IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

SPECIAL SECTION ON MOBILE EDGE COMPUTING

Received November 1, 2017, accepted December 12, 2017, date of publication December 25, 2017, date of current version February 14, 2018.

Digital Object Identifier 10.1109/ACCESS.2017.2786662

Full-Duplex Aided User Virtualization for Mobile Edge Computing in 5G Networks

MING LIU, YUMING MAO, SUPENG LENG, AND SUN MAO
School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China

Corresponding author: Supeng Leng (spleng@uestc.edu.cn)

This work was supported in part by the Fundamental Research Funds for the Central Universities, China, under Grant ZYGX2016Z011, in part by the National Natural Science Foundation of China under Grant 61374189, and in part by the Joint Fund of the Ministry of Education of China and China Mobile under Grant MCM20160304.

ABSTRACT Driven by the increasing demand of intensive computing services and the resource limitation of mobile devices at the edge of mobile networks, mobile edge computing (MEC) is concerned with being an emerging paradigm towards the 5G communications. A main issue of MEC is the coordination of communication, computation, and storage. In this paper, we propose a novel MEC framework with a user virtualization scheme in the software-defined network virtualization cellular network, in which radio resources are virtualized along with computation and storage resources to cooperatively finish MEC services. Besides, we introduce the user virtualization assisted by the full-duplex communication which can extend the radio resource of wireless networks and provide the potential to increase system performance. Moreover, enabled by user virtualization, users can offload the edge computation tasks directly to the MEC server implemented at infrastructure providers or via the virtualized mobile device attributed to different mobile virtual network operators (MVNOs). Under this MEC framework, we formulate the virtual resource allocation as a joint optimization problem. A distributed resource allocation algorithm based on an alternating direction method of multipliers is proposed which can reduce computational complexity and signaling overhead. We evaluate the proposed algorithm through extensive simulations, and the results show that the total utility of MVNOs can be improved significantly which benefited from user virtualization.

INDEX TERMS Full-duplex, mobile edge computing (MEC), resource allocation, software defined network virtualization (SDNV), user virtualization.

I. INTRODUCTION

The increasing demand of computation intensive and latency critical mobile services, such as video format transformation, voice recognition, virtual reality and augmented reality, has attracted great attentions [1]. These services promote the requirement for communication and computation capabilities [2], [3]. Although the performance of mobile devices are benefited from the evolution of the CPU and flash memory technologies, their computation capabilities are still limited for future computation tasks. Furthermore, the battery of mobile devices is a main constraint, which indicates plethoric energy consumption from local intensive computing is unacceptable for user experience.

Motivated by mobile cloud computing, mobile edge computing (MEC) is an emerging technology to address the intensive computation issues [4]–[6]. By pushing mobile computing, network control and storage to the network edges, MEC enable large-scale computation and storage resources at network infrastructures. With MEC, the computation-intensive and latency-critical applications can be achieved at the resource-limited mobile devices [7]. Without relying on the remote cloud server, the demands of rapid response from edge users can be accomplished as long as the computation and storage collaborate with communication.

The heterogeneity of computing, storage and communication resources, and the diversity requirement of network services bring new challenges. Current cellular networks are becoming incapable to meet the growing demand [8]. However, software defined network (SDN) and network function virtualization (NFV) possess potential to address these challenges in future 5G networks [9], [10]. The combination of SDN and NFV has attracted attention from both academia and industry, which is named as software-defined network virtualization (SDNV) [11]. Although large amounts of work has been done in this area, current investigations customarily following former researches by concentrating on the wireless virtualization of the infrastructures, generally ignore the potential capabilities from mobile devices.

To exploit the potential capability of mobile devices, we introduce the full-duplex (FD) communication into the SDNV. The FD communication supports both transmission and receiving in the same band simultaneously [12]. Nodes implemented with full-duplex support amplify and forward (FD-AF) transmission, achieve lower delay and higher spectrum efficiency [13], [14]. Offloading the computation task through FD-AF relay increases the communication capability without extra memory occupation and delay augment. The combination of FD communication and SDNV framework may trigger innovative network design that fully exploit the advantages of both two techniques. Besides, extra loop-interference generates from simultaneous transmission and receiving [15]. Furthermore, the extended virtual resources and user mobility make it is difficult to accomplish the resource management for computation offloading.

To the best knowledge of us, so far no investigation has concerned the virtualization of mobile devices in SDNV network with MEC. In this paper, we investigate the MEC offloading with FD-assisted user virtualization in the SDNV cellular network. Contributions of this paper are summarized as follows:

1) We propose a novel MEC framework with user virtualization scheme in SDNV cellular network, where radio resources are virtualized along with computation and storage resources to cooperatively accomplish MEC tasks. Moreover, enabled by user virtualization, user can offload the edge computation tasks directly to the MEC server implemented via base stations (BSs) or via the virtualized mobile device attributed to different mobile virtual network operators (MVNOs).

2) In order to expand the virtual resources for MEC in the SDNV network, we introduce the FD-assisted user virtualization which enables amplify and forward (AF) transmission for computation task offloading. The mobile devices are virtualized along with infrastructures and controlled by the SDN controller. The user virtualization scheme provides the potential to increase the communication capability for offloading without extra memory occupation and additional delay augment.

3) We formulate the virtual resource allocation strategy as a joint optimization problem. Due to the multi-dimensional heterogeneous resources and the imperfect global channel status indicator (CSI), the control and management of the SDNV cellular network with user virtualization is intractable. A distributed resource allocation algorithm based on alternating direction method of multipliers (ADMM) is proposed which can reduce the computational complexity and signaling overhead.

4) We evaluate the proposed algorithm through extensive simulations with different system parameters. Simulation results show that the total utility of MVNOs can significantly benefit from the user virtualization. Moreover, the proposed distributed algorithm exhibits almost same performance compared to the centralized algorithm while keeping a distributed manner.

The rest of this paper is organized as follows. Section II presents the related work on MEC and SDNV. We introduce the system model in Section III. In Section IV and V, we formulate the joint resource allocation problem and solve it by proposing a distributed algorithm for reducing the complexity and signaling overhead. Section VI provides simulation based performance evaluation. Finally, we conclude this paper in Section VII.

II. RELATED WORK

In this section, we generally summarize the related work on software-defined network virtualization and mobile edge computing.

A. SOFTWARE-DEFINED NETWORK VIRTUALIZATION

With the insulation of data and control and the abstraction of PHY resources and internet services, SDN and NFV provide more flexible access management and service scheduling. One of the key issues in SDN is separating network control and data forwarding functionalities, which lead to centralized and programmable network control [16]. Benefits introduced by SDN include enhanced network control, flexible and efficient network management, and improved network service performance [17]–[19]. Key issue of NFV is providing specific mechanisms to decouple service functions from infrastructure, leveraging virtualization technologies to transfer network function from hardware appliances to software applications. Advantages promised by NFV include simplified service development, more flexible service delivery, and reduced network capital and operation costs [20]–[23]. Both SDN and NFV share common goals and similar technical ideas, and are complementary to each other. The combination of SDN and NFV in future network has drawn attention from both academia and industry [24]–[27]. A framework named software-defined network virtualization (SDNV) combines the SDN principle of separating data and control planes with the NFV principle of decoupling service functions from infrastructures, which gives a clear holistic vision of SDN and NFV integration in 5G network [11]. Merging SDN and NFV allows innovative network designs to fully exploit the advantages of both paradigms. A number of research efforts have been devoted to this area, current investigations customarily following former researches by concentrating on the wireless virtualization of the infrastructures, generally ignore the potential capabilities from mobile devices. Consider the gigantic quantities of mobile devices in future 5G network which may consist a large amount of network resources, how to effectively organize these resources is still a problem has not been exploit yet.

