^D_{i,n}} - \frac{I^{D,A}_{i,n,m}+I^G_{i,n,m}+N_0}{g^n_{i,n}} \tag{55}$$

Next, the $\hat{p}^D_{i,n,m}$ corresponding to case (a) (b) and (c) in subsection C of section V, which can be obtained by the similar discussion about different cases in (54-6) and (54-8), are given by

$$\hat{p}^D_{i,n,m} = \frac{\eta log_2 e}{e^D_{i,n}+\eta\delta^D_{i,n}-\eta\varepsilon^D_{i,n}} - \frac{I^{D,A}_{i,n,m}+I^G_{i,n,m}+N_0}{g^n_{i,n}} \tag{56}$$

$$\hat{p}^D_{i,n,m} = \frac{\left(1+\beta^D_{i,n}\right)\eta log_2 e}{e^D_{i,n}+\eta\delta^D_{i,n}-\eta\varepsilon^D_{i,n}} - \frac{I^{D,A}_{i,n,m}+I^G_{i,n,m}+N_0}{g^n_{i,n}} \tag{57}$$

$$\hat{p}^D_{i,n,m} = \frac{\left(1-\gamma^D_{i,n}\right)\eta log_2 e}{e^D_{i,n}+\eta\delta^D_{i,n}-\eta\varepsilon^D_{i,n}} - \frac{I^{D,A}_{i,n,m}+I^G_{i,n,m}+N_0}{g^n_{i,n}} \tag{58}$$

Then, taking (57) for example, there are also different cases in (54-8) and (54-9). Note that $\beta^D_{i,n}$ is known.

(a) If $\delta^D_{i,n}=\varepsilon^D_{i,n}=0$, $0<\hat{p}^D_{i,n,m}<p^{D,max}_{i,n}$ should hold. In this case, (56) is rewritten as

$$\hat{p}^D_{i,n,m} = \frac{\left(1+\beta^D_{i,n}\right)\eta log_2 e}{e^D_{i,n}} - \frac{I^{D,A}_{i,n,m}+I^G_{i,n,m}+N_0}{g^n_{i,n}} \tag{59}$$

If $0<\frac{\left(1+\beta^D_{i,n}\right)\eta log_2 e}{e^D_{i,n}}-\frac{I^{D,A}_{i,n,m}+I^G_{i,n,m}+N_0}{g^n_{i,n}}<p^{D,max}_{i,n}$ holds, $\hat{p}^D_{i,n,m}$ is given by (59).

(b) If $\delta^D_{i,n}>0$, $\varepsilon^D_{i,n}=0$, $\hat{p}^D_{i,n,m}=p^{D,max}_{i,n}>0$ should hold. In this case, (56) is rewritten as

$$\hat{p}^D_{i,n,m} = \frac{\left(1+\beta^D_{i,n}\right)\eta log_2 e}{e^D_{i,n}+\eta\delta^D_{i,n}} - \frac{I^{D,A}_{i,n,m}+I^G_{i,n,m}+N_0}{g^n_{i,n}} \tag{60}$$

If there is a $\delta^D_{i,n}>0$ making $\frac{\left(1+\beta^D_{i,n}\right)\eta log_2 e}{e^D_{i,n}+\eta\delta^D_{i,n}}-\frac{I^{D,A}_{i,n,m}+I^G_{i,n,m}+N_0}{g^n_{i,n}}=p^{D,max}_{i,n}$ hold, $\hat{p}^D_{i,n,m}=p^{D,max}_{i,n}$.

(c) If $\delta^D_{i,n}=0$, $\varepsilon^D_{i,n}>0$, $\hat{p}^D_{i,n,m}=0<p^{D,max}_{i,n}$ should hold. In this case, (56) is rewritten as

$$\hat{p}^D_{i,n,m} = \frac{\left(1+\beta^D_{i,n}\right)\eta log_2 e}{e^D_{i,n}-\eta\varepsilon^D_{i,n}} - \frac{I^{D,A}_{i,n,m}+I^G_{i,n,m}+N_0}{g^n_{i,n}} \tag{61}$$

If there is a $\varepsilon^D_{i,n}>0$ making $\frac{\left(1+\beta^D_{i,n}\right)\eta log_2 e}{e^D_{i,n}-\eta\varepsilon^D_{i,n}}-\frac{I^{D,A}_{i,n,m}+I^G_{i,n,m}+N_0}{g^n_{i,n}}=0$ hold, $\hat{p}^D_{i,n,m}=0$.

