Received November 9, 2018, accepted December 18, 2018, date of publication January 14, 2019, date of current version January 29, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2890820

Two-Timescale Resource Allocation for Wireless Powered D2D Communications With Self-Interested Nodes

LI FENG[1], QINGHAI YANG[2,3], KYEHYUN KIM[4], AND KYUNG SUP KWAK[1]

[1]Department of Information and Communication Engineering, Inha University, Incheon 22212, South Korea
[2]State Key Laboratory on ISN, School of Telecommunications Engineering, Xidian University, Xi'an 710071, China
[3]Collaborative Innovation Center of Information Sensing and Understanding, Xidian University, Xi'an 710071, China
[4]Department of Geoinformatic Engineering, Inha University, Incheon 22212, South Korea

Corresponding author: Kyung Sup Kwak (kskwak@inha.ac.kr)

This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korea Government (MSIT) under Grant NRF-2017R1A2B2012337 and Grant 61801365.

ABSTRACT Wireless powered device-to-device (D2D) communications can empower D2D communications via a power station (PS) in a self-sustainable manner. In this paper, aided by an incentive mechanism, we investigate the dynamic resource allocation for content transmission in wireless powered D2D communications with self-interested nodes including D2D transmitters (D2D-Txs) and PS. Stochastic optimization is developed to maximize the average network utility under the constraints of the limited data buffer, energy capacity, and incentives. Then, to coordinate the different timescales for the network state variation, the problem of stochastic optimization is converted into two subproblems via two-timescale Lyapunov optimization technique. Specifically, at the large timescale where the D2D-Tx's resource states (data queue, energy resource, and incentive) change slowly, we obtain the solution of a joint rate adaption and energy trading problem. While, at the small timescale where the channel state experiences rapid variation, we develop a closed-form expression of power factor by solving a transmit power control problem, which is non-convex owing to the interference among D2D-Txs. Additionally, an online two-timescale resource allocation (OTTRA) algorithm is proposed, and the performance bounds of the algorithm are characterized theoretically in terms of the utility-delay tradeoff. The numerical results exhibit that the OTTRA algorithm not only encourages the cooperative content transmission and ET among self-interested nodes, but also ensures a satisfied network performance in the long term.

INDEX TERMS Wireless powered D2D communications, two timescales, self-interested nodes.

I. INTRODUCTION

With the upsurge growth of data traffic and the increasing number of smart phones, device-to-device (D2D) communications have recently been widely studied as one of the key technologies to realize the 5G communication networks. By taking advantage of cooperative transmitters, D2D communications can improve the network performance (e.g., transmission capacity, delay, etc.) through direct communications between D2D devices without traversing the base station (BS) [1]. In such D2D networks, most D2D transmitters (D2D-Txs) are powered by limited battery capacity, and shall be recharged or replaced when their energy resource is depleted, which will hinder the fruitful development of D2D communication networks. Moreover, the devices in practical D2D communications are always endowed with smart autonomic functions, which are natural self-interested nodes [2]. These self-interested nodes just care about their own profits and will present selfishness, i.e., be unwilling to participate in cooperative content transmission. In D2D communication system, there are many kinds of selfish behaviors for the self-interested nodes, e.g., the hesitating caching for different social ties [3], unfaithful reporting for privacy protection and so on. The most common is that the D2D-Tx as an individual may refuse to forward the data for others to save their finite resource (the limited energy, valuable CPU resources, etc.) in the content dissemination process. In [4], the self-interested mobile terminals might be unwilling to form a virtual MIMO and relay the data for others via D2D links, since they would incur power costs. To save the valuable energy resource, the self-interested users in [5] wouldn't want to spread the contents through cooperative D2D communications. The existing schemes proposed in [6] and [7] which implicitly assume the D2D-Txs are altruist, i.e., they act as relays whenever required, will not work and lead to the overestimation of the network performance.

Recently, the wireless power transfer (WPT) techniques, which act as a convenient and perpetual power supply solution for wireless devices in the air, have attracted much attention. The employment of WPT into D2D networks leads to the emergence of wireless power D2D communications. Although wireless powered D2D communications can improve the network capacity and prolong the network lifetime, there are many research challenges remained to be addressed for the resource allocation. Firstly, owing to the introduction of the WPT technology, the autonomous power station (PS) may also not want to transmit its own power resource to D2D-Tx without any directly benefit [8]. The self-interested nodes (i.e., D2D-Txs and PS) in the wireless powered D2D communications will impair the effectiveness of content transmission through wireless links and further degrade the network performance. As a result, the incentive mechanism needs to be proposed to stimulate the cooperative transmission for D2D-Txs, and the energy trading (ET) between PS and D2D-Txs. Secondly, D2D-Tx's resources vary with its content transmissions, and the wireless link for D2D communication pair may experience fading. To copy with these dynamic network states, a dynamic resource allocation scheme jointly performing rate adaption (RA), ET and transmit power controls (TPCs) shall be raised under the constraints of limited data buffer, energy capacity and incentives. Thirdly, D2D-Tx's resource states in the practical systems always change much slower than the channel fading [9]. Hence, a two-timescale network decisions shall be designed to coordinate the different timescale network states, which further increase the difficulty of the scheme. Besides, the coupling among D2D communication pairs caused by the interference may lead to a non-convex optimization problem, which is difficult to solve. Based on the analysis above, it is necessary to propose a dynamic resource allocation scheme based on different timescale network states, to improve the long-term network performance for wireless powered D2D communications with self-interested nodes.

In this paper, we focus on the effective content transmission in wireless powered D2D communications where the D2D-Tx with limited energy capacity can harvest wireless power from a PS and then forward the contents to others. In such networks, the nodes are self-interested, i.e., the PS is unwilling to transmit power to D2D-Txs and the D2D-Tx will do not want to forward the contents for others. Moreover, the variation of D2D-Tx's resource state is much slower than that of channel state. To deal with the different timescale network states and decrease the negative influence of the self-interested nodes on content transmission, we develop a stochastic optimization to maximize the average network utility under the constraints of limited data buffer, energy capacity and incentives. By virtue of two-timescale Lyapunov optimization technique, the optimization problem is decomposed into two subproblems in terms of different timescale network states, which are the joint RA and ET problem for the large timescale, and the TPC problem for the small timescale.

The main contributions of this paper are outlined as follows.

- We develop an incentive mechanism, to encourage the cooperative content transmission for D2D-Txs and the ET between PS and D2D-Txs.
- Under the constraints of limited data buffer, energy capacity and incentives, a stochastic optimization is developed to maximize the average network utility in the long term.
- To cope with the different timescale network states, we present an online two-timescale resource allocation (OTTRA) algorithm, and give the performance bounds of the algorithm theoretically for the utility-delay tradeoff. In OTTRA algorithm, the resource allocation is executed based solely on the current network states, and the communication overhead imposed is shown to be greatly reduced.

The rest of the paper is organized as follows. Section II presents the related works. The system model, incentive mechanism and problem formulation are given in Section III. In Section IV, an OTTRA algorithm is proposed, while Section V presents the network performance of the proposed scheme. Section VI introduces the simulation results to evaluate the proposed scheme. The conclusion is drawn in Section VII.

II. RELATED WORKS

Recently, many resource allocations including RA and TPC for D2D communications were studied and analyzed [10]–[13]. An iterative power allocation algorithm was proposed in [10] for optimizing the energy efficiency (EE) in D2D communications under the transmit power constraint. When considering the interference among D2D communication pairs, a stable matching approach was adopted in [11] to improve the EE in D2D enabled cellular networks. However, they all focus on the static situation, e.g., the network states including network nodes' resource states and channel state, keep constant. By jointly considering the time-varying channels and energy resource, a distributed power allocation scheme was designed in [12] for overlay D2D communications, while a joint resource allocation and power control strategies was investigated in [13] for D2D communications. Nevertheless, none of the literatures considers the battery capacity limitation of D2D-Tx which is unavoidable in the practical D2D communication networks.

To address the energy bottleneck in D2D communication networks, energy harvesting has been considered as a promising solution to provide power for the D2D-Txs [14]. By introducing energy harvesting mechanism into the traditional D2D communications, a solar energy harvesting scheme was exploited in [15] to execute power allocation for the D2D system. Since energy can be harvested from radio-frequency (RF) signals over small distances in the air, WPT technology is more reliable than natural renewable resources and fully controllable [16]. By employing WPT technology into D2D communications, a harvest-transmit-store model and resource allocation problem were built in [17] to maximize the average throughput for D2D communication. The research employs WPT technology for resource allocation in D2D communication networks, but it neglects the potential self-interested nodes in the cooperative content transmission.

Owing to the precondition of the selfish player, game theories were employed by some literatures [18], [19] to deal with the self-interested nodes when executing resource allocation in D2D communications. A non-cooperative game was employed in [18] for solving the resource allocation problem with self-interested user equipments (UEs), while an ET based Stackelberg game was studied in [19] to analyze the hierarchical interaction between the selfish BS and D2D-Txs. For energy-efficient resource allocation with intra-cell and inter-cell interference, a noncooperative game was introduced in [20] to model the strategy iterations of selfish UEs in D2D communications underlaying LTE-A networks. The games above indeed stimulate the selfish nodes to cooperate with each other, but the assumption that nodes' available resources keep constant, may not be valid in the practical D2D systems. Jointly considering the selfish D2D-Txs and dynamic network states, some selfishness management schemes have been studied for efficient resource allocation in wireless networks. A dynamic social trust associations protocol was designed in [21] to deal with the presence of selfish D2D-Txs over D2D communications, while it just paid attention to the trust's aggregation based on the observation. Markov Decision Process (MDP) as one of the dynamic programming method, was employed in [22] to map D2D-Txs' discrete resource states onto its selfishness levels for D2D communications. However, when the number of system states increases, the process is difficult to converge [23]. Lyapunov optimization [24] is an effective technique for solving problems of long-term performance optimizations. Compared with MDP, Lyapunov based scheme dose not require the prior knowledge of the network states and has been widely applied in practical networks. The previous works [25], [26] have made some attempts to deal with nodes' dynamic selfishness in the wireless networks via stochastic optimization. A virtual selfishness queue and an application example were given in [25] to regulate node's dynamic selfish property, but it did not analyze the proposed application example in detail. A stochastic differential equation and an adaptive-compensation algorithm were presented in [26] to handle nodes' continuous selfishness based on continuous-time Lyapunov optimization, while it ignored the interference among nodes due to spectrum reuse. Moreover, none of them employs the WPT technology for sustainable communication and considers the influence of data queue

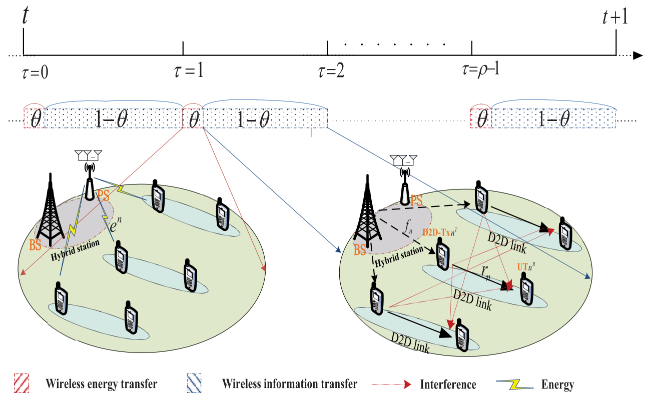


FIGURE 1. Wireless-powered D2D communication network with selfish nodes.

dynamics on the network performance, which may lead to unacceptable degradation in transmission delay and the network unstability [27].