B. MOBILE EDGE COMPUTING

The high-rate and highly-reliable air interface allows to run computing services of mobile devices at the remote cloud data center, resulting in the research area named Mobile

Cloud Computing (MCC) [2]. However, MCC architecture is inefficient to accomplish the computation-intensive and latency-critical applications. MEC promises dramatic reduction in latency and mobile energy consumption, tackling key challenges for materializing 5G vision [28]. A principal emphasis on MEC investigation is to seamlessly merge the two disciplines of wireless communications and mobile computing, producing a wide-range of new designs ranging from computation offloading techniques to network architectures [30]. Generally, if a mobile device demands high computation rate, or when executing a given task on the mobile device consumes higher energy than executing on an MEC server, computation offloading is performed. Meanwhile, MEC is expected to enable the network to support extensive innovative services and applications. In the case of LTE, the MEC server can be integrated into the eNodeB directly [30]. Recent years, a lot of research efforts have been devoted to MEC to enable real-time application and reducing the energy consumption [31]–[35]. It should be noted that as the MEC is a immature technology, there are still many issues need to be addressed, such as the coordination of communication, computation, and storage resources, computation resource allocation and offloading decision.

III. SYSTEM MODEL

In this section, we present the system model for SDNV with user virtualization, full-duplex communication and computation.

A. SDVN WITH USER VIRTUALIZATION

Offloading the computation tasks to the base station (BS) may cost quite a few energy for communication when the CSI is poor for users at the edge network. Moreover, energy has been concerned as a key parameter in 5G green networks. Computation offloading via virtualized users may reduce the energy consumption and improve the communication capability. That is cost-effective for offloading and motivates us to exploit the better communication architecture. FD communication achieves better spectrum efficiency than conventional half duplex communication, which enables transmission and receiving in the same band simultaneously [13]. Thus, we introduce the concept named user virtualization assisted by FD communication which enables users to amplify and forward (AF) transmission for computation task offloading. User virtualization provides the abstraction of the mobile devices which might attribute to different MVNOs.

As shown in Fig. 1, in our SDNV-MEC framework, users attribute to different MVNOs are virtualized as complementary resources for InPs. Thus, mobile devices can access cellular network through either network infrastructures or the virtualized mobile devices. The control and management of both infrastructure and virtualized users are implemented by the SDN controller. Consider the mobility of users, the mobile management function should be supplemented for the SDN controller, that is not difficult by network function virtualization [20]. The SDNV-MEC framework with user virtualization assisted by FD communication promotes the computation capability without extra memory occupation and additional delay augment for computation offloading. Accomplished by the MEC server, computation tasks will be transmitted by the BS back to the offloading UE, via the RSs or virtualized users. In order to focus on the computation offloading in MEC with user virtualization, the results feedback of computation tasks are not investigated in out work.

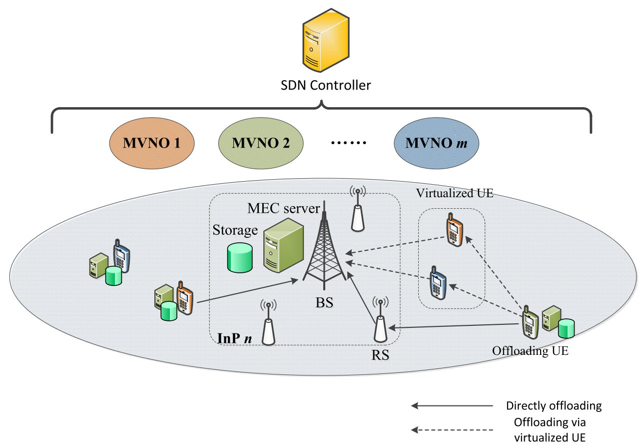

FIGURE 1. SDNV-MEC framework with user virtualization.

It is assumed that a SDNV network contains multiple relay stations (RSs) coexisted within the coverage of a number of BSs as the infrastructure of InPs. The BS and RSs deployed by InP are leased to multiple MVNOs through the network function virtualization and management of virtualized wireless resources. The SDN controller is deployed at the BS possesses the virtualized resource management function. The set of MVNOs is indicated as $\mathcal{M}$, where $\mathcal{M} = \{0, 1, \ldots, M\}$. For the sake of brevity, we denote $\mathcal{K} = \sum k_M, M \in \mathcal{M}$ as the set of users attributed to different MVNOs. For all the users with mobile edge computing tasks, we denote $\mathcal{K} = \{0, 1, \ldots, k\}$ as the set of users. Let $\mathcal{N}$ denotes the set of InPs, where each InP consist a BS along with several RSs. Then, the set of BSs is indicated as $\mathcal{N}_{bs} = \{0, 1, \ldots, N\}$. Assume there are few UEs are virtualized which are capable of relaying the offloading data for other users. Thus, other UEs are enabled to access the wireless network via virtualized UEs in a FD-AF way without extra memory occupation and additional delay augment. Let $\mathcal{N}_{vu} = \{1, \ldots N_{vu}\}$ be the set of virtualized users who potentially perform a relay function for InP. We assume there are at most two-hops relay transmission for simplifying the resource management. It should be noted that virtualized users are only allowed to access the BS when playing the relay role. We denote $\mathcal{N}_{rs}$ as the set of FD-RSs, where $\mathcal{N}_{rs} = \{0, 1, \ldots, N_{rs}\}$. Let $\mathcal{J}_n = \{0, \ldots J\} = \mathcal{N}_{vu} \cup \mathcal{N}_{rs} \cup \mathcal{N}_{bs}$ represents the set of potentially offloading receivers of the network including BSs, RSs and virtualized UEs in the nth InP. In the context of SDNV networks, user k_M of the Mth MVNO is associate with either the BS $\mathcal{N}_{bs}$ by one-hop transmissions or the virtualized users $\mathcal{N}_{vu}$ and RSs $\mathcal{N}_{rs}$ via two-hop relay transmissions.

B. FULL-DUPLEX COMMUNICATION MODEL

Consider the spatial locations of UEs directly effect the distribution of distance $d_{i,j}$ between node i and node j, we employ the combination of both large scale attenuation and small scale attenuation as our channel model. The standard power loss propagation model is used with a path loss exponent $\sigma > 2$. As far as random channel effects such as fading and shadowing, we assume, unless otherwise noted, that the tagged base station and tagged user experience only Rayleigh fading with mean 1. Then the receiving signal at a typical node which is a distance $d_{i,j}$ from its base station or paired UE is given by

$$y(t) = x(t)h(t)d_{i,j}^{-\sigma} \tag{1}$$

The spectrum efficiency of the FD relay link between UE to BS via the RSs is given by

$$r_{ue,rs} = \min\{\log 2(1 + \frac{p_{ue}|h_{ue,rs}|^2}{N_0 + p_{rs}\eta}), \times \log 2(1 + \frac{p_{rs}|h_{rs,bs}|^2}{N_0 + p_{ue}|h_{ue,bs}|^2})\} \tag{2}$$

where p_{ue} and p_{rs} denote the transmission power of user and RSs, respectively. $h_{ue,rs}$ represents the channel gain between UE and RS. $h_{rs,bs}$ represents the channel gain between RS and BS. $h_{ue,bs}$ represents the channel gain between UE and BS. N_0 denotes the power spectrum density (PSD) of the additive white Gaussian noise (AWGN). η denotes the self-interference factor of FD communication [36].

The spectrum efficiency of the directly access link between UE and BS is given by

$$r_{ue,bs} = \log 2(1 + \frac{p_{ue}|h_{ue,bs}|^2}{N_0}) \tag{3}$$

where p_{ue} denotes the transmission power of UE. $h_{ue,bs}$ represents the channel gain between UE and the BS.