(d) If $\delta^D_{i,n}>0$, $\varepsilon^D_{i,n}>0$, $\hat{p}^D_{i,n,m}=p^{D,max}_{i,n}$ and $\hat{p}^D_{i,n,m}=0$ should hold. Clearly, this is contradictory case since $p^{D,max}_{i,n}\neq 0$.

In general, one of cases (a) (b) and (c) should hold. That is, if $0<\frac{\left(1+\beta^D_{i,n}\right)\eta log_2 e}{e^D_{i,n}}-\frac{I^{D,A}_{i,n,m}+I^G_{i,n,m}+N_0}{g^n_{i,n}}<p^{D,max}_{i,n}$ does not hold, then $\hat{p}^D_{i,n,m}=p^{D,max}_{i,n}$ or $\hat{p}^D_{i,n,m}=0$ must hold.

According to the Lagrange duality theory, function (53) is still a convex function of $\hat{p}^D_{i,n,m}$. Thus, if $\frac{\left(1+\beta^D_{i,n}\right)\eta log_2 e}{e^D_{i,n}}-\frac{I^{D,A}_{i,n,m}+I^G_{i,n,m}+N_0}{g^n_{i,n}}\geq p^{D,max}_{i,n}$, the $\min_{P^D_{i,n}} L^D_{i,n}$ must appear on $p^{D,*}_{i,n,m}=p^{D,max}_{i,n}$. On the other hand, if $\frac{\left(1+\beta^D_{i,n}\right)\eta log_2 e}{e^D_{i,n}}-\frac{I^{D,A}_{i,n,m}+I^G_{i,n,m}+N_0}{g^n_{i,n}}\leq 0$, then $\min_{P^D_{i,n}} L^D_{i,n}$ must appear on $p^{D,*}_{i,n,m}=0$. Accordingly, we have

$$\hat{p}^D_{i,n,m} = \left[\frac{\left(1+\beta^D_{i,n}\right)\eta log_2 e}{e^D_{i,n}} - \frac{I^{D,A}_{i,n,m}+I^G_{i,n,m}+N_0}{g^n_{i,n}}\right]^{p^{D,max}_{i,n}}_0 \tag{62}$$

which is exactly (37). Note that it's unnecessary to figure out the values of $\delta^D_{i,n}$ or $\varepsilon^D_{i,n}$. (36) and (39) can be proved in the similar way. That is, Theorem 2 is proved.

ACKNOWLEDGMENT

This work was supported in part by the National Natural Science Foundation of China (61873352, 61572528). Jinsong Gui is the corresponding author.