III. SYSTEM MODEL

A. NETWORK MODEL

We consider a wireless powered D2D communication network, in which a BS, who acts as a content server, pushes the contents to the network users. Due to the limited transmission range, BS can not directly serve some users which are denoted as user terminals (UTs). In this situation, we assume that the intermediate user who receives the contents from BS can cache the contents in its finite buffer, and then acts as a D2D-Tx to forward the contents to a specifically requested UT. Each D2D-Tx and its corresponding UT make up a D2D communication pair. In the considered scenario, there exist N D2D communication pairs in total, which share the same frequency resource and introduce interference to each other, as shown in Fig. 1. The set of communication pairs is denoted by $\mathcal{N}$, $\mathcal{N} = \{1, 2, \cdots, N\}$. The notation $n^T \in \mathcal{N}^T$ denotes the D2D-Tx which forwards the contents for communication pair n, and $n^R \in \mathcal{N}^R$ is the UT[1] which receives the contents via n^T. Moreover, there exist a set of $\mathcal{F}$ contents which are required to be transmitted from BS to D2D-Txs, in which f_n denotes the content session requested by communication pair n. We also assume that the D2D-Tx has no embedded power supply available and needs to harvest energy from the RF signals transmitted by a PS which is powered constantly by a national grid. Then, the "harvest-then-transmit" protocol is employed in this paper, which is divided into two parts: (a) Wireless energy transfer (WET), i.e., the PS provides power for the D2D-Txs via RF signals and the D2D-Txs store the harvested energy in their rechargeable batteries; (b) Wireless information transfer (WIT), i.e., the D2D-Txs employ the battery to forward the contents in their buffers to UTs. Additionally, as pointed in [28], the PS and BS can be integrated into one hybrid station, which can

[1] The notations $\mathcal{N}^T$ and $\mathcal{N}^R$ denote the sets of D2D-Txs and UTs, respectively.

switch between power transfer and content transmission over time.[2]

B. INCENTIVE MECHANISM

Note that, the success of each D2D communication depends on the willingness of D2D-Tx to forward the contents to the corresponding UT. However, in practice, the D2D-Txs are self-interested, which are different from the infrastructure D2D-Txs which are owned by the operator [22]. Since the forwarding actions consume a specified energy resource, the D2D-Tx may be unwilling to forward the received contents to UT in order to save its own precious battery energy. In this paper, the rational D2D-Txs can control their transmit power automatically to present their forwarding willingness based on preferences. Meanwhile, the smart PS is also self-interested, i.e. it does not want to transmit wireless power to D2D-Txs for free. These selfish behaviors for both D2D-Txs and PS may greatly degrade the network capacity and lifetime. To overcome this problem, a virtual token exchange mechanism [29] is exploited to stimulate the forwarding behaviors of selfish D2D-Txs and the ET between PS and D2D-Txs. When the BS wants to deliver the contents to UTs via D2D-Txs, it will pay some tokens to stimulate D2D-Txs to participate in the cooperative forwarding. Specifically, if the D2D-Tx forwards the contents to the corresponding UT, it will receive some tokens, and can spend these tokens on buying energy from the PS for empowering themselves. Otherwise, the D2D-Tx receives nothing of the reward. For the sake of fairness, the amount of tokens that the D2D-Tx obtains from the BS is related to its contribution to improve the network capacity. When D2D-Tx n^T participates in the cooperative content forwarding with transmission rate r_n, it will receive $\mathcal{R}^g r_n$ tokens as the reward, where $\mathcal{R}^g$ is the number of the received tokens for unit data session. The specific expression of transmission rate r_n for communication pair n will be given in the following subsection. Additionally, when the D2D-Tx harvests e_n, $e_n \in \{1, 2, \cdots, E\}$, energy from the PS through the ET, it must pay $\mathcal{R}^p e_n$ tokens. When the D2D-Tx exhausts its limited battery energy and has no rechargeable energy, it will be a disable node. To enable the network nodes (BS, PS and D2D-Txs) to pay and earn tokens, a Credit Clearance Center (CCC) [30] is employed to manage the tokens for the system.

C. TWO-TIMESCALE COORDINATION FOR RA, ET AND TPC

In such a network, the BS determines its outgoing flow rates regarding D2D-Txs' resource states (data queue, residual energy and holding tokens), to push the contents into networks as much as possible. While, the selfish-interested D2D-Tx decides the amount of the energy bought from PS based on its utility (related to its resource states). Besides, each D2D-Tx has to control its transmit power for forwarding UT's contents to maximize its own profit

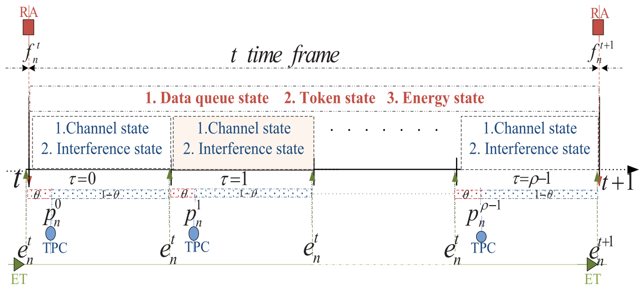


FIGURE 2. Two-timescale RA, EH and TPCs.

(depending on the channel state). Noting that the variation of D2D-Tx's available resource (energy, tokens and data queue) is much slower than that of its channel state [9]. The different changing rates of the above factors lead to different timescale network states. Moreover, both RA of BS and ET of D2D-Txs, which are implemented based on D2D-Txs' resource states, cannot afford to run at each timeslot due to the high complexity and communication overhead in collecting D2D-Txs' relevant resource information. As a result, we consider a two-timescale system to dynamically coordinate the RA for BS, ET and TPCs for D2D-Txs based on different timescale network states. Specifically, the time horizon is divided into time frames, and each time frame contains ρ short timeslots indexed by τ. We assume that the resource states of D2D-Txs keep constant during one time frame and may vary over the frame boundaries, while the channel state changes at each timeslot. At the beginning of the time frame, the RA $\mathbf{F}^t = [f_n^t]_{n\in\mathcal{N}}$ is operated by the BS and the amount of harvested energy e_n^t via energy trading in WET is decide by each D2D-Tx, according to D2D-Txs' resource states. Then, at timeslot τ, $\tau \in \{0, 1, \cdots, \rho - 1\}$, within a specific time frame t, the "harvest-then-transmit" protocol is employed. During θ duration of timeslot τ (WET stage), D2D-Tx n^T buys e_n^t energy[3] from PS using tokens. In the rest of the time $(1-\theta)\tau$ (WIT stage), the D2D-Txs may utilize the harvested energy and forward the contents to UTs for maximizing their profits. When the D2D-Txs' resource states change and the next time frame $t + 1$ comes, a new coordinate period for the RA of BS, ET and TPCs of each D2D-Tx starts, as shown in Fig. 2. In the coordinate process above, D2D-Tx's transmit power strategies for various timeslots within a time frame will impact its resource states and further its ET and BS's RA at the next time frame. The BS's RA and D2D-Txs' ETs at the current time frame also influence the network states and the transmit power strategies at the next time frame.

D. DATA QUEUING MODEL

In this paper, the channel state information (CSI) for D2D pair n' link at timeslot τ is defined as the channel gain h_{nn}^τ. We assume that h_{nn}^τ is independently and identically distributed (i.i.d.) over different timeslots. Moreover, we use

[2]The time-dependent schedule will be addressed in the following subsection.

[3]The amount of harvesting energy is decided by D2D-Tx n^T via ET at the beginning of the time frame t.

p_n^τ to designate the transmit power of D2D-Tx n^T at timeslot τ. Then, considering the interference introduced by the other D2D communication pairs, the received signal to interference-plus-noise ratio (SINR) of D2D communication pair n, $\gamma_n^\tau(a_n, \mathbf{a}_{-n})$, at timeslot τ is given by

$$\gamma_n^\tau(p_n^\tau, \mathbf{p}_{-n}^\tau) = \frac{h_{nn}^\tau p_n^\tau}{\sigma^2 + \sum_{m\in\mathcal{N}\backslash\{n\}} h_{mn}^\tau p_m^\tau}, \tag{1}$$

where σ^2 is the variance of the additive white Gaussian noise, $\mathbf{p}_{-n}^\tau = \{p_1^\tau, \cdots, p_{n-1}^\tau, p_{n+1}^\tau, \ldots, p_N^\tau\}$, h_{mn}^τ is the channel gain of the link between D2D-Tx m^T, $m^T \in \mathcal{N}^T\backslash\{n^T\}$, and UT n^R at timeslot τ. Without loss of generality and under the framework of the Shannon formula, the transmission rate for communication pair n at timeslot τ is

$$r_n^\tau = W \log\left(1 + \zeta\gamma_n^\tau(p_n^\tau, \mathbf{p}_{-n}^\tau)\right), \quad \forall t > 0,\ n\in\mathcal{N},$$

where W is the bandwidth. Then, the dynamics of the queue backlog Q_n^{t+1} ($\forall t > 0, n \in \mathcal{N}$) over different time frames is

$$Q_n^{t+1} = \max[Q_n^t - \sum_{\tau=0}^{\rho-1} r_n^\tau, 0] + f_n^t, \tag{2}$$

where $\sum_{\tau=0}^{\rho-1} r_n^\tau$ is the accumulative transmission rate, and f_n^t is the arriving flow rate within time frame t. Noting that the first term in Eq. (2) corresponds to the departure process and the second term corresponds to the arrival process. Due to the time-varying variables, both the departure and arrival process are stochastic, hence the data queue backlogs are changing over the time frames. Moreover, network stability is a natural constraint in the dynamic system [27]. To ensure that the network is strongly stable and further the data queue is bounded within its maximal capacity, we obtain the following constraint

$$\frac{1}{T}\sum_{t=0}^{T-1}\sum_{n=1}^{N} \mathbb{E}[Q_n^t] < \infty, \tag{3}$$