C. COMPUTATION MODEL

Assume the k_Mth user generate a computation task $A_{k_M}(L_{k_M}, \tau_{k_M}, c_{k_M})$ to be assigned to the MEC server n, where L_{k_M} denotes the task input-data size of the kth user of the Mth MVNO (in bits), τ_{k_M} denotes the completion deadline (in second), and c_{k_M} denotes the computing ability of the user k_M required for accomplishing task, which can be quantized by the amount of CPU cycles per bit. Let $f_{k_M,n}$ denotes the computation capability of the BS n for the task from the k_Mth user, which is quantized by the total number of CPU cycles per second. Thus, execution latency for task $A_{k_M}(L_{k_M}, \tau_{k_M}, c_{k_M})$ at the nth MEC server can be calculated accordingly to

$$d_{k_M,n} = \frac{c_{k_M}L_{k_M}}{f_{k_M,n}} \tag{4}$$

Assume the time slot is t, then we have the total computation cycles of each slot at BS is $f_{k_M,n}t$, which is denoted as the computation capability of the MEC server. The computation requirement of task k_M quantized by the total number of CPU cycles is given by

$$\omega_{k_M} = L_{k_M}c_{k_M} \tag{5}$$

Consider a MEC server that handle several computation tasks from UEs and the k_Mth task is allocated with ω_{k_M} CPU cycles with CPU-cycle frequency $f_{k_M,n}$. The power consumption of CPU can be divided into several factors including the dynamic, short-circuit, and leakage power consumption, where the dynamic power consumption dominates the others [37]. The energy consumption of a CPU cycle is given by ξf^2, where ξ is a constant related to the hardware architecture [38]. Hence, the total energy consumed by CPU of the MEC server n is denoted by E_n can be expressed as

$$E_n = \sum_{k_M=1}^{\mathcal{K}} \xi\omega_{k_M}f_{n,i}^2 \tag{6}$$

where $\omega_{k_M} = L_{k_M}c_{k_M}$ denotes the total number of CPU cycles which was given in (5).

IV. PROBLEM FORMULATION

In this section, we formulate the user association, radio resource allocation and computation offloading as a joint optimization problem.

A. CONSTRAINTS

Assume that each user has a computation task to be completed with a certain requirement of computation rate. Let binary $a_{k_M,n}$ denotes the user association indicator, ie., $a_{k_M,n} = 0$ represents that the k_Mth user from the Mth MVNO is associate with the nth potentially receiver and 0 otherwise. Practically, each user can be associated with only one receiver, thus

$$\sum_{n\in\mathcal{N}} a_{k_M,n} = 1, \quad \forall k_M \in \mathcal{K},\ M \in \mathcal{M} \tag{7}$$

We employ $x_{k_M,n}$ indicates the allocated band from spectrum provider to user i via the nth potentially receiver, where $x_{k_M,n} \in [0, 1]$. Due to the limitation of the spectrum provider, total allocated bandwidth to users is constrained as

$$\sum_{k_M\in\mathcal{K}}\sum_{M\in\mathcal{M}}\sum_{j\in\mathcal{J}} a_{k_M,n_j}x_{k_M,n_j}B_n \le B_n, \quad \forall n \in \mathcal{N}_{bs} \tag{8}$$

The communication data rate for offloading the computation task A_{k_M} should be guaranteed no less than the data rate requirement.

$$\sum_{n\in\mathcal{N}} a_{k_M,n}x_{k_M,n}Br_{k_M,n} \ge R_{k_M,n}^{min}, \quad \forall k_M \in \mathcal{K},\ M \in \mathcal{M} \tag{9}$$

The latency of computation task should be guaranteed less than the completion deadline τ_{k_M}, we have

$$\sum_{n\in\mathcal{N}_{bs}} d_{k_M,n} \le \tau_{k_M}, \quad \forall k_M \in \mathcal{K},\ M \in \mathcal{M} \tag{10}$$

Let $y_{k_M,n}$ denotes the computation resource of MEC server n allocated to the k_Mth task, where $y_{k_M,n} \in [0, 1]$.

Then, $y_{k_M,n}f_{k_M,n}t$ denotes the total computation cycles allocated to task k_M. By substituting $y_{k_M,n}f_{k_M,n}t$ into (10), we have

$$\sum_{n\in\mathcal{N}} a_{k_M,n}y_{k_M,n}f_{k_M,n}t \geq \omega_{k_M}, \quad \forall k_M \in \mathcal{K},\ M \in \mathcal{M} \tag{11}$$

where $\omega_{k_M} = L_{k_M}c_{k_M}$. This constraints the computation rate is no less than task requirement. It should be noted that the computation ability at each BS is limited, the total request computation resources cannot exceed the total amount of the computation resource of the MEC server.

$$\sum_{k_M\in\mathcal{K}}\sum_{M\in\mathcal{M}}\sum_{j\in\mathcal{J}} a_{k_M,n_j}y_{k_M,n_j} \leq 1, \quad \forall n \in \mathcal{N}_{bs} \tag{12}$$

Each task A_{k_M} consists data size L_{k_M} which cause to memory occupation, but the memory at each MEC server is not infinity. Thus, the total memory occupation of all the tasks offloaded at the MEC server is strictly no more than the maximum memory size C_{bs}.

$$\sum_{k_M\in\mathcal{K}}\sum_{M\in\mathcal{M}}\sum_{j\in\mathcal{J}} a_{k_M,n_j}L_{k_M} \leq C_n, \quad \forall n \in \mathcal{N}_{bs} \tag{13}$$

B. UTILITY FUNCTION

In this paper, we set the maximization of the revenue of MVNOs as our goal. The utility for the potential transmission between users k_M and BS ($n \in \mathcal{N}_{bs}$) can be defined as

$$\begin{aligned} U_{k_M,n} = {} & a_{k_M,n}\alpha_n x_{k_M,n}Br_{k_M,n} + a_{k_M,n}\phi_n R_{k_M,n} \\ & - a_{k_M,n}x_{k_M,n}\beta_n - a_{k_M,n}\psi_n E_{k_M,n} - a_{k_M,n_j}L_{k_M}\delta_n \end{aligned} \tag{14}$$

where term $a_{k_M,n}\alpha_n x_{k_M,n}Br_{k_M,n}$ denotes the income of MVNO from users to access the SDNV networks, α denotes the revenue per unit of the data rate. $a_{k_M,n}x_{k_M,n}\beta_n$ denotes the expense of MVNOs to pay for the usage of spectrum bandwidth, β is the price per unit of the consumed spectrum. $a_{k_M,n}\phi_n R_{k_M,n}$ denotes the income of MVNOs from users to access the MEC servers for executing the offloaded computation tasks, ϕ is the price per unit of the computation rate. $a_{k_M,n}\psi_n E_{k_M,n}$ denotes the cost of MVNOs to pay for the usage of MEC servers, ψ is the cost per unit of the computation energy. The last term $a_{k_M,n_j}L_{k_M}\delta_n$ represents the fee from MVNOs on caching the computation tasks, δ is cost per unit of the storage occupation.

The utility for the potential transmission between users k_M and RSs or virtualized UE ($n \in \mathcal{N}_{vu} \cup \mathcal{N}_{rs}$) is

$$\begin{aligned} U_{k_M,n} = {} & a_{k_M,n}\alpha_n x_{k_M,n}Br_{k_M,n} + a_{k_M,n}\phi_n R_{k_M,n} \\ & - a_{k_M,n}x_{k_M,n}\beta_n - a_{k_M,n}\varphi_n - a_{k_M,n}\psi_n E_{k_M,n} \\ & - a_{k_M,n_j}L_{k_M}\delta_n \end{aligned} \tag{15}$$

where the term $a_{k_M,n}\varphi_n$ evaluating the expense of MVNO to pay for the usage of RS or virtualized UEs, where φ_n denotes the unit fee for user virtualization.