REFERENCES

[1] *Cisco Visual Networking Index (VNI) Complete Forecast for 2017–2022*, Cisco, San Jose, CA, USA, 2018.
[2] N. Xia, H.-H. Chen, and C.-S. Yang, "Radio resource management in machine-to-machine communications—A survey," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 1, pp. 791–828, 1st Quart., 2018.
[3] H. Teng, Y. Liu, A. Liu, N. N. Xiong, Z. Cai, T. Wang, and X. Liu, "A novel code data dissemination scheme for Internet of Things through mobile vehicle of smart cities," *Future Gener. Comput. Syst.*, vol. 94, pp. 351–367, May 2019.
[4] Y. Liu, A. Liu, X. Liu, and X. Huang, "A statistical approach to participant selection in location-based social networks for offline event marketing," *Inf. Sci.*, vol. 480, pp. 90–108, Apr. 2019.
[5] J. Tan, W. Liu, M. Xie, H. Song, A. Liu, M. Zhao, and G. Zhang, "A low redundancy data collection scheme to maximize lifetime using matrix completion technique," *EURASIP J. Wireless Commun. Netw.*, vol. 2019, Jan. 2019, Art. no. 5. doi: 10.1186/s13638-018-1313-0.
[6] X. H. Deng, J. Luo, L. F. He, Q. Liu, X. Li, and L. Cai, "Cooperative channel allocation and scheduling in multi-interface wireless mesh networks," *Peer-Peer Netw. Appl.*, vol. 12, no. 1, pp. 1–12, Jan. 2019.
[7] J. Luo, X. Deng, H. Zhang, and H. Qi, "QoE-driven computation offloading for edge computing," *J. Syst. Archit.*, vol. 97, pp. 34–39, Aug. 2019. doi: 10.1016/j.sysarc.2019.01.019.
[8] S. Andreev, O. Galinina, A. Pyattaev, and M. Gerasimenko, "Understanding the IoT connectivity landscape: A contemporary M2M radio technology roadmap," *IEEE Commun. Mag.*, vol. 53, no. 9, pp. 32–40, Sep. 2015.
[9] A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, and M. Ayyash, "Internet of Things: A survey on enabling technologies, protocols, and applications," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 4, pp. 2347–2376, 4th Quart., 2015.
[10] W. Zhang, W. Liu, T. Wang, A. Liu, Z. Zeng, H. Song, and S. Zhang, "Adaption resizing communication buffer to maximize lifetime and reduce delay for WVSNs," *IEEE Access*, vol. 7, pp. 48266–48287, 2019.
[11] J. Li, W. Liu, T. Wang, H. Song, X. Li, F. Liu, and A. Liu, "Battery-friendly based relay selection scheme to prolong lifetime for sensor nodes in Internet of Things," *IEEE Access*, vol. 7, no. 1, pp. 33180–33201, 2019.
[12] F. Ghavimi and H.-H. Chen, "M2M communications in 3GPP LTE/LTE-A networks: Architectures, service requirements, challenges, and applications," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 2, pp. 525–549, 2nd Quart., 2015.

[13] Y. Zhang, R. Yu, S. Xie, W. Yao, Y. Xiao, and M. Guizani, "Home M2M networks: Architectures, standards, and QoS improvement," *IEEE Commun. Mag.*, vol. 49, no. 4, pp. 44–52, Apr. 2011.

[14] S.-Y. Lien, K.-C. Chen, and Y. Lin, "Toward ubiquitous massive accesses in 3GPP machine-to-machine communications," *IEEE Commun. Mag.*, vol. 49, no. 4, pp. 66–74, Apr. 2011.

[15] G. Zhang, A. Li, K. Yang, L. Zhao, Y. Du, and D. Cheng, "Energy-efficient power and time-slot allocation for cellular-enabled machine type communications," *IEEE Commun. Lett.*, vol. 20, no. 2, pp. 368–371, Feb. 2016.

[16] M. Chen, J. Wan, S. Gonzalez, X. Liao, and V. C. M. Leung, "A survey of recent developments in home M2M networks," *IEEE Commun. Surveys Tuts.*, vol. 12, no. 1, pp. 98–114, 1st Quart., 2014.

[17] A. G. Gotsis, A. S. Lioumpas, and A. Alexiou, "M2M scheduling over LTE: Challenges and new perspectives," *IEEE Veh. Technol. Mag.*, vol. 7, no. 3, pp. 34–39, Sep. 2012.

[18] G. Zhang, J. Qian, S. Xiao, and J. Gu, "Hierarchical resource allocation scheme for M2M communications enabled by cellular networks," in *Proc. 16th Int. Symposium Modeling Optim. Mobile, Ad Hoc, Wireless Netw. (WiOpt)*, Shanghai, China, May 2018, pp. 1–6.

[19] J. Gui, L. Hui, and X. Zhou, "Improving lifetime of cell-edge smart sensing devices by incentive architecture based on dynamic charging," *IEEE Access*, vol. 7, no. 1, pp. 72703–72715, 2019.