E. ENERGY CONSUMPTION AND SUPPLY MODEL FOR D2D-Txs

Noting that during θ duration of timeslot τ within time frame t (WET stage), D2D-Tx n^T buys e_n^t energy from the PS via wireless power transfer. The energy received and stored in its battery is $\eta e_n^t g_n$, where ϑ, $0 \leq \vartheta < 1$, is the energy harvesting efficiency and g_n is the channel power gain from BS to D2D-Tx n^T. Then, at the rest time of timeslot τ (WIT stage), D2D-Tx n^T uses power p_n^τ to accomplish its forwarding tasks. The energy consumption of D2D-Tx n^T for forwarding its received contents to UT n^R at timeslot τ is p_n^τ. Thus, the total energy consumption $P_n^{Total,t}$ of D2D-Tx n^T at time frame t is $P_n^{Total,t} = \sum_{\tau=0}^{\rho-1} p_n^\tau$, where $\sum_{\tau=0}^{\rho-1} p_n^\tau$ is the cumulative energy consumption of D2D-Tx n^T during time frame t. We assume each n^T knows its own current energy availability. By denoting the energy resource for D2D-Tx n^T at time frame t as E_n^t, the energy dynamic equation is

$$E_n^{t+1} = E_n^t + \vartheta\rho e_n^t g_n - P_n^{Total,t}, \quad \forall n \in N, \tag{4}$$

where $\vartheta\rho e_n^t g_n$ is the cumulative harvested energy for D2D-Tx n^T during time frame t. At any time frame t, the total energy consumption at D2D-Tx n^T can not exceed the current energy resource. Then, we obtain following energy-availability constraint,

$$E_n(t) \geq P_n^{Total,t}, \quad \forall n \in N. \tag{5}$$

We assume that D2D-Tx n^T is equipped with a battery having the limited capacity $\mathcal{E}_n^{max}$. At any time frame t, the total energy volume stored in the battery is limited by the battery capacity, thus the following inequality must be satisfied,

$$E_n(t) + \vartheta\rho e_n^t g_n \leq \mathcal{E}_n^{max}, \quad \forall n \in N. \tag{6}$$

F. TOKEN CONSUMPTION AND SUPPLY MODEL

Based on the proposed incentive mechanism, each D2D-Tx earns tokens by forwarding received contents, and spends tokens on buying wireless energy from PS to empower themselves via ET. Specifically, if D2D-Tx n^T forwards the contents to its corresponding UT with transmission rate r_n^τ, it will receive $\mathcal{R}^g r_n^\tau$ tokens at timeslot τ. Additionally, when the D2D-Tx buys e_n^t energy from PS, it will spend $\mathcal{R}^p e_n^t$ tokens at timeslot τ, where $\mathcal{R}^p$ is the spent tokens of the D2D-Tx for unit data session. Then, the dynamics of token queue M_n^{t+1} for D2D-Tx n^T at time frame $t+1$ is formulated as

$$M_n^{t+1} = M_n^t - \rho\mathcal{R}^p e_n^t + \mathcal{R}^g \sum_{\tau=0}^{\rho-1} r_n^\tau, \tag{7}$$

where $\rho\mathcal{R}^p e_n^t$ and $\mathcal{R}^g \sum_{\tau=0}^{\tau=\rho-1} r_n^\tau$ are the cumulative tokens that the D2D-Tx n^T spends and earns during time frame t, respectively. Similarly, the total token consumption at D2D-Tx n^T must satisfy the following constraint for time frame t

$$M_n(t) \geq \rho\mathcal{R}^p e_n^t, \quad \forall n \in N. \tag{8}$$

At any time frame t, the total token volume stored is limited by the token capacity $\mathcal{M}_n^{max}$, which is

$$M_n(t) + \mathcal{R}^g \sum_{\tau=0}^{\rho-1} r_n^\tau \leq \mathcal{M}_n^{max}, \quad \forall n \in N. \tag{9}$$

G. PROBLEM FORMULATION

To deal with the coupled system policies (i.e., RAs, ETs and TPCs) and ensure a satisfied long-term network performance, the objective of the two-timescale scheme is achieved based on an incentive mechanism. That is, the BS adjusts its outgoing data rates at each time frame to maximize the time-average network utility, and at the same time considers the self-interested nodes' strategies under the constraints of limited data buffer, energy capacity and incentives. Here, the network utility at time frame t can be expressed by choosing a function $U^t(\mathbf{F}^t)$, defined as $U^t(\mathbf{F}^t) = \sum_n \varpi_n \log(e + f_n^t)$, where e and log are the natural constant and logarithm, respectively. Noting that the utility $U^t(\mathbf{F}^t)$ is a weighted proportional fairness function, which can maximize the sum data rates outgoing from the BS while at the same time allowing all D2D communication pairs to receive at least a minimal level of data rates [31]. Then, we have following time-average optimization problem,

$$\max_{\mathbf{F}^t,\mathbf{e}^t,\mathbf{p}^t} \overline{U} = \lim_{T\to\infty} \frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}[U^t(\mathbf{F}^t)] \tag{10}$$

$$\text{s.t.}\quad 0 \le \sum_n f_n^t \le C^t, \quad \forall t \ge 0, \tag{C1}$$

$$0 \le p_n^\tau \le P_n^{max}, \quad \forall 0 \le \tau < \rho - 1, \tag{C2}$$

$$0 \le e_n^t \le o_n, \quad \forall t \ge 0, \tag{C3}$$

$$\frac{1}{T}\sum_{t=0}^{T-1}\sum_{n=1}^{N}\mathbb{E}[Q_n^t] < \infty, \tag{C4}$$

$$E_n(t) \ge P_n^{Total,t}, \quad \forall n \in N, \tag{C5}$$

$$E_n(t) + \vartheta\rho e_n^t g_n \le \mathcal{E}_n^{max}, \quad \forall n \in N, \tag{C6}$$

$$M_n(t) \ge \rho\mathcal{R}^p e_n^t, \quad \forall n \in N, \tag{C7}$$

$$M_n(t) + \mathcal{R}^g\sum_{\tau=0}^{\rho-1} r_n^\tau \le \mathcal{M}_n^{max}, \quad \forall n \in N, \tag{C8}$$

where $\mathbf{e}^t = [e_n^t]_{\{n\in N\}}$, $\mathbf{p}^t = [p_n^t, p_n^{t+1}, \cdots, p_n^{t+\rho-1}]_{\{n\in N\}}$, C^t is the total link rate budget for the BS, $P_n^{\max}$ and o_n is the maximum transmit power and harvested energy for D2D-Tx n^T, respectively. Constraint (C1) in Eq. (10) shows the total rate constraint of the links outgoing from the BS, (C2) and (C3) are the power and energy harvesting constraints for D2D-Tx; (C4) is the network stability constraint; (C5) is the energy-availability constraint which denotes that the total energy consumption for every D2D-Tx shall be below the current energy resource; (C6) is the stored energy constraint which shows that the available energy resource is bounded by the battery capacity; Similarly, (C7) is the token-availability constraint, and (C8) is the stored tokens constraint at each time frame.

Eq. (10) can be viewed as a stochastic programming. The solution is to propose an algorithm to decide flow rate $\mathbf{F}^t$, trading energy $\mathbf{e}^t$ and transmit power $\mathbf{p}^t$, such that all constraints are satisfied and the utility is maximized as large as possible. Noting that for constraint (C8) in Eq. (10), there is a tight coupling between the different timescale network states.[4] The system policies (i.e., RAs, ETs and TPCs) cannot be optimized independently. Moreover, the optimization problem in Eq. (10) involves long-term averaging of both objective function and constraint (i.e., (C4)), which cannot be directly solved by the traditional optimization techniques. To track the optimization problem, we will convert Eq. (10) into a serious of different timescale solvable subproblems over the time frames based on the two-timescale Lyapunov optimization theory [32], and propose a two-timescale resource allocation algorithm which pushes the average network utility to the optimal solution of Eq. (10).

[4]There exists a coupling between long timescale token state and short timescale channel state.

The detailed process of Eq. (10) to be solved will be described in the following section.

IV. TWO-TIMESCALE RESOURCE ALLOCATION SCHEME

We observe that the challenge behind Eq. (10) is that we shall find a dynamic resource allocation decision $\{\mathbf{F}^t, \mathbf{e}^t, \mathbf{p}^t\}$ for maximizing the average network utility under the constraints of limited data buffer, energy capacity and holding tokens. Motivated by [33], we regard the decrease of available resource (residual energy or/and holding tokens) as queue's departure process, and the increase as queue's arrival process. Then, we can construct two virtual queues, called virtual energy queue and token queue, to model the dynamics of D2D-Tx's available tokens and energy over time frame, respectively. By controlling the arrival and departure processes of the queues (data queues, virtual energy queues and token queues) appropriately via two-timescale Lyapunov *drift-plus-penalty* method, we can limit the lengths of queues, whilst maximizing the average network utility. For simplicity, in this paper, the virtual energy queue and token queues of D2D-Tx n^T are denoted as E_n^t and M_n^t, respectively. Then, let Θ^t denote a matrix containing the queues $\{(Q_n^t, E_n^t, M_n^t)|n \in \mathcal{N}\}$. We can define the quadratic Lyapunov function at time frame t as $L(\Theta^t) = \frac{1}{2}\sum_n[Q_n^t]^2 + \sum_n[-\tilde{E}_n^t]^2 + \sum_n[-\tilde{M}_n^t]^2$, where $\tilde{E}_n^t = \mathcal{E}_n^{max} - E_n^t \ge 0$ and $\tilde{M}_n^t = \mathcal{M}_n^{max} - M_n^t \ge 0$. The conditionally expected Lyapunov drift at time frame t is

$$\Delta(\Theta^t) :\triangleq \mathbb{E}[L(\Theta^{t+1})|\Theta^t] - \mathbb{E}[L(\Theta^t)], \tag{11}$$

where the expectation is taken over the randomness of departure and arrival processes of the queues.