C. PROBLEM FORMULATION

We adopt the utility function proposed in (15) as the objective function of our optimization problem, and the problem is formulated as

$$\begin{aligned} P1: & \\ & \max_{a_{k_M,n},x_{k_M,n},y_{k_M,n}} \sum_{k_M\in\mathcal{K}}\sum_{M\in\mathcal{M}}\sum_{n\in\mathcal{N}} U_{k_M,n} \\ & s.t.\ C1: \sum_{n\in\mathcal{N}} a_{k_M,n} = 1, \quad \forall k_M \in \mathcal{K},\ M \in \mathcal{M} \\ & \qquad C2: \sum_{k_M\in\mathcal{K}}\sum_{M\in\mathcal{M}}\sum_{j\in\mathcal{J}} a_{k_M,n_j}x_{k_M,n_j} \leq 1, \quad \forall n \in \mathcal{N}_{bs} \\ & \qquad C3: \sum_{n\in\mathcal{N}} a_{k_M,n}x_{k_M,n}Br_{k_M,n} \geq R^{min}_{k_M}, \\ & \qquad\quad \forall k_M \in \mathcal{K}, \quad M \in \mathcal{M} \\ & \qquad C4: \sum_{n\in\mathcal{N}} a_{k_M,n}y_{k_M,n}f_{k_M,n}t \geq \omega_{k_M}, \\ & \qquad\quad \forall k_M \in \mathcal{K}, \quad M \in \mathcal{M} \\ & \qquad C5: \sum_{k_M\in\mathcal{K}}\sum_{M\in\mathcal{M}}\sum_{j\in\mathcal{J}} a_{k_M,n_j}y_{k_M,n_j} \leq 1, \\ & \qquad\quad \forall n \in \mathcal{N}_{bs} \\ & \qquad C6: \sum_{k_M\in\mathcal{K}}\sum_{M\in\mathcal{M}}\sum_{j\in\mathcal{J}} a_{k_M,n_j}L_{k_M} \leq C_n, \quad \forall n \in \mathcal{N}_{bs} \end{aligned} \tag{16}$$

where $U_{k_M,n}$ is the potential utility of user k_M associating with receiver n. According to constraint C1, each user can be associated with only one receiver. Constraint C2 denotes that the total allocated bandwidth to all UEs associated with BS n is unable to exceed the total bandwidth of BS n. In C3, we guarantee the communication data rate requirement for each task. C4 indicates that the executing latency cannot exceed the task requirement. The computation and memory resources of each MEC server at BS n is guaranteed by C5 and C6, respectively.

D. PROBLEM REFORMULATION

Problem **P1** is difficult to solve due to the following observations: Due to the fact that $a_{k_M,n}$ is a binary variable, the feasible set of problem **P1** is non-convex. There exist product relationships between $a_{k_M,n}$ and linear function of $x_{k_M,n}$, as well as $y_{k_M,n}$. Thus the objective function of problem **P1** is not a convex function.

The problem has a quite large size due to the characteristics of the SDNV network. If we assume that the average number of UEs of one MVNO is k, the number of variables in this problem could reach kN, and the complexity for a central algorithm to find a globally optimal solution will be $O((kN)^x)$ ($x > 0$, $x = 1$ implies a linear algorithm while $x > 1$ implies a polynomial time algorithm) even if we simply consider all the variables as binary variables. In addition, the number of MVNOs in the future SDNV network is increasing as time goes on, which results in an even more radically increasing complexity in our problem.

As is shown, problem **P1** is a mixed integer and nonconvex optimization problem [39], [40], and such problems are usually considered as NP-hard problems [41]. Therefore, a transformation and simplification of the original problem is composed of the following two steps:

1) Binary variables relaxation: In order to transform the nonconvex feasible set of original problem into a convex set, we relax the binary association variable $a_{k_M,n}$ into a continuous variable such that $0 \leq a_{k_M,n} \leq 1$, which can be interpreted as the k_Mth user associating with multiple receivers in a time-division multiple access way [42].

2) Substitution of the product terms: Due to the nonconvex objective function, the problem is still intractable even though we relax the binary variable. Then we will propose a proposition of equivalent problem of **P1** to solve it.

Proposition 1: If we substitute $\hat{x}_{k_M,n} = a_{k_M,n}x_{k_M,n}$, and $\hat{y}_{k_M,n} = a_{k_M,n}y_{k_M,n}$ into the original problem **P1**, there exists an equivalent formulation, which is given by (17).

$P2$:

$$\begin{aligned}
&\max_{a_{k_M,n},\tilde{x}_{k_M,n},\tilde{y}_{k_M,n}} \sum_{k_M\in\mathcal{K}}\sum_{M\in\mathcal{M}}\sum_{n\in\mathcal{N}} \overline{U}_{k_M,n}\\
&s.t.\ C1: \sum_{n\in\mathcal{N}} a_{k_M,n} = 1, \quad \forall k_M \in \mathcal{K},\ M \in \mathcal{M}\\
&\quad C2: \sum_{k_M\in\mathcal{K}}\sum_{M\in\mathcal{M}}\sum_{j\in\mathcal{J}} \tilde{x}_{k_M,n_j} \leq 1, \quad \forall n\in\mathcal{N}_{bs}\\
&\quad C3: \sum_{n\in\mathcal{N}} \tilde{x}_{k_M,n} Br_{k_M,n} \geq R_{k_M}^{min},\\
&\qquad \forall k_M \in \mathcal{K}, \quad M \in \mathcal{M}\\
&\quad C4: \sum_{n\in\mathcal{N}} \tilde{y}_{k_M,n} f_{k_M,n} t \geq \omega_{k_M},\\
&\qquad \forall k_M \in \mathcal{K},\ M \in \mathcal{M}\\
&\quad C5: \sum_{k_M\in\mathcal{K}}\sum_{M\in\mathcal{M}}\sum_{j\in\mathcal{J}} \tilde{y}_{k_M,n_j} \leq 1, \quad \forall n\in\mathcal{N}_{bs}\\
&\quad C6: \sum_{k_M\in\mathcal{K}}\sum_{M\in\mathcal{M}}\sum_{j\in\mathcal{J}} a_{k_M,n_j} L_{k_M} \leq C_n, \quad \forall n \in \mathcal{N}_{bs}\\
&\quad C7: a_{k_M,n} \geq \tilde{x}_{k_M,n}, \quad a_{k_M,n} \geq \tilde{y}_{k_M,n},\\
&\qquad \forall k_M \in \mathcal{K}, \quad M \in \mathcal{M} n \in \mathcal{N}
\end{aligned} \tag{17}$$

Where constraint C7 enforces that the association indicator is no less than the spectrum resource allocation indicator and computing resource indicator. Intuitively, if $a_{k_M,n} > 0$, $\tilde{x}_{k_M,n}$ can be any nonnegative; if $a_{k_M,n} = 0$, $\tilde{x}_{k_M,n} = 0$ must hold. The relaxed problem **P1** can be directly recovered through substituting the variables $\tilde{x}_{k_M,n} = a_{k_M,n}x_{k_M,n}$, and $\tilde{y}_{k_M,n} = a_{k_M,n}y_{k_M,n}$ into problem **P2**. Afterwards, the mapping between $\{a_{k_M,n}, x_{k_M,n}, y_{k_M,n}\}$ and $\{a_{k_M,n}, \tilde{x}_{k_M,n}, \tilde{y}_{k_M,n}\}$ can be obtained as

$$x_{k_M,n} = \begin{cases} \dfrac{\tilde{x}_{k_M,n}}{a_{k_M,n}}, & a_{k_M,n} > 0 \\ 0, & \text{otherwise} \end{cases} \tag{18}$$

$$y_{k_M,n} = \begin{cases} \dfrac{\tilde{y}_{k_M,n}}{a_{k_M,n}}, & a_{k_M,n} > 0 \\ 0, & \text{otherwise} \end{cases} \tag{19}$$

Therefore, original problem **P1** is transformed into a convex problem **P2**. A lot of mathematical tools on convex optimization can be used to solve this type of problems, e.g. interior point method and dual decomposition [38]. However, the size of our problem will greatly increase when the number of InPs and MVNOs grows in future 5G networks. Besides, the signaling overhead of local information (such as Channel Status Indicator (CSI)) acquirement could be extremely high when the size of problem is large. Thus, it is more efficient and practically to implement a distributed optimization algorithm executing on each BS, and this method significantly reduce the signaling overhead of the SDNV wireless network.