[20] L. Yin, J. Gui, and Z. W. Zeng, "Improving energy efficiency of multimedia content dissemination by adaptive clustering and D2D multicast," *Mobile Inf. Syst.*, vol. 2019, Mar. 2019, Art. no. 5298508.

[21] J. Gui, Z. Li, and Z. Zeng, "Improving energy-efficiency for resource allocation by relay-aided in-band D2D communications in C-RAN-based systems," *IEEE Access*, vol. 7, pp. 8358–8375, 2018.

[22] Z. Li, J. Gui, N. Xiong, and Z. Zeng, "Energy-efficient resource sharing scheme with out-band D2D relay-aided communications in C-RAN-based underlay cellular networks," *IEEE Access*, vol. 7, pp. 19125–19142, 2019.

[23] Y. Saito, Y. Kishiyama, A. Benjebbour, T. Nakamura, A. Li, and K. Higuchi, "Non-orthogonal multiple access (NOMA) for cellular future radio access," in *Proc. IEEE Veh. Technol. Conf.*, Dresden, Germany, Jun. 2013, pp. 1–5.

[24] 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.

[25] Z. Yang, W. Xu, and Y. Li, "Fair non-orthogonal multiple access for visible light communication downlinks," *IEEE Wireless Commun. Lett.*, vol. 6, no. 1, pp. 66–69, Feb. 2017.

[26] Z. Yang, W. Xu, H. Xu, J. Shi, and M. Chen, "Energy efficient non-orthogonal multiple access for machine-to-machine communications," *IEEE Commun. Lett.*, vol. 21, no. 4, pp. 817–820, Apr. 2017.

[27] Z. Ding, Y. Liu, J. Choi, Q. Sun, M. Elkashlan, C.-L. I, and H. V. Poor, "Application of non-orthogonal multiple access in LTE and 5G networks," *IEEE Commun. Mag.*, vol. 55, no. 2, pp. 185–191, Feb. 2017.

[28] K. Higuchi and A. Benjebbour, "Non-orthogonal multiple access (NOMA) with successive interference cancellation for future radio access," *IEICE Trans. Commun.*, vol. 98, no. 3, pp. 403–414, Mar. 2015.

[29] K. Higuchi and Y. Kishiyama, "Non-orthogonal access with random beamforming and intra-beam SIC for cellular MIMO downlink," in *Proc. IEEE Veh. Technol. Conf. (VTC Fall)*, Las Vegas, NV, USA, Sep. 2013, pp. 1–5.

[30] Z. Ding, L. Dai, R. Schober, and H. V. Poor, "NOMA meets finite resolution analog beamforming in massive MIMO and millimeter-wave networks," *IEEE Commun. Lett.*, vol. 21, no. 8, pp. 1879–1882, Aug. 2017.

[31] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Nonorthogonal multiple access for 5G and beyond," *Proc. IEEE*, vol. 105, no. 12, pp. 2347–2381, Dec. 2017.

[32] M. F. Kader, M. B. Shahab, and S. Y. Shin, "Exploiting non-orthogonal multiple access in cooperative relay sharing," *IEEE Commun. Lett.*, vol. 21, no. 5, pp. 1159–1162, May 2017.

[33] A. Zafar, M. Shaqfeh, M.-S. Alouini, and H. Alnuweiri, "On multiple users scheduling using superposition coding over rayleigh fading channels," *IEEE Commun. Lett.*, vol. 17, no. 4, pp. 733–736, Apr. 2013.

[34] F. Ghavimi, Y.-W. Lu, and H.-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.

[35] A. Aijaz, M. Tshangini, M. R. Nakhai, X. Chu, and A.-H. Aghvami, "Energy-efficient uplink resource allocation in LTE networks with M2M/H2H co-existence under statistical QoS guarantees," *IEEE Trans. Commun.*, vol. 62, no. 7, pp. 2353–2365, Jul. 2014.