Following from the Lyapunov optimization framework, we add the penalty term $-V\mathbb{E}[U^t(\mathbf{F}^t)|\Theta^t]$ to Eq. (11) to obtain the following *drift-plus-penalty* term

$$\Delta_V(\Theta^t) = \Delta(\Theta^t) - V\mathbb{E}[U^t(\mathbf{F}^t)|\Theta^t]. \tag{12}$$

Here, $V > 0$ is a control parameter. Through minimizing *drift-plus-penalty* term at each time frame, we can limit the increases of queues' lengths, and also improve the network utility. The network's objective is further achieved. Then, we have the following theorem regarding the *drift-plus-penalty* term.

Theorem 1: For any feasible resource allocation decision that can be implemented during time frame t, we have Eq. (13), as shown at the top of the next page, where B is an upper bound on the term $\frac{1}{2}[(\rho\vartheta g_n e_n^t)^2 + (P_n^{Total,t})^2 + (\rho\mathcal{R}^p e_n^t)^2 + (\rho\mathcal{R}^g)^2 + (\sum_{\tau=0}^{\rho-1} r_n^\tau)^2 + (f_n^t)^2]$.

Proof: See Appendix A ■

Our dynamic resource allocation policy is designed to observe the data queue state information (QSI) $\mathbf{Q}^t = \{Q_n^t|n \in \mathcal{N}\}$, virtual token queue $\mathbf{M}^t = \{M_n^t, n \in \mathcal{N}\}$ and energy queue $\mathbf{E}^t = \{E_n^t, n \in \mathcal{N}\}$, and as well to make the system policies $\{\mathbf{F}^t, \mathbf{e}^t, \mathbf{p}^t\}$ for minimizing the right-hand-side in Eq. (13) for current time. The non-constant part of the right-hand-side in Eq. (13) can be rewritten as Eq. (14), as

$$\Delta_V(\Theta^t \leq B - V\,\mathbb{E}[U^t(\mathbf{F}^t)|\Theta^t] + Q_n^t(f_n^t - \sum_{\tau=0}^{\rho-1} r_n^\tau) - \tilde{M}_n^t(\mathcal{R}^g \sum_{\tau=0}^{\rho-1} r_n^\tau - \rho\mathcal{R}^p e_n^t) - \tilde{E}_n^t(\rho\vartheta g_n e_n^t - P_n^{Total,t}). \tag{13}$$

$$\begin{aligned} &V\mathbb{E}[U^t(\mathbf{F}^t)|\Theta^t] - Q_n^t(f_n^t - \sum_{\tau=0}^{\rho-1} r_n^\tau) + \tilde{M}_n^t(\mathcal{R}^g \sum_{\tau=0}^{\rho-1} r_n^\tau - \rho\mathcal{R}^p e_n^t) + \tilde{E}_n^t(\rho\vartheta g_n e_n^t - P_n^{Total,t}) \\ &= \left\{V\sum_n \varpi_n \log(f_n^t + e) - \sum_n Q_n^t f_n^t + \sum_n \vartheta\rho g_n \tilde{E}_n^t e_n^t - \rho\mathcal{R}^p \sum_n \tilde{M}_n^t e_n^t\right\} + \left\{\sum_n \sum_{\tau=0}^{\rho-1} Q_n^t r_n^\tau + \sum_n \sum_{\tau=0}^{\rho-1} \tilde{M}_n^t \mathcal{R}^g r_n^\tau - \sum_n \sum_{\tau=0}^{\rho-1} \tilde{E}_n^t p_n^\tau\right\}. \end{aligned} \tag{14}$$

shown at the top of the page. Then, Eq. (10) is translated into a series of optimization problems over the time frames. Since the functions in Eq. (14) and constraints in original problem (Eq. (10)) for the system policies are independent with each other, the converted problem are divided into two subproblems based on different timescale network states, which are joint RA and ET problem (at each time frame) and TPC problem (at each timeslot).

A. JOINT RA AND ET PROBLEM FOR EACH TIME FRAME

From Eq. (14), we obtain the joint RA and ET problem for time frame t as

$$\begin{aligned} \max_{\mathbf{F}^t, \mathbf{e}^t} & \sum_n \left(V\varpi_n \log(f_n^t + e) - Q_n^t f_n^t\right) \\ & + \rho \sum_n \left(\vartheta g_n \tilde{E}_n^t e_n^t - \mathcal{R}^p \tilde{M}_n^t e_n^t\right) \\ \text{s.t. } & \text{(C1), (C3), (C6) and (C7).} \end{aligned} \tag{15}$$

Since there also exist no coupling between RA and ET for their functions and constraints, we further separate the problem in Eq. (15), and solve RA for BS and ET for each D2D-Tx at time frame t independently.

1) RA FOR BS

From Eq. (15), the RA of the BS at time frame t is

$$\begin{aligned} \max_{\mathbf{F}^t} & \sum_n V\varpi_n \log(f_n^t + e) - Q_n^t f_n^t] \\ \text{s.t. } & \text{(C1).} \end{aligned} \tag{16}$$

Eq. (16) is a strictly concave function, which can be solved efficiently by the gradient descent method [34]. Then, the rate f_n^t allocated to D2D-Tx n^R can be calculated by

$$f_n^t = \left[\frac{V\varpi_n}{Q_n^t + \eta^t} - e\right]^{[0, C^t]}, \tag{17}$$

where $[x]^{[0,C^t]}$ denotes the projection of x onto $[0, C^t]$, and η^t is the optimal Lagrange multiplier.

2) ET DECISIONS FOR D2D-Txs

To obtain the optimal trading energy e_n^t for D2D-Tx n^T, we have

$$\begin{aligned} \max_{\mathbf{e}^t} & \sum_n \rho(\vartheta g_n \tilde{E}_n^t e_n^t - \mathcal{R}^p \tilde{M}_n^t e_n^t) \\ \text{s.t. } & \text{(C3), (C6) and (C7).} \end{aligned} \tag{18}$$

Since Eq. (18) is a linear equation, if $\vartheta g_n \tilde{E}_n^t \geq \mathcal{R}^p \tilde{M}_n^t$, the D2D-Tx n^T shall buy energy from the PS as much as possible to recharge its battery for optimizing Eq. (18). Otherwise, the D2D-Tx n^T's residual energy is enough and it does not need to buy energy at the cost of its holding tokens. We have

$$e_n^t = \begin{cases} \min[o_n, \dfrac{\mathcal{E}_n^{\max} - E_n^t}{\vartheta\rho g_n}, \dfrac{M_n(t)}{\rho\mathcal{R}^p}] & \text{if } \vartheta g_n \tilde{E}_n^t \geq \mathcal{R}^p \tilde{M}_n^t, \\ 0 & \text{otherwise.} \end{cases} \tag{19}$$

B. TPC PROBLEM OF D2D-Txs FOR EACH TIMESLOT

The adaptive TPC problem can be described as the second term in Eq. (14). Considering the power constrains, i.e., (C2), (C5) and (C8) in Eq. (10), there is no coupling in the term $\sum_{n\in\mathcal{N}}$. Then, with $r_n^\tau = W\log_2(1 + \zeta\frac{h_{nn}^\tau p_n^\tau}{\sigma^2 + \sum_{m\in\mathcal{N}\setminus\{n\}} h_{mn}^\tau p_m^\tau})$, the TPC problem for D2D-Tx n^T can be rewritten as

$$\begin{aligned} \max_{\mathbf{p}^t} & \begin{cases} (Q_n^t + \tilde{M}_n^t \mathcal{R}^g)W \\ \quad \log_2(1 + \zeta\dfrac{h_{nn}^\tau p_n^\tau}{\sigma^2 + \sum_{m\in\mathcal{N}\setminus\{n\}} h_{mn}^\tau p_m^\tau}) \\ -\tilde{E}_n^t p^\tau \end{cases} \\ \text{s.t. } & 0 \leq p_n^\tau \leq \mathcal{P}_n, \quad \forall 0 \leq \tau < \rho - 1, \end{aligned} \tag{20}$$

where $\log_2$ is the binary logarithm. The value $\mathcal{P}_n$ is the maximum transmit power that D2D-Tx n^T can take, which is obtained based on the power constrains in Eq. (10).

Remark 1: Based on the system model, we know that D2D-Tx n^T obtains $W\mathcal{R}^g \log_2(1 + \zeta\frac{h_{nn}^\tau p_n^\tau}{\sigma^2 + \sum_{m\in\mathcal{N}\setminus\{n\}} h_{mn}^\tau p_m^\tau})$ tokens and spends p_n^τ energy when providing forwarding service to UT n^T. In the practical system, the D2D-Tx n^T also focuses *on its long-term revenue, and as well it may consider its resource states except for its current reward and cost when taking the power strategy to UT n^R at timeslot τ. It is noted that, the less $\tilde{M}^t_{n,m}$ for D2D-Tx n^T means the more tokens ($\tilde{M}^t_n = \mathcal{M}^{max}_n - M^t_n$). At this time, the tokens that the D2D-Tx earns by providing forwarding service to UT n^R are less valuable. Moreover, the larger $\tilde{E}^t_{n,m}$ for D2D-Tx n^T means the less residual energy, then the energy resource that the D2D-Tx consumes for forwarding contents to n^R is expensive. Additionally, the QSI contributes the extra utility to D2D-Tx to make sure the normal operation of the network.*

Eq. (20) is a non-convex objective function due to the interference among D2D-Txs. In the following, we will approximate the non-convex objective function in the exponential domain of power. The resultant approximation problem facilitates to devise a fast algorithm that provably achieves at least local optimality.