V. ADMM-BASED DISTRIBUTED RESOURCE ALLOCATION ALGORITHM

In this section, we introduce a distributed algorithm named alternating direction method of multipliers (ADMM) [43] to solve the optimization problem. Due to the coupling relationship of global variables $\{a, x, y\}$ and constraints $C1$, $C3$, and $C4$, $P2$ is inseparable to be executed on each BS. In order to apply the distributed algorithm, we first decouple the problem **P2**.

A. PROBLEM DECOMPOSITION

In order to decouple the coupling variables, we introduce local copies of $\{a_{k_M,n}, \tilde{x}_{k_M,n}$, and $\tilde{y}_{k_M,n}\}$ for each BS $n \in \mathcal{N}_{bs}$, which can be locally determined at each BS. To lighten the notation, k is employed to denote each user instead of k_M. Let $\{\hat{a}^n_{k,i}\}$, $\{\hat{x}^n_{k,i}\}$, and$\{\hat{y}^n_{k,i}\}$ be the local copies of $\{a_{k,n}\}$, $\{\tilde{x}_{k,n}\}$, and$\{\tilde{y}_{k,n}\}$, respectively. Then, the feasible set of local variables for each BS n can be defined as

$$\chi_n = \left\{ \begin{matrix} \{\hat{a}^n_{k,i}\} \\ \{\hat{x}^n_{k,i}\} \\ \{\hat{y}^n_{k,i}\} \end{matrix} \left| \begin{aligned}
&\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{J}_n} \hat{a}^n_{k,i_j} = 1, \forall k \in \mathcal{K}\\
&\sum_{k\in\mathcal{K}} \hat{x}^n_{k,i} + \sum_{k\in\mathcal{K}}\sum_{j\in\mathcal{J}_n} \hat{x}^n_{k,i_j} \leq 1, \forall i \in \mathcal{N}_{\lfloor\int}\\
&\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{J}_n} \hat{x}^n_{k,i_j} Br_{k,i_j} \geq R^{min}_k, \forall k \in \mathcal{K}\\
&\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{J}_n} \hat{y}^n_{k,i_j} f_i t \geq \omega_k, \forall k \in \mathcal{K}\\
&\sum_{k\in\mathcal{K}}\sum_{j\in\mathcal{J}_n} \hat{y}^n_{k,i_j} \leq 1, \forall i \in \mathcal{N}_{bs}\\
&\sum_{k\in\mathcal{K}}\sum_{j\in\mathcal{J}_n} \hat{a}^n_{k,i_j} L_k \leq C_i, \forall i \in \mathcal{N}_{bs}\\
&\hat{a}^n_{k,i_j} \geq \hat{x}^n_{k,i_j}, \hat{a}^n_{k,i_j} \geq \hat{y}^n_{k,i_j},\\
&\forall k \in \mathcal{K}, j \in \mathcal{J}_n, i \in \mathcal{N}
\end{aligned} \right. \right\} \tag{20}$$

The local utility function of each BS $n \in \mathcal{N}_{bs}$ is given by

$$f_n = \begin{cases} -\sum_{k\in\mathcal{K}} \hat{U}_{k,n}, (\{\hat{a}^n_{k,i}\}, \{\hat{x}^n_{k,i}\}, \{\hat{y}^n_{k,i}\}) \in \chi_n \\ 0, \text{otherwise}. \end{cases} \quad (21)$$

where $\hat{U}_{k,n}$ is obtained by substituting $\{\hat{a}^n_{k,i}\}$, $\{\hat{x}^n_{k,i}\}$, and$\{\hat{y}^n_{k,i}\}$ into $\overline{U}_{k,n}$ in (17). The object function is obviously convex with respect to $(\{\hat{a}^n_{k,i}\}, \{\hat{x}^n_{k,i}\}, \{\hat{y}^n_{k,i}\})$. Problem **P2** can be equivalently rewritten as

$$\begin{aligned} P3: \quad & \min \sum_{n\in\mathcal{N}_{bs}} f_n(\hat{a}^n_{k,i}, \hat{x}^n_{k,i}, \hat{y}^n_{k,i}) \\ & s.t.\ \{\hat{a}^n_{k,i}\} = \{a_{k,i}\}, \quad \{\hat{x}^n_{k,i}\} = \{\tilde{x}_{k,i}\}, \ \{\hat{y}^n_{k,i}\} = \{\tilde{y}_{k,i}\} \\ & \quad \forall i, k, n \end{aligned} \quad (22)$$

It can be observed that, the objective function is separable across all BSs, but the consensus constraints are still coupled on BSs.

B. PROBLEM SOLVING VIA ADMM

After the decomposition of problem, we intend to derive a distributed consensus optimization via ADMM. Firstly, we formulate the augmented Lagrangian [43] for (22) which is given by

$$\begin{aligned} &\mathcal{L}_\rho(\{\hat{\boldsymbol{a}}, \hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}\}, \{\boldsymbol{a}, \tilde{\boldsymbol{x}}, \tilde{\boldsymbol{y}}\}, \boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{\nu}) \\ &= \sum_{n\in\mathcal{N}_{bs}} f_n(\hat{a}^n_{k,i}, \hat{x}^n_{k,i}, \hat{y}^n_{k,i}) + \sum_{n\in\mathcal{N}_{bs}} \sum_{\substack{i\in\mathcal{N}_{bs}\\k\in\mathcal{K}}} \lambda^n_{k,i}(\hat{a}^n_{k,i} - a_{k,i}) \\ &\quad + \sum_{n\in\mathcal{N}_{bs}} \sum_{\substack{i\in\mathcal{N}_{bs}\\k\in\mathcal{K}}} \mu^n_{k,i}(\hat{x}^n_{k,i} - \tilde{x}_{k,i}) + \sum_{n\in\mathcal{N}_{bs}} \sum_{\substack{i\in\mathcal{N}_{bs}\\k\in\mathcal{K}}} \nu^n_{k,i}(\hat{y}^n_{k,i} - \tilde{y}_{k,i}) \\ &\quad + \frac{\rho}{2} \sum_{n\in\mathcal{N}_{bs}} \sum_{\substack{i\in\mathcal{N}_{bs}\\k\in\mathcal{K}}} (\hat{a}^n_{k,i} - a_{k,i})^2 + \frac{\rho}{2} \sum_{n\in\mathcal{N}_{bs}} \sum_{\substack{i\in\mathcal{N}_{bs}\\k\in\mathcal{K}}} (\hat{x}^n_{k,i} - \tilde{x}_{k,i})^2 \\ &\quad + \frac{\rho}{2} \sum_{n\in\mathcal{N}_{bs}} \sum_{\substack{i\in\mathcal{N}_{bs}\\k\in\mathcal{K}}} (\hat{y}^n_{k,i} - \tilde{y}_{k,i})^2 \end{aligned} \quad (23)$$

where ρ is named the penalty parameter, $\hat{\boldsymbol{a}} = \{\hat{a}^n_{k,i}\}$, $\hat{\boldsymbol{x}} = \{\hat{x}^n_{k,i}\}$, $\hat{\boldsymbol{y}} = \{\hat{y}^n_{k,i}\}$. $\boldsymbol{a} = \{a_{k,i}\}$, $\tilde{\boldsymbol{x}} = \{\tilde{x}_{k,i}\}$, $\tilde{\boldsymbol{y}} = \{\tilde{y}_{k,i}\}$. $\boldsymbol{\lambda} = \{\lambda^n_{k,i}\}$, $\boldsymbol{\mu} = \{\mu^n_{k,i}\}$, and $\boldsymbol{\nu} = \{\nu^n_{k,i}\}$ are the associated dual variables. The convergence of the ADMM-based algorithm is effected by the penalty parameter ρ. A larger value of ρ will make the primal dual quick converge to zero, but it will also result in an increased dual residual [44]. Thus, a proper value of ρ is of enormous important to control the process of the ADMM algorithm. Compared to the standard Lagrangian, the performance of the distributed algorithm can be promoted by adding the quadratic penalty term. The solution of ADMM algorithm is equivalent to problem **P3** due to the penalty term is generally zero for any feasible solution.