[36] U. Tefek and T. J. Lim, "Relaying and radio resource partitioning for machine-type communications in cellular networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 2, pp. 1344–1356, Feb. 2017.

[37] M. Elkourdi, A. Mazin, E. Balevi, and R. D. Gitlin, "Enabling slotted Aloha-NOMA for massive M2M communication in IoT networks," in *Proc. IEEE 19th Wireless Microw. Technol. Conf. (WAMICON)*, Sand Key, FL, USA, Apr. 2018, pp. 1–4.

[38] A. D. Shoaei, M. Derakhshani, and T. Le-Ngoc, "A NOMA-enhanced reconfigurable access scheme with device pairing for M2M networks," *IEEE Access*, vol. 7, pp. 32266–32275, 2019.

[39] T. Lv, Y. Ma, J. Zeng, and P. T. Mathiopoulos, "Millimeter-wave noma transmission in cellular M2M communications for Internet of Things," *IEEE Internet Things J.*, vol. 5, no. 3, pp. 1989–2000, Jun. 2018.

[40] Y. Wu, G. Kang, and N. Zhang, "Random access and resource allocation for the coexistence of NOMA-based and OMA-based M2M communications," *China Commun.*, vol. 14, no. 6, pp. 43–53, 2017.

[41] A. B. Rozario and M. F. Hossain, "An architecture for M2M communications over cellular networks using clustering and hybrid TDMA-NOMA," in *Proc. 6th Int. Conf. Inf. Commun. Technol. (ICoICT)*, Bandung, Indonesia, May 2018, pp. 18–23.

[42] Z. Yang, W. Xu, Y. Pan, C. Pan, and M. Chen, "Energy efficient resource allocation in machine-to-machine communications with multiple access and energy harvesting for IoT," *IEEE Internet Things J.*, vol. 5, no. 1, pp. 229–245, Feb. 2018.

[43] B. Gu, C. Zhang, H. Wang, Y. Yao, and X. Tan, "Power control for cognitive M2M communications underlaying cellular with fairness concerns," *IEEE Access*, vol. 7, pp. 80789–80799, 2019.

[44] H. Kwon and T. G. Birdsall, "Channel capacity in bits per joule," *IEEE J. Ocean. Eng.*, vol. OE-11, no. 1, pp. 97–99, Jan. 1986.

[45] Z. Zhou, M. Dong, K. Ota, G. Wang, and L. T. Yang, "Energy-efficient resource allocation for D2D communications underlaying cloud-RAN-based LTE-A networks," *IEEE Internet Things J.*, vol. 3, no. 3, pp. 428–438, Jun. 2016.

[46] C. Y. Wong, R. S. Cheng, K. B. Lataief, and R. D. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *IEEE J. Sel. Areas Commun.*, vol. 17, no. 10, pp. 1747–1758, Oct. 1999.

[47] W. Dinkelbach, "On nonlinear fractional programming," *Manage. Sci.*, vol. 13, no. 7, pp. 492–498, Mar. 1967.

[48] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[49] K. T. K. Cheung, S. Yang, and L. Hanzo, "Achieving maximum energy-efficiency in multi-relay OFDMA cellular networks: A fractional programming approach," *IEEE Trans. Commun.*, vol. 61, no. 8, pp. 2746–2757, Jul. 2013.

[50] M. J. Osborne and A. Rubinstein, *A Course in Game Theory*. Cambridge, MA, USA: MIT Press, 1994.

[51] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. S. Berger, and L. Dittmann, "Cloud RAN for mobile networks—A technology overview," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 405–426, 1st Quart., 2015.

ZEMING LI is currently pursuing the master's degree with the Department of Computer Science and Technology, School of Computer Science and Engineering, Central South University, China. His research interests include the Internet of Things, wireless sensor networks, network simulation, and performance evaluation.

JINSONG GUI received the M.S. and Ph.D. degrees from Central South University, China, in 2004 and 2008, respectively, where he is currently a Professor with the School of Computer Science and Engineering. His research interests include the general areas of distributed systems and related fields such as wireless network topology control, performance evaluation, and network security.