After implementing a variable change $\hat{p}^\tau_n = log(p^\tau_n)$, Eq. (20) can be rewritten as

$$\max_{\hat{\mathbf{p}}^t} \begin{cases} (Q^t_n + \tilde{M}^t_n \mathcal{R}^g) \log_2(\sigma^2 + \Psi(\mathbf{p}^\tau_{-n}) + \zeta h^\tau_{nn} \exp^{\hat{p}^\tau_n}) \\ -(Q^t_n + \tilde{M}^t_n \mathcal{R}^g) \log_2(\sigma^2 + \Psi(\mathbf{p}^\tau_{-n})) - \tilde{E}^t_n \exp^{\hat{p}^\tau_n} \end{cases}$$
$$\text{s.t. } 0 \le \hat{p}^\tau_n \le log(\mathcal{P}_n), \quad \forall 0 \le \tau < \rho - 1, \tag{21}$$

where $\Psi(\mathbf{p}^\tau_{-n}) = \sum_{m \in \mathcal{N} \setminus \{n\}} h^\tau_{mn} p^\tau_m$. In Eq. (21), both $(Q^t_n + \tilde{M}^t_n \mathcal{R}^g) \log_2(\sigma^2 + \Psi(\mathbf{p}^\tau_{-n}))$ and $\tilde{E}^t_n \exp^{\hat{p}^\tau_n}$ are convex functions. If the first term $(Q^t_n + \tilde{M}^t_n \mathcal{R}^g) \log_2(\sigma^2 + \Psi(\mathbf{p}^\tau_{-n}) + \zeta h^\tau_{nn} \exp^{\hat{p}^\tau_n})$ can be approximated by some linear functions, Eq. (21) can be reformulated as a convex problem. Motivated by this consideration, we make some linear approximations based on the successive approximation method [35]. Specifically, if function $f(x)$ is non-concave for variable x, we can employ the first-order Taylor expansion at the fixed point x_p to approximate $f(x)$, i.e.,

$$f(x) \approx f(x_p) + (\frac{\partial f(x)}{\partial x})_{|x=x_p}(x - x_p). \tag{22}$$

In this paper, by setting the point x_p as 0, we can obtain the following approximation,

$$\log_2(\sigma^2 + \Psi(\mathbf{p}^\tau_{-n}) + \zeta h^\tau_{nn} \exp^{\hat{p}^\tau_n})$$
$$\approx \underbrace{\log_2(\sigma^2 + \Psi(\mathbf{p}^\tau_{-n}) + \zeta h^\tau_{nn})} + \underbrace{\frac{\zeta h^\tau_{nn}}{\sigma^2 + \Psi(\mathbf{p}^\tau_{-n}) + \zeta h^\tau_{nn}} \hat{p}^\tau_n}.$$

Then, Eq. (21) can be reformulated as Eq. (23), as shown at the bottom of the next page. Since its objective function is concave function and its constraint is convex set, thus Eq. (23) is a strictly concave function in $\hat{p}^\tau$. Then, based on gradient descent method, the optimal power control for D2D-Tx n^T at timeslot τ is

$$p^\tau_n = \left[\frac{(Q^t_n + \tilde{M}^t_n \mathcal{R}^g) \zeta h^\tau_{nn}}{(\tilde{E}^t_n + \mu^\tau_n)(\sigma^2 + \sum_{m \in \mathcal{N} \setminus \{n\}} h^\tau_{mn} p^\tau_m + \zeta h^\tau_{nn})} \right]^{[0,\mathcal{P}_n]}, \tag{24}$$

where μ^τ_n is the optimal Lagrange multiplier.

Algorithm 1 OTTRA

Initialization: Iteration index $k = 1$, $i = 1$.
At each time frame t, BS observes D2D-Txs' data queue states $\mathbf{Q}^t$, energy states $\mathbf{E}^t$ and token states $\mathbf{M}^t$;
for $n = 1 : N$ **do**
  $f^t_n(k) = \left[\frac{V \varpi_n}{Q^t_n + \eta^t(k)} - e \right]^{[0, C^t]}$;
  if $\sum_N \parallel f^t_n(k) - f^t_n(k-1) \parallel \le \varrho$ **then**
    Updates $\eta^t(k)$; $k = k + 1$;
  else
    break;
  end if
end for
for $n = 1 : N$ **do**
  Calculates ET decision e^t_n based on Eq. (19);
end for
for $\tau = 0 : (\rho - 1)$ within time frame t **do**
  for $n = 1 : N$ **do**
    D2D-Tx n observes channel state and other D2D-Txs' transmit power strategies $\tilde{\mathbf{p}}^\tau_{-n} = (\tilde{p}^\tau_1, \cdots, \tilde{p}^\tau_{n-1}, \tilde{p}^\tau_{n+1}, \cdots, \tilde{p}^\tau_N)$;
    $p^\tau_n(i) = \left[\frac{(Q^t_n + \tilde{M}^t_n \mathcal{R}^g) h^\tau_{nn}}{(\tilde{E}^t_n + \mu^\tau_n(i))(\sigma^2 + \Psi(\tilde{\mathbf{p}}^\tau_{-n}) + \zeta h^\tau_{nn})} \right]^{[0, \mathcal{P}_n]}$;
    if $\sum_N \parallel p^\tau_n(i) - p^\tau_n(i-1) \parallel \le \phi$ **then**
      Updates $\mu^\tau_n(i)$; $i = i + 1$;
    else
      break;
    end if
  end for
end for
D2D-Txs update $\mathbf{Q}^t$, $\mathbf{E}^t$ and $\mathbf{M}^t$.

Based on the analysis above, we propose an online two-timescale resource allocation (OTTRA) algorithm for joint RA of BS, ET and TPCs of D2D-Txs in terms of different timescale network states, as shown in **Algorithm 1**. Noting that TPC for D2D-Tx n^T at timeslot τ (Eq. 24) is related to other D2D-Txs' transmit power. In **Algorithm 1**, we employ the block coordinate descent (BCD) method [36] to obtain the optimal transmit power for each D2D-Tx. Specifically, at each iteration for timeslot τ, each D2D-Tx observes others' transmit power. Then, the single block of power variables for D2D-Txs is optimized orderly while the remaining blocks are held fixed. Since Eq. (18) is a linear equation in e^t_n, Eq. (23) and Eq. (16) are strongly convex in $\hat{p}^\tau_n$ and f^t_n respectively, **Algorithm 1** can converge to the global optimum for two subproblems (joint RA and ET problem and TPC problem). Then, the network objective can be achieved and D2D communication pairs can receive desirable contents in the long term. Additionally, by referring to [37], we know that the iteration complexities of RA and TPC in **Algorithm 1** are $O(\frac{1}{\varrho^2})$ and $O(\frac{1}{\phi^2})$, when setting their tolerance deviations as ϱ and ϕ respectively. Similarly, ET's iteration complexity is $O(1)$. Since there are two loops (two timescales) executed nested

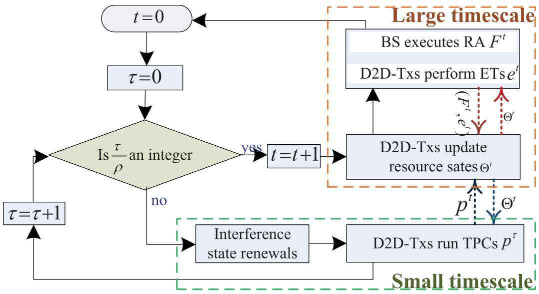


FIGURE 3. Flowchart of resource allocation scheme.

for **Algorithm 1**, the iteration complexity of **Algorithm 1**[5] is $O(\frac{1}{\varrho^2\phi^2})$.

V. PERFORMANCE ANALYSIS FOR THE SCHEME

Noting that, there may exist conflicting interests between the BS and the self-interested nodes for the cooperative content transmission over wireless powered D2D networks. Moreover, in the practical communication scenario, the node's source states may change much slower than channel fading. To deal with the self-interested nodes, we first propose an incentive mechanism to encourage the cooperative content forwarding for D2D-Txs and the ET between PS and D2D-Txs. Then, for different timescale network states, we propose a two-timescale resource allocation scheme to maximize the average network utility. Fig. 3 summarizes the overall solution and the interrelationship of the resource allocation for our two-timescale scheme. From Fig. 3, we observe that the decisions of RA and ET are made by the BS and D2D-Txs at each large time frame, while TPC strategy is taken by each D2D-Tx at each timeslot for maximizing its own profit. Specifically, at the beginning of each large time frame, the BS first executes RA to allocate the content flow to D2D-Txs based on D2D-Txs' data queue state, such that the BS can push the contents requested to network as much as possible. Meanwhile, each D2D-Tx performs ET with PS according to its holding tokens and residual energy. Then, for τ timeslot within a specific time frame, each D2D-Tx will buy energy from PS using tokens at WET stage, and run TPC to forward the contents to UT based on the interference state among D2D-Txs at WIT stage. Noting that BS's RA, D2D-Txs' ET and TPCs affect each other over time via network states. Then, by using OTTRA algorithm, both BS and D2D-Txs can achieve their objectives in an orderly manner based on different timescale network states.

A. COMMUNICATION OVERHEAD OF TWO-TIMESCALE SCHEME

In the two-timescale resource allocation scheme, only at each time frame (large timescale), the RA, ET and information collection[6] need to be performed. The proposed scheme benefits from low communication overhead which includes the signalling overhead and computational complexity. Motivated by [38], we elaborate the communication overhead for our two-timescale scheme at each time frame. For simplicity, each data packet is assumed to be quantized by ξ bits.

- For RA problem at each time frame, BS needs to observe data queue states $\mathbf{Q}^t$ of D2D-Txs. Then, the total number of observed bits is $N\xi$, and hence, the total signalling overhead is $N\xi$. Meanwhile, the BS has to calculate the data rate for each D2D-Tx. By using primal-dual interior point method, its computational complexity for each time frame is $O(N^3)$ [39].
- For ET problem at each time frame, each D2D-Tx decides the amount of the required ET based on its local information and sends the decision to the PS. Then, the signalling overhead of ET problem is $N\xi$. Besides, each D2D-Tx needs to calculate the weight of utility function (Eq. (18) in Section IV) and make the decision. Then, the computational complexity of ET problem is N.
- For TPC problem at each timeslot, apart from its local resource states, the D2D-Tx has to observe the channel states of system and others' power strategies. Then, the signalling overhead of TPC problems for each time frame is $(2N-1)\xi\rho$. Since the TPC problem can be approximated as a convex problem, the computational complexity for each time frame is $O(N^3\rho)$.