The process for solving the problem **P3** with ADMM algorithm consists the following iterations:

1) Local variables update

$$\begin{aligned} &\{\hat{\boldsymbol{a}}^n, \hat{\boldsymbol{x}}^n, \hat{\boldsymbol{y}}^n\}^{[t+1]}_{n\in\mathcal{N}_{bs}} \\ &= \arg\min \Bigg\{ f_n(\hat{a}^n_{k,i}, \hat{x}^n_{k,i}, \hat{y}^n_{k,i}) \\ &\quad + \sum_{\substack{i\in\mathcal{N}_{bs}\\k\in\mathcal{K}}} \lambda^{n[t]}_{k,i}(\hat{a}^n_{k,i} - a^{[t]}_{k,i}) + \sum_{\substack{i\in\mathcal{N}_{bs}\\k\in\mathcal{K}}} \mu^{n[t]}_{k,i}(\hat{x}^n_{k,i} - \tilde{x}^{[t]}_{k,i}) \\ &\quad + \sum_{\substack{i\in\mathcal{N}_{bs}\\k\in\mathcal{K}}} \nu^{n[t]}_{k,i}(\hat{y}^n_{k,i} - \tilde{y}^{[t]}_{k,i}) + \frac{\rho}{2} \sum_{\substack{i\in\mathcal{N}_{bs}\\k\in\mathcal{K}}} (\hat{a}^n_{k,i} - a^{[t]}_{k,i})^2 \\ &\quad + \frac{\rho}{2} \sum_{\substack{i\in\mathcal{N}_{bs}\\k\in\mathcal{K}}} (\hat{x}^n_{k,i} - \tilde{x}^{[t]}_{k,i})^2 + \frac{\rho}{2} \sum_{\substack{i\in\mathcal{N}_{bs}\\k\in\mathcal{K}}} (\hat{y}^n_{k,i} - \tilde{y}^{[t]}_{k,i})^2 \Bigg\} \end{aligned} \quad (24)$$

where superscript $[t]$ denotes the iteration index. After eliminating the constant term, the local variables can be updated by solving the follow problem at iteration $t + 1$;

$$\begin{aligned} P4: \quad & \min -\sum_{k\in\mathcal{K}} \hat{U}_{k,n} + \sum_{\substack{i\in\mathcal{N}_{bs}\\k\in\mathcal{K}}} \lambda^{n[t]}_{k,i} \hat{a}^n_{k,i} + \sum_{\substack{i\in\mathcal{N}_{bs}\\k\in\mathcal{K}}} \mu^{n[t]}_{k,i} \hat{x}^n_{k,i} \\ & \quad + \sum_{\substack{i\in\mathcal{N}_{bs}\\k\in\mathcal{K}}} \nu^{n[t]}_{k,i} \hat{y}^n_{k,i} + \frac{\rho}{2} \sum_{\substack{i\in\mathcal{N}_{bs}\\k\in\mathcal{K}}} (\hat{a}^n_{k,i} - a^{[t]}_{k,i})^2 \\ & \quad + \frac{\rho}{2} \sum_{\substack{i\in\mathcal{N}_{bs}\\k\in\mathcal{K}}} (\hat{x}^n_{k,i} - \tilde{x}^{[t]}_{k,i})^2 + \frac{\rho}{2} \sum_{\substack{i\in\mathcal{N}_{bs}\\k\in\mathcal{K}}} (\hat{y}^n_{k,i} - \tilde{y}^{[t]}_{k,i})^2 \\ & s.t.\ (\hat{\boldsymbol{a}}^n, \hat{\boldsymbol{x}}^n, \hat{\boldsymbol{y}}^n) \in \boldsymbol{\chi}_n \end{aligned} \quad (25)$$

It can be proved **P4** is a convex problem as its objective function and feasible set are convex. The primal dual interior-point method is an efficient way to solve this problem. The details of primal dual interior-point method are omitted due to the limited space.

2) Global variables update

In this step, global variables are updated according to

$$\begin{aligned} \{\boldsymbol{a}\}^{[t+1]} = \arg\min \Bigg\{ & \sum_{n\in\mathcal{N}_{bs}} \sum_{\substack{i\in\mathcal{N}_{bs}\\k\in\mathcal{K}}} \lambda^{n[t]}_{k,i}(\hat{a}^{n[t+1]}_{k,i} - a_{k,i}) \\ & + \frac{\rho}{2} \sum_{n\in\mathcal{N}_{bs}} \sum_{\substack{i\in\mathcal{N}_{bs}\\k\in\mathcal{K}}} (\hat{a}^{n[t+1]}_{k,i} - a_{k,i})^2 \Bigg\} \end{aligned} \quad (26)$$

$$\begin{aligned} \{\tilde{\boldsymbol{x}}\}^{[t+1]} = \arg\min \Bigg\{ & \sum_{n\in\mathcal{N}_{bs}} \sum_{\substack{i\in\mathcal{N}_{bs}\\k\in\mathcal{K}}} \mu^{n[t]}_{k,i}(\hat{x}^{n[t+1]}_{k,i} - \tilde{x}_{k,i}) \\ & + \frac{\rho}{2} \sum_{n\in\mathcal{N}_{bs}} \sum_{\substack{i\in\mathcal{N}_{bs}\\k\in\mathcal{K}}} (\hat{x}^{n[t+1]}_{k,i} - \tilde{x}_{k,i})^2 \Bigg\} \end{aligned} \quad (27)$$

$$\begin{aligned} \{\tilde{\boldsymbol{y}}\}^{[t+1]} = \arg\min \Bigg\{ & \sum_{n\in\mathcal{N}_{bs}} \sum_{\substack{i\in\mathcal{N}_{bs}\\k\in\mathcal{K}}} \nu^{n[t]}_{k,i}(\hat{y}^{n[t+1]}_{k,i} - \tilde{y}_{k,i}) \\ & + \frac{\rho}{2} \sum_{n\in\mathcal{N}_{bs}} \sum_{\substack{i\in\mathcal{N}_{bs}\\k\in\mathcal{K}}} (\hat{y}^{n[t+1]}_{k,i} - \tilde{y}_{k,i})^2 \Bigg\} \end{aligned} \quad (28)$$

Owning to the quadratic regularization term has been added to the augmented Lagrangian, the unconstraint problems (26)-(28) are strictly convex with respect to

($\boldsymbol{a}$, $\tilde{\boldsymbol{x}}$, and $\tilde{\boldsymbol{y}}$). After setting the gradients to zero, we can have

$$\begin{aligned} a_{k,i}^{[t+1]} &= \frac{1}{N}\sum_{n\in\mathcal{N}_{bs}}(\hat{a}_{k,i}^{n[t+1]}+\frac{1}{\rho}\lambda_{k,i}^{n[t]}) \\ \tilde{x}_{k,i}^{[t+1]} &= \frac{1}{N}\sum_{n\in\mathcal{N}_{bs}}(\hat{x}_{k,i}^{n[t+1]}+\frac{1}{\rho}\mu_{k,i}^{n[t]}) \\ \tilde{y}_{k,i}^{[t+1]} &= \frac{1}{N}\sum_{n\in\mathcal{N}_{bs}}(\hat{y}_{k,i}^{n[t+1]}+\frac{1}{\rho}\nu_{k,i}^{n[t]}) \end{aligned} \tag{29}$$