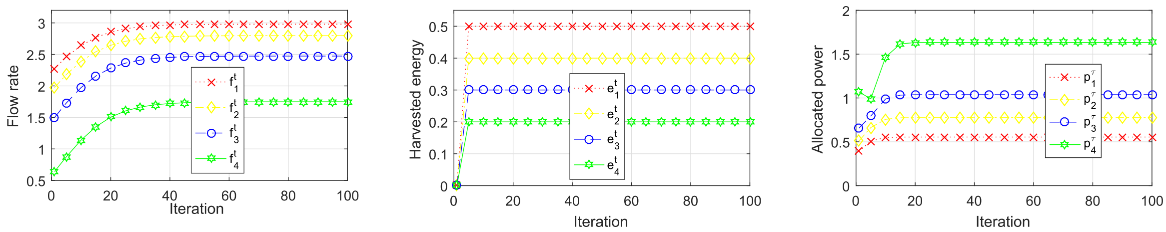
From above, we know that the signalling overhead and computational complexity of the overall two-timescale scheme at each time frame are $O((2N+(2N-1)\rho)\xi)$ and $O(N^3+N+N^3\rho)$ respectively, which are at most polynomials in terms of the system parameters. Compared with the one-timescale resource allocation scheme which operates system policies (RAs, ETs and TPCs) and information collection at each short timescale,[7] the communication overhead imposed in our scheme is trivial. Accordingly, the proposed two-timescale scheme is cost-effective and easy-to-realized, which are well scalable with the size of the problem.

[5] The algorithm complexity is defined as the maximum number of algorithm iterations before agent obtains the optimal strategy at a specific system state.

[6] BS needs to collect D2D-Txs' resource states for RA at large timescale.

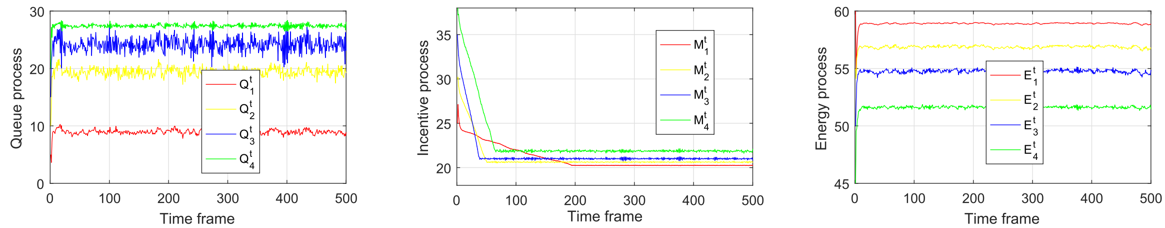
[7] The signaling overhead and computational complexity in one-timescale scheme are $O((4N-1)\xi\rho)$ and $O((N^3+N+N^3)\rho)$, respectively.

$$\max_{\hat{\mathbf{p}}^t} \begin{cases} (Q_n^t+\tilde{M}_n^t\mathcal{R}^g)\left[\log_2(\sigma^2+\Psi(\mathbf{p}_{-n}^\tau)+\zeta h_{nn}^\tau)+\dfrac{h_{nn}^\tau}{\sigma^2+\Psi(\mathbf{p}_{-n}^\tau)+\zeta h_{nn}^\tau}\hat{p}_n^\tau\right] \\ -(Q_n^t+\tilde{M}_n^t\mathcal{R}^g)\log_2(\sigma^2+\Psi(\mathbf{p}_{-n}^\tau))-\tilde{E}_n^t\exp^{\hat{p}_n^\tau} \end{cases}$$
$$\text{s.t. } 0\le\hat{p}_n^\tau\le log(\mathcal{P}_n),\quad \forall 0\le\tau<\rho-1. \tag{23}$$



(a) (b) (c)

FIGURE 4. Dynamic resource allocation process for contents transmission. (a) RA for BS. (b) ET for D2D-Txs. (c) TPCs for D2D-Txs.



(a) (b) (c)

FIGURE 5. Dynamic queue processes across the time frame. (a) Date queue process. (b) Virtual token queue process. (c) Virtual energy queue process.

B. NETWORK PERFORMANCE OF OTTRA ALGORITHM

Now, we will analyze the network performance of our OTTRA algorithm for the two-timescale scheme.

Theorem 2: For implementing the algorithm with any fixed parameter $V > 0$ for all time frames, we have the following performance guarantees:

a) The queue upper bounds are given as follows,

$$\mathcal{Q}_n^{max} = V\varpi_n \nu_U + C_n, \quad n \in N; \tag{25}$$

$$\mathcal{M}_n^{max} = \frac{V\varpi_n \nu_U + C_n}{\mathcal{R}^p - \mathcal{R}^g} + \mathcal{M}_t^0, \quad n \in N; \tag{26}$$

$$\mathcal{E}_n^{max} = \frac{\vartheta g_n(V\varpi_n \nu_U + C_n)}{\mathcal{R}^p(\mathcal{R}^p - \mathcal{R}^g)} + \mathcal{E}_t^0, \quad n \in N; \tag{27}$$

b) The objective function value of the problem achieved by the proposed algorithm satisfies the bound

$$\overline{U} \geq O^* - \frac{B + \Upsilon}{V}, \tag{28}$$

where Υ is a positive parameter, C_n is the maximal rate of the link between BS and D2D-Tx n^T, O^* is the optimal value of the built optimization problem.

Proof: Please see Appendix B.

Theorem 2 gives the upper bounds of both queues and average network utility, which is related to control parameter V. Specifically, for any $V > 0$, the proposed algorithm can achieve a time-average utility that is within $O(1/V)$ shown in Eq. (28), while ensuring that the average data queues, available energy resource and tokens have upper bounds of $O(V)$ shown in Eqs. (25)-(27). Since the average transmission delay is proportional to the average length [40], we can depict the average delay by the average length. Then, based on **Theorem 2**, we know that parameter V enables an explicit trade-off between the average objective value and network delay. Therefore, we can obtain a significant rule for engineering design to flexibly balance the network utility-delay performance.

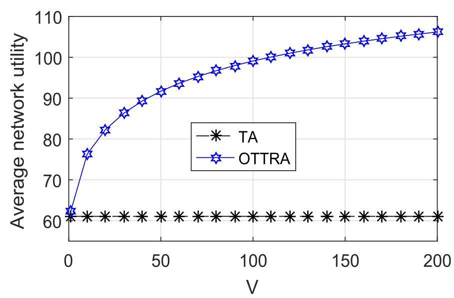


FIGURE 6. Average network utility vs control parameter V.

VI. SIMULATION RESULTS

In this section, we present simulation results to illustrate the performance for our scheme. We present the dynamic processes of the resource allocation, and show the characteristics of network utility and queue lengths of our scheme.

A. PARAMETERS SETTING

There are four D2D communication pairs in the simulation system. For simplicity of simulations, we consider a normalized bandwidth spacing, i.e. $W = 1$. We model the channel process as Gaussian random variables which are i.i.d over different timeslots. Then, setting the following parameters: maximum transmit power $P^{\max} = 3$, spent tokens $\mathcal{R}^p = 0.42$ and earned tokens $\mathcal{R}^g = 0.1$. Moreover, the maximum rate for

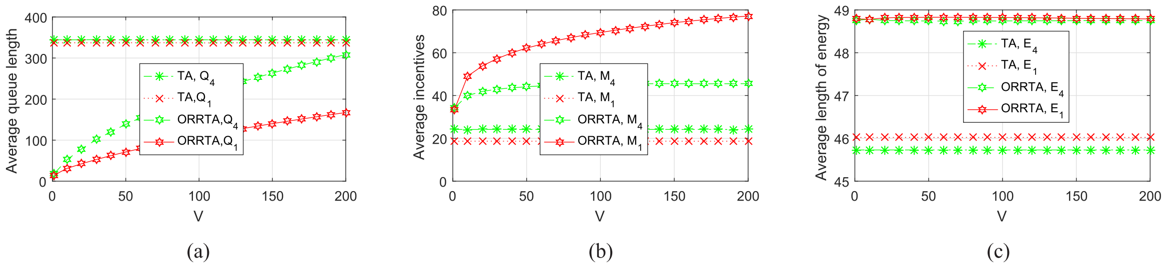


FIGURE 7. Average queue lengths vs control parameter V. (a) Length of average data queue. (b) Length of average virtual token queue. (c) Length of average virtual energy queue.

link between BS and D2D-Tx n^T, $C^t=20$, $\forall t \geq 0$. We also assume that there are 30 timeslots during each time frame. The resource states (data queue, residual energy and holding tokens) of D2D-Txs keep content within one time frame and change on the frame boundaries (the large timescale). On the contrary, the channel state on the link of D2D communication pair keeps constant within one timeslot and changes on the time boundaries (the small timescale).

B. DYNAMIC PROCESSES OF THE PROPOSED SCHEME

In this subsection, we exhibit the performance of the proposed algorithm over different timescale network states. Fig. 4 depicts the resource allocation process for OTTRA algorithm. Fig. 4 (a) exhibits the dynamics of flow rate for BS and Fig. 4 (b) plots the D2D-Txs' ET decisions for a specific time frame, while Fig. 4 (c) plots the updates of D2D-Txs' transmit power strategies for one timeslot. Fig. 4 indicates the convergence of our algorithm, and also shows the influences of queues $\mathbf{Q}^t$, $\mathbf{M}^t$ and $\mathbf{E}^t$ on the resource allocation decisions. Fig. 5 plots the queue processes for OTTRA algorithm over the time frames. Figs. 5 (a), 5 (b) and 5 (c) exhibit the dynamic data queue process, holding token process and residual energy process for D2D-Txs, respectively. In Fig. 5, we see that the lengths of queues will tend to be stable and stay around the positive values, which verifies to be in the boundaries of the queues proved in **Theorem 2**.

C. AVERAGE QUEUE LENGTH AND NETWORK THROUGHPUT FOR DIFFERENT APPROACHES

To verify the effectiveness of our scheme, we compare our algorithm with a traditional approach (TA) which executes resource allocation to D2D-Txs equally. Fig. 6 depicts that the average network utility versus the control parameter V for different approaches. We notice that the average network utility in our scheme increases with V and that the average utility achieved by our scheme is larger than that of TA. Fig. 7 shows the time-average lengthes of queues versus the control parameter V. Specifically, Figs. 7 (a), 7 (b) and 7 (c) plot the average lengths of data queue, holding tokens and energy resource for D2D-Txs 1^T and 4^T. Since the network delay can be depicted by the average data queue length. Fig. 7 (a) also illustrates that the network delay for our scheme is much smaller than that of TA scheme. Moreover, from Fig. 6 and Fig. 7 (a), we can see that our scheme obtains a tradeoff between average network utility and delay, which proves the conclusion of **Theorem 2**. In Figs. 7 (b) and 7 (c), the values of D2D-Tx's holding tokens and residual energy are larger than those of TA scheme. It is shown that, our scheme performs better in storing more energy resource and improving the positivity of the self-interested nodes for the content transmission participation.

VII. CONCLUSIONS

In this paper, we studied the dynamic resource allocation for content transmission in wireless powered D2D communication with self-interested D2D-Txs and PS. Firstly, a stochastic optimization was built to maximize the average network utility under the constraints of the limited data buffer, energy capacity and tokens. Then, the problem of stochastic optimization is converted into two subproblems operating at two timescales via Lyapunov optimization technique in terms of different timescale network states. We obtained the solutions of a joint RA and ET problem at the large timescale and a non-convex TPC problem at the small timescale. Finally, an OTTRA algorithm was proposed which introduced small communication overhead, and its performance bounds was obtained for the utility-delay tradeoff.

APPENDIX A
PROOF OF THEOREM 1

Followed from Eq. (12), we have

$$\frac{1}{2}[(-\tilde{E}_n^{t+1})^2 - (-\tilde{E}_n^t)^2]$$
$$\leq \frac{1}{2}[(\rho\vartheta g_n e_n^t)^2 + (P_n^{Total,t})^2] + \tilde{E}_n^t(\rho\vartheta g_n e_n^t - P_n^{Total,t});$$
$$\frac{1}{2}[(-\tilde{M}_n^{t+1})^2 - (-\tilde{M}_n^t)^2]$$
$$\leq \frac{1}{2}[(\rho\mathcal{R}^p e_n^t)^2 + (\rho\mathcal{R}^g)^2] + \tilde{M}_n^t(\mathcal{R}^g \sum_{\tau=0}^{\rho-1} r_n^\tau - \mathcal{R}^p e_n^t);$$
$$\frac{1}{2}[(Q_n^{t+1})^2 - (Q_n^t)^2]$$
$$\leq \frac{1}{2}[(\sum_{\tau=0}^{\rho-1} r_n^\tau)^2 + (f_n^t)^2] + Q_n^t(\sum_{\tau=0}^{\rho-1} r_n^\tau - f_n^t).$$

Then, we have Eq. (29), as shown at the top of the next page, where B is an upper bound on the term $\frac{1}{2}[(\rho\vartheta g_n e_n^t)^2 + (P_n^{Total,t})^2 + (\rho\mathcal{R}^p e_n^t)^2 + (\rho\mathcal{R}^g)^2 + (\sum_{\tau=0}^{\rho-1} r_n^\tau)^2 + (f_n^t)^2]$, which holds under the fact that the resource allocation variables

$$L(\Theta^{t+1}) - L(\Theta^t) \leq \frac{1}{2}[(\rho\vartheta g_n e_n^t)^2 + (P_n^{Total,t})^2 + (\rho\mathcal{R}^p e_n^t)^2 + (\rho\mathcal{R}^g)^2 + (\sum_{\tau=0}^{\rho-1} r_n^\tau)^2 + (f_n^t)^2]$$
$$+ Q_n^t(f_n^t - \sum_{\tau=0}^{\rho-1} r_n^\tau) - \tilde{M}_n^t(\mathcal{R}^g \sum_{\tau=0}^{\rho-1} r_n^\tau - \rho\mathcal{R}^p e_n^t) + \tilde{E}_n^t(\rho\vartheta g_n e_n^t - P_n^{Total,t})$$
$$\leq B + Q_n^t(f_n^t - \sum_{\tau=0}^{\rho-1} r_n^\tau) - \tilde{M}_n^t(\mathcal{R}^g \sum_{\tau=0}^{\rho-1} r_n^\tau - \rho\mathcal{R}^p e_n^t) - \tilde{E}_n^t(\rho\vartheta g_n e_n^t - P_n^{Total,t}). \quad (29)$$

satisfy the properties of bounds, i.e. $0 \leq \sum_n f_n^t \leq C^t$, $0 \leq p_n^\tau \leq P_n^{max}$ and $0 \leq e_n^t \leq o_n$. Adding $-V\mathbb{E}[U^t(\mathbf{F}^t, \mathbf{h}^t)|\Theta^t]$ to both sides of Eq. (29). This completes the proof of **Theorem 2**.

APPENDIX B
PROOF OF THEOREM 2

a) Since Eq. (25) holds at time frame $t = 0$, we show that if Eq. (25) holds at time frame t, i.e., $Q_n^t < V\varpi_n\nu_U + C_n$, then it also holds at time frame $t+1$. If $Q_n^t \leq V\varpi_n\nu_U$, then we have $Q_n^t < V\varpi_n\nu_U + C_n$, where C_n denotes the maximal rate on link between BS and D2D-Txs n^T and ν_U is the maximal network utility for the time frame. For the case of $V\nu_U < Q_n^t \leq V\varpi_n\nu_U$, we show that $f_n^t = 0, \forall n \in \mathcal{N}$, is the optimal solution to RA problem. Then, $Q_n^{t+1} \leq Q_n^t \leq V\varpi_n\nu_U$. Thus, Eq. (25) is proved. Similarly, if $\vartheta g_n \tilde{E}_n^t - \mathcal{R}^p \tilde{M}_n^t \leq 0$, and $(Q_n^t + \tilde{M}_n^t\mathcal{R}^g) - \tilde{E}_n^t \leq 0$, the optimal $e_n^t = 0$ and $\sum_\tau p_n^\tau = 0$. Based on the Equation above, we obtain that $\tilde{M}_n^t \geq \frac{Q_n^t}{\mathcal{R}^p - \mathcal{R}^g}$. That is, $\mathcal{M}_{\max} > \frac{Q_n^t}{\mathcal{R}^p - \mathcal{R}^g} + \mathcal{M}_t^n$. Since $Q_n^t \leq V\varpi_n\nu_U + C_n$ and $\mathcal{M}_t^n \leq \mathcal{M}_0^n$, we set $\mathcal{M}_{\max}$ as

$$\mathcal{M}_{\max} = \frac{V\varpi_n\nu_U + C_n}{\mathcal{R}^p - \mathcal{R}^g} + \mathcal{M}_t^0.$$

Then, for the $\mathcal{E}_{\max}$,

$$\mathcal{E}_{\max} = \frac{\vartheta g_n(V\varpi_n\nu_U + C_n)}{\mathcal{R}^p(\mathcal{R}^p - \mathcal{R}^g)} + \mathcal{E}_t^0.$$

b) Suppose that there exists a $\mathbf{h}$-only policy that satisfies $\tilde{U}^t(\tilde{\mathbf{F}}^t) = O^*$, $\mathbb{E}[\sum_{\tau=0}^{\rho-1} \tilde{r}_n^\tau - \tilde{f}_n^t] \geq \varepsilon$, $\mathbb{E}[\sum_{\tau=0}^{\rho-1} \tilde{p}_n^\tau - \vartheta\rho\tilde{e}_n^t g_n] \geq \sigma$ and $\mathbb{E}[\rho\mathcal{R}^p\tilde{e}_n^t - \mathcal{R}^g \sum_{\tau=0}^{\rho-1} \tilde{r}_n^\tau] \geq \iota$, where $\tilde{f}_n^t, \tilde{e}_n^t$, $\tilde{r}_n^\tau$ and $\tilde{p}_n^\tau$ are the resulting values under $\mathbf{h}$-only policy. Then, we have $\Delta(\Theta)^t - V\mathbb{E}[U^t(\mathbf{F}^t)|\Theta^t] \leq B + \Upsilon - V\tilde{U}^t(\tilde{\mathbf{F}}^t) - \sum_n(\varepsilon Q_n^t + \sigma E_n^t + \iota W_n^t) \leq B + \Upsilon - VO^*$, where Υ is a positive parameter. After having summed over $t \in \{0, 1, \cdots, \varsigma - 1\}$ for above inequality, we get $\mathbb{E}[L(\Theta^{\varsigma-1})|\Theta^t] - \mathbb{E}[L(\Theta^0)] - V\sum_{t=0}^{\varsigma-1} \mathbb{E}[U^t(\mathbf{F}^t)|\Theta^t] \leq (B+\Upsilon)\varsigma - V\varsigma O^*$. Then, following from the general derivation, we get

$$-\frac{\mathbb{E}[L(\Theta^0)]}{\varsigma V} + O^* \leq \frac{B+\Upsilon}{V} + \frac{1}{\varsigma}\sum_{t=0}^{\varsigma-1} \mathbb{E}[U^t|\Theta^t].$$

Taking limit as $\varsigma \to \infty$, we comes to the solution, $\overline{U} \geq O^* - \frac{B+\Upsilon}{V}$.

REFERENCES

[1] L. Song, D. Niyato, Z. Han, and E. Hossain, "Game-theoretic resource allocation methods for device-to-device communication," *IEEE Wireless Commun.*, vol. 21, no. 3, pp. 136–144, Jun. 2014.

[2] C. Yi, S. Huang, and J. Cai, "An incentive mechanism integrating joint power, channel and link management for social-aware D2D content sharing and proactive caching," *IEEE Trans. Mobile Comput.*, vol. 17, no. 4, pp. 789–802, Apr. 2018.

[3] D. Wu, L. Zhou, and Y. Cai, "Social-aware rate based content sharing mode selection for D2D content sharing scenarios," *IEEE Trans. Multimedia*, vol. 19, no. 11, pp. 2571–2582, Nov. 2017.

[4] M. N. Soorki, M. H. Manshaei, B. Maham, and H. Saidi, "On uplink virtual MIMO with device relaying cooperation enforcement in 5G networks," *IEEE Trans. Mobile Comput.*, vol. 17, no. 1, pp. 155–168, Jan. 2017.

[5] Z. Yang, H. Tian, S. Fan, and G. Chen, "Dynamic incentive design in content dissemination process through D2D communication," *IEEE Commun. Lett.*, vol. 21, no. 8, pp. 1799–1802, Aug. 2017.

[6] Z. Ye, Q. Cui, N. Hu, and F. Yang, "A dynamic bandwidth and power allocation scheme for cooperative D2D communications," in *Proc. IEEE 28th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Oct. 2017, pp. 1–6.