While the dual variables are initialized as zeros, we have $\sum_{n\in\mathcal{N}_{bs}}\lambda_{k,i}^{n[t]}=0$, $\sum_{n\in\mathcal{N}_{bs}}\mu_{k,i}^{n[t]}=0$, and $\sum_{n\in\mathcal{N}_{bs}}\nu_{k,i}^{n[t]}=0$, $\forall k, i$, at each iteration t [42]. Then, (29) is reduced to

$$\begin{aligned} a_{k,i}^{[t+1]} &= \frac{1}{N}\sum_{n\in\mathcal{N}_{bs}}\hat{a}_{k,i}^{n[t+1]} \\ \tilde{x}_{k,i}^{[t+1]} &= \frac{1}{N}\sum_{n\in\mathcal{N}_{bs}}\hat{x}_{k,i}^{n[t+1]} \\ \tilde{y}_{k,i}^{[t+1]} &= \frac{1}{N}\sum_{n\in\mathcal{N}_{bs}}\hat{y}_{k,i}^{n[t+1]} \end{aligned} \tag{30}$$

We can observe that global variables are the average of all the updated local copies in (30). This step of iteration is implemented by the SDN controller of the network. Without involving the dual variables, this step can significantly reduce the signaling overhead.

3) Lagrange multipliers update

In the final step, the Lagrange multipliers is updated in this step by gathering the values from the two steps before.

$$\{\boldsymbol{\lambda}^n\}_{n\in\mathcal{N}_{bs}}^{[t+1]} = \boldsymbol{\lambda}^{n[t]}+\rho\left\{\hat{\boldsymbol{a}}^{n[t+1]}-\boldsymbol{a}^{[t+1]}\right\} \tag{31}$$

$$\{\boldsymbol{\mu}^n\}_{n\in\mathcal{N}_{bs}}^{[t+1]} = \boldsymbol{\mu}^{n[t]}+\rho\left\{\hat{\boldsymbol{x}}^{n[t+1]}-\tilde{\boldsymbol{x}}^{[t+1]}\right\} \tag{32}$$

$$\{\boldsymbol{\nu}^n\}_{n\in\mathcal{N}_{bs}}^{[t+1]} = \boldsymbol{\nu}^{n[t]}+\rho\left\{\hat{\boldsymbol{y}}^{n[t+1]}-\tilde{\boldsymbol{y}}^{[t+1]}\right\} \tag{33}$$

where the augmented Lagrangian parameter ρ represents the step size for updating. The first step and the third step can be completely separable across BSs, when finding the optimal local variables and local dual variables. The second step for global update is implemented in the network manager for SDNV.

The rational stopping criterion from [42] is introduced as the stopping criterion for our algorithm. It has been demonstrated that the solution satisfies the second dual feasibility condition after the Lagrange multipliers updating. The second dual feasibility condition is obtained by getting gradients with respect to global variables. In fact, the first dual feasibility and primal feasibility actually do not hold. However, the first dual residual and primal residual are able to converge to zero, which indicates the first dual feasibility and the primal feasibility are achieved when t towards infinity.

In previously problem transforming, we have relaxed binary variables to continuous variables. After the ADMM-based optimization, we intend to recover the association indicator $\boldsymbol{a}$ from the optimal continuous variables to binary variables. This procedure is computing the marginal benefit for BS n on user k [45]. The binary association indicator can be recovered as

$$a_{k,n}^{*} = \begin{cases} 1, & \text{If} \quad Q_{k,n}=\max_k Q_{k,n},\, n\in\mathcal{N}_{bs} \\ 0, & \text{otherwise} \end{cases} \tag{34}$$

where $Q_{k,n}$ denotes the first partial derivation of $U_{k,n}$ with respect to $a_{k,n}$.

The convergence of ADMM algorithm has been proved in [43]. The computational complexity of the centralized algorithm with dual interior point method is $\mathcal{O}(((N+1)k)^i)$. For the proposed ADMM-based distributed resource allocation algorithm, BSs only need to deal the local optimization problem. Then, the computation complexity during local optimization at each BS is $\mathcal{O}(k^i)$, which is a polynomial time algorithm when $i > 1$, otherwise, a linear algorithm when $i = 1$. After local copy updating, SDN controller need to gather all the results from each BS. Then, update global variables and dual variables during each iteration, the complexity of this step is $\mathcal{O}((N+1)k)$ for both global variables and dual variables. Assuming P stands for the number of iterations needed for the algorithm convergence, we summarize the overall time complexity of the proposed distributed algorithm as $P(\mathcal{O}((N+1)k^i+2(N+1)k))$. It will be shown in simulation, the number of iterations for algorithm convergence is not large. The time complexity of the proposed distributed algorithm is significantly smaller rather than the centralized algorithm.

VI. SIMULATION RESULTS AND DISCUSSIONS

We implement a SDNV cellular network in a field of size $1000m \times 1000m$, where there are three InPs, each consists a BS with a MEC server implemented. The computing and storage capability of MEC server is limited which is practical. The location of BS is fixed, and several users around follow the random uniform distributions. Among these users, we assume there are a few users support user virtualization for computation offloading. We adopt certain models for the wireless channels in the proposed SDNV cellular network [46]. The fading channels are assumed exhibit Rayleigh fading, the channel coefficients are distributed as $\mathcal{CN}(0, 1/(1+d)^{\gamma})$ with a path loss exponent $\gamma = 4$, where d denotes the distance. The noise equals to the additive Gaussian noise. The proposed ADMM-based distributed resource allocation algorithm has been implemented at the SDN controller for all the users attributed to different MVNOs and InPs. In order to highlight our work, we assume that each UE has subscribed one content from content provider at the core network, and all the contents can be delivered to the cellular network.

In this paper, we evaluate the performance of the proposed distributed virtual resource allocation algorithm by simulation results and study the impact of the following parameters: 1) the number of users; 2) the number of virtualized users; 3) the average required data rate per task; 4) the average size of computation tasks; 5) the computation capability of MEC servers and 6) the self-interference cancellation factor. We employ the total utility of the MVNOs to measure the performance of the proposed algorithm. Meanwhile, for performance comparison, the other two algorithms are also evaluated. These algorithms are listed as follows:

1) the centralized algorithm with user virtualization, in which the global CSI from all users is collected and the resource allocation executes in a centralized manner;

2) the centralized algorithm without user virtualization, in which computation tasks are only allowed to directly offloaded to the BS.

TABLE 1. Simulation parameters.

Parameters	Value
Bandwidth	10MHz
Transmission power	20dBm
Self interference cancelling	110dB
PSD of AWGN	-174dBm/Hz
Numbers of users	[10,40]
Memory size per MEC server	100Mb
Communication revenue α_n	10 units/Kbps
Communication fee from users β_n	3 units/KHz
Computation revenue ϕ_n	10 units/Kb
Computation fee from MVNOsψ_n	40^{-6} units/J
Storage fee of computation tasksδ_n	30 units/Mb

Our simulation is implemented on the MATLAB, and the Monte Carlo simulation is used for evaluating the performance of the proposed algorithm. The locations of users are changed in each simulation loop by the Monte Carlo simulation. After a number of loops, the average value is obtained to lessen the randomness effects. Simulation parameters of our work are given in Table 1.