[7] C. Xu, J. Feng, Z. Zhou, Z. Chang, Z. Han, and S. Mumtaz, "Two-stage matching for energy-efficient resource management in D2D cooperative relay communications," in *Proc. IEEE Global Commun. Conf.*, Dec. 2017, pp. 1–6.

[8] Z. Hou, H. Chen, Y. Li, Z. Han, and B. Vucetic, "A contract-based incentive mechanism for energy harvesting-based Internet of Things," in *Proc. IEEE Int. Conf. Commun.*, May 2017, pp. 1–6.

[9] J. Gong, Z. Zhou, and S. Zhou, "Analysis and optimization of wireless transmissions over fast fading channels with slow time-varying energy arrival," in *Proc. IEEE Int. Conf. Commun.*, May 2017, pp. 1–6.

[10] Z. Zhou, G. Ma, M. Dong, K. Ota, C. Xu, and Y. Jia, "Iterative energy-efficient stable matching approach for context-aware resource allocation in D2D communications," *IEEE Access*, vol. 4, pp. 6181–6196, Jul. 2016.

[11] Z. Zhou, K. Ota, M. Dong, and C. Xu, "Energy-efficient matching for resource allocation in D2D enabled cellular networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 6, pp. 5256–5268, Jun. 2017.

[12] A. Abrardo and M. Moretti, "Distributed power allocation for D2D communications underlaying/overlaying OFDMA cellular networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1466–1479, Mar. 2017.

[13] Y. Jiang, Q. Liu, F. Zheng, X. Gao, and X. You, "Energy-efficient joint resource allocation and power control for D2D communications," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6119–6127, Aug. 2016.

[14] Y. Liu, Y. Zhang, R. Yu, and S. Xie, "Integrated energy and spectrum harvesting for 5G wireless communications," *IEEE Netw.*, vol. 29, no. 3, pp. 75–81, May/Jun. 2015.

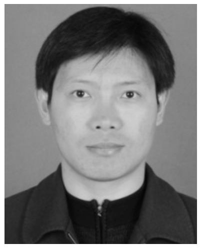
[15] U. Saleem, H. K. Qureshi, S. Jangsher, and M. Saleem, "Transmission power management for throughput maximization in harvesting enabled D2D network," in *Proc. IEEE Symp. Comput. Commun.*, Jun. 2016, pp. 1078–1083.

[16] M.-L. Ku and J.-W. Lai, "Joint beamforming and resource allocation for wireless-powered device-to-device communications in cellular networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 11, pp. 7290–7304, Nov. 2017.

[17] H. Wang, J. Wang, G. Ding, L. Wang, T. A. Tsiftsis, and P. K. Sharma, "Resource allocation for energy harvesting-powered D2D communication underlaying UAV-assisted networks," *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 1, pp. 14–24, Mar. 2018.

[18] Z. Zhou, M. Dong, K. Ota, R. Shi, Z. Liu, and T. Sato, "Game-theoretic approach to energy-efficient resource allocation in device-to-device underlay communications," *IET Commun.*, vol. 9, no. 3, pp. 375–385, Feb. 2015.
[19] B. Shang, L. Zhao, K.-C. Chen, and X. Chu, "An economic aspect of device-to-device assisted offloading in cellular networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 4, pp. 2289–2304, Apr. 2018.
[20] 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.
[21] J. Urama *et al.*, "Dynamic social trust associations over D2D communications: An implementation perspective," in *Proc. IEEE Int. Conf. Mobile Services*, Jun./Jul. 2016, pp. 186–189.
[22] N. Mastronarde, V. Patel, J. Xu, L. Liu, and M. van der Schaar, "To relay or not to relay: Learning device-to-device relaying strategies in cellular networks," *IEEE Trans. Mobile Comput.*, vol. 15, no. 6, pp. 1569–1585, Jun. 2015.
[23] M. Neely, "Dynamic optimization and learning for renewal systems," *IEEE Trans. Autom. Control*, vol. 58, no. 1, pp. 32–46, Jun. 2013.
[24] C. Qiu, Y. Hu, and Y. Chen, "Lyapunov optimized cooperative communications with stochastic energy harvesting relay," *IEEE Internet Things J.*, vol. 5, no. 2, pp. 1323–1333, Apr. 2018.
[25] L. Feng, Q. Yang, and J. Li, "Virtual selfishness queue and its application," *IEEE Commun. Lett.*, vol. 21, no. 2, pp. 262–265, Feb. 2017.
[26] L. Feng, Q. Yang, K. Kim, and K. Kwak, "Dynamic rate allocation and forwarding strategy adaption for wireless networks," *IEEE Signal Process. Lett.*, vol. 25, no. 7, pp. 1034–1038, Jul. 2018.
[27] X. Ma, M. Sheng, and Y. Zhang, "Green communications with network cooperation: A concurrent transmission approach," *IEEE Commun. Lett.*, vol. 16, no. 12, pp. 1952–1955, Dec. 2012.
[28] J. Yang, Q. Yang, K. S. Kwak, and R. R. Rao, "Power–delay tradeoff in wireless powered communication networks," *IEEE Trans. Veh. Technol.*, vol. 66, no. 4, pp. 3280–3292, Apr. 2017.
[29] H. Zhu, X. Lin, R. Lu, Y. Fan, and X. Shen, "SMART: A secure multilayer credit-based incentive scheme for delay-tolerant networks," *IEEE Trans. Veh. Technol.*, vol. 58, no. 8, pp. 4628–4639, Oct. 2009.
[30] Q. Xu, Z. Su, and S. Guo, "A game theoretical incentive scheme for relay selection services in mobile social networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6692–6702, Aug. 2016.
[31] S. A. Chowdhur, A. Benslimane, and F. Akhter, "Fair and efficient weighted sum rate maximization for multi-rate secondary users in cognitive radio network," in *Proc. IEEE ICC*, May 2017, pp. 1–6.
[32] H. Yu, M. H. Cheung, L. Huang, and J. Huang, "Power-delay tradeoff with predictive scheduling in integrated cellular and Wi-Fi networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 735–742, Apr. 2016.
[33] W. Xu, Y. Zhang, Q. Shi, and X. Wang, "Energy management and cross layer optimization for wireless sensor network powered by heterogeneous energy sources," *IEEE Trans. Wireless Commun.*, vol. 14, no. 5, pp. 2814–2826, May 2015.
[34] D. P. Palomar and M. Chiang, "A tutorial on decomposition methods for network utility maximization," *IEEE J. Sel. Areas Commun.*, vol. 24, no. 8, pp. 1439–1451, Aug. 2006.
[35] J. Li, M. Peng, Y. Yu, and Z. Ding, "Energy-efficient joint congestion control and resource optimization in heterogeneous cloud radio access networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 12, pp. 9873–9887, Dec. 2016.
[36] A. Beck and L. Tetruashvili, "On the convergence of block coordinate descent type methods," *SIAM J. Optim.*, vol. 23, no. 4, pp. 2037–2060, 2013.
[37] W. Wu, Q. Yang, P. Gong, and K. S. Kwak, "Adaptive multi-homing resource allocation for time-varying heterogeneous wireless networks without timescale separation," *IEEE Trans. Commun.*, vol. 64, no. 9, pp. 3794–3807, Sep. 2016.
[38] N. Omidvar, A. Liu, V. Lau, F. Zhang, D. H. K. Tsang, and M. R. Pakravan, "Optimal hierarchical radio resource management for HetNets with flexible backhaul," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4239–4255, Jul. 2018.
[39] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2009.
[40] D. P. Bertsekas and R. G. Gallager, *Data Networks*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1987.

LI FENG received the B.S. degree in electronic information engineering from the Zhengzhou University of Aeronautics, in 2011, and the M.E. and Ph.D. degrees in telecommunications engineering from Xidian University, in 2014 and 2017, respectively. Since 2018, she has been a Postdoctoral Fellow with the Ultra Wide Band Wireless Communications Research Center, Department of Information and Communication Engineering, Inha University, South Korea. Her research interests include wireless resource allocation, stochastic network optimization, node-selfishness analysis, and nano-communication.



QINGHAI YANG received the B.S. degree in communication engineering from the Shandong University of Technology, China, in 1998, the M.S. degree in information and communication systems from Xidian University, China, in 2001, and the Ph.D. degree in communication engineering from Inha University, South Korea, in 2007. From 2007 to 2008, he was a Research Fellow with Ultra Wide Band-ITRC, South Korea. Since 2008, he has been with Xidian University. His current research interests include autonomic communication, content delivery networks, and long term evolution advanced techniques. He was a recipient of the University-President Award for his Ph.D. degree.

KYEHYUN KIM is a Professor with the Department of Geoinformatic Engineering, Inha University, South Korea. His research interests include advanced GIS applications for water resource and quality management, environmental information systems, and natural hazard mitigation.

KYUNG SUP KWAK received the Ph.D. degree from the University of California at San Diego. He was with Hughes Network Systems and with the IBM Network Analysis Center, USA. Since then, he has been with the School of Information and Communication Engineering, Inha University, South Korea, as a Professor. Since 2003, he has been the Dean of the Graduate School of Information Technology and Telecommunications and has been the Director of the UWB Wireless Communications Research Center, IT Research Center, South Korea. In 2006, he was the President of the Korean Institute of Communication Sciences (KICS), and in 2009, the President of the Korea Institute of Intelligent Transport Systems. In 2008, he was elected as the Inha Fellow Professor, and currently, as the Inha Hanlim Fellow Professor. He has authored or co-authored more than 200 peer-reviewed journal papers. His research interests include multiple access communication systems, mobile and UWB radio systems, the future IoT, wireless body area network: nano network and molecular communications. He is a member of the IEEE, IEICE, KICS, and KIEE. In 1993, he received the Engineering College Achievement Award from Inha University and the Service Award from the Institute of Electronics Engineers of Korea, in 1996 and 1999. He was a recipient of the Distinguished Service Awards from the KICS. He has also received the LG Paper Award, in 1998, and Motorola Paper Award, in 2000. He received the Official Commendations for UWB radio technology research and development from the Ministry of Information and Communication, the Prime Minister, and the President of South Korea, in 2005, 2006, and 2009, respectively. In 2007, he received the Haedong Paper Award, and in 2009, the Haedong Scientific Award of research achievement. He served as the TPC and Track Chair/Organizing Chair for several IEEE related conferences.