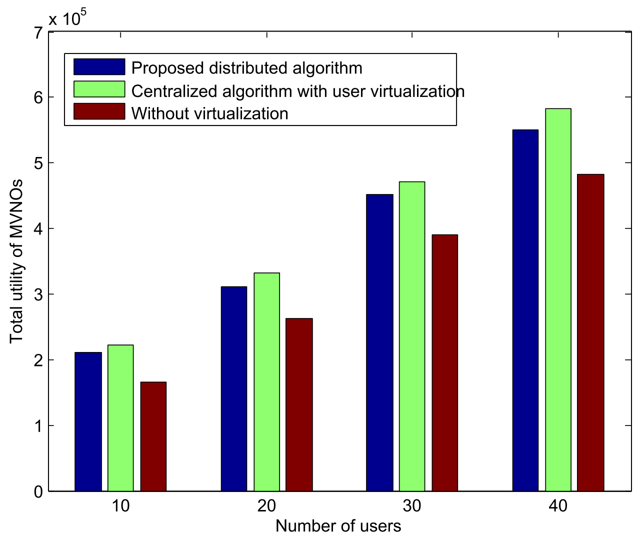

FIGURE 2. Total utility of MVNOs under different number of users.

Fig. 2 represents the impact of the number of users on the performance of different algorithms. In this scenario, users are attributed to 4MVNOs, and there are 3 InPs. The average data rate requirement is 1Mb/s, and the SIC factor of FD transmission is 10^{-11}. As shown in Fig. 2, the total utility obtained by three algorithms increase with the growth of the number of users. This is because that a network incorporating more receivers will introduce multiuser diversity gain. Note that in the conventional cellular network without user virtualization, the users are only available to offload the computation tasks directly to the BS. Thus, centralized algorithm without user virtualization will result in the least utility compared with the other two algorithms. The proposed algorithm achieves almost same utility compared to the optimal centralized algorithm, the dual residual cause to the gap between these two algorithms. However, the distributed algorithm can significantly reduce the computational complexity and signaling overhead without gathering the global CSI.

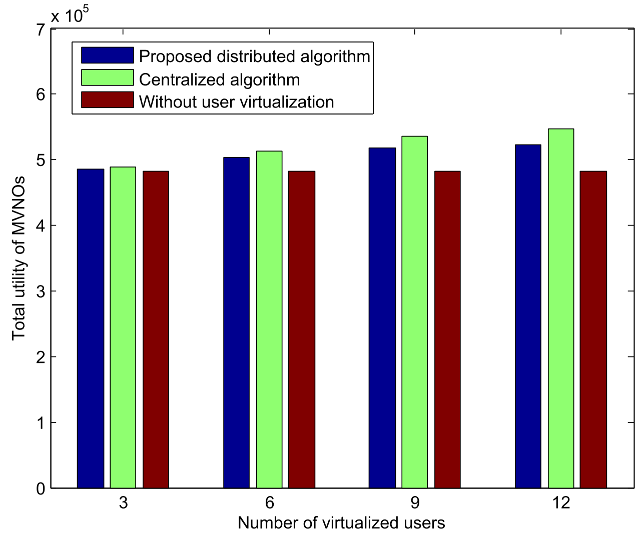

FIGURE 3. Total utility of MVNOs via different number of virtualized users.

In order to evaluate the performance gain from the proposed user virtualization, we compared the system utility under different setting of the virtualized user numbers in Fig. 3. In following simulations, we set the same basic scenario with Fig. 2, that there are 4MVNOs and 3 InPs. The total number of users is 30. The average data rate requirement is 1Mbps, and the SIC factor of FD transmission is 10^{-11}. As shown in Fig. 3, the total utility obtained by centralized algorithm and the proposed distributed algorithm increase with the growth of the number of virtualized users. This is because that more virtualized users lead to better association choices for offloading, thus the communication cost is reduced. The conventional cellular network without user virtualization archives lower utility than other two algorithms due to the users are only available to offload the computation tasks directly to the BS.

In Fig. 4, we compare the total utility of MVNOs at different average date rate requirements per tasks. The computation capability of each MEC server is 10GHz, the average size of compuation task is 1Mb, and the SIC factor of FD transmission is 10^{-11}. The total utility obtained by two algorithms declines when the average required data rate

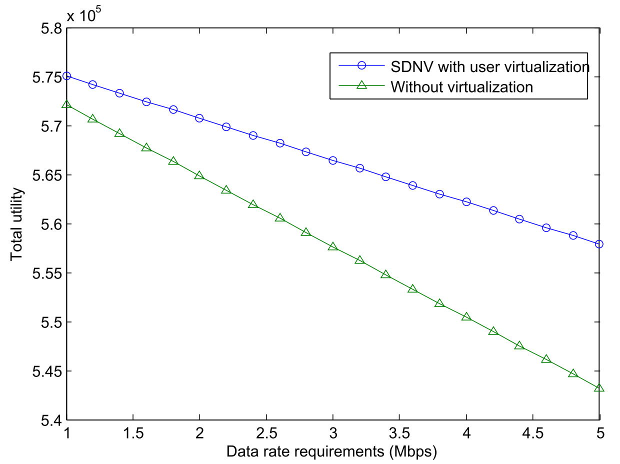

FIGURE 4. Performance evaluation under different parameter setting of average date rate requirements.

increases. The reason for this trend is that larger average data requirement will cost more for communication resources. It should be noted that our proposed algorithm achieves better performance than conventional cellular network without user virtualization, because the communication capability of our framework is larger.

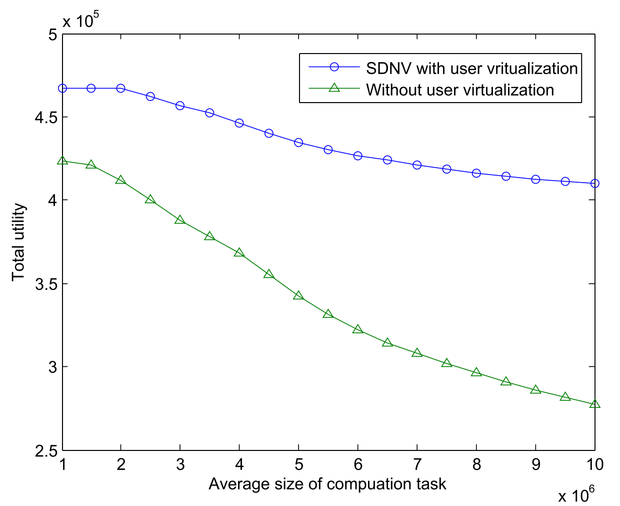

FIGURE 5. Performance evaluation under different parameter setting of average size of computation tasks.

In Fig. 5, the behavior of two algorithms at different average size of computation task is investigated, where the users number is 30, and computation capability of each MEC server is 10GHz. The SIC factor of FD transmission is 10^{-11}. The total utility obtained by two algorithms declines when the average size of computation task increases. That is because computation tasks occupy the memory of MEC servers which cause to more caching fee for MVNOs, besides, some users are incapable of accessing the cost-optimal MEC servers. In Fig. 5, the proposed SDNV framework performs better than the network without user virtualization due to the larger revenue from communication.

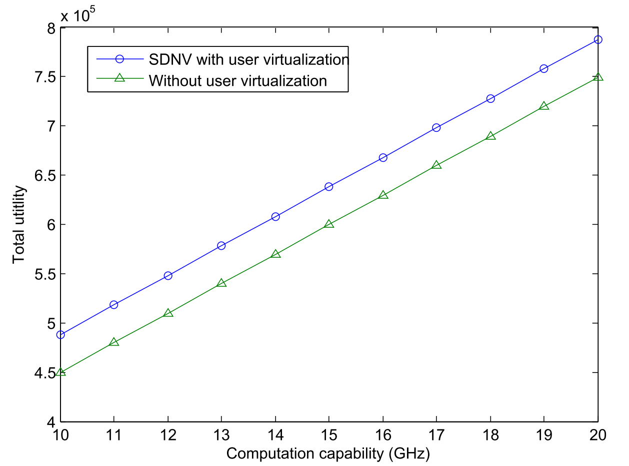

FIGURE 6. Performance evaluation under different parameter setting of computation capability of MEC servers.

In Fig. 6, we change the computation capability of each MEC server to study the total utility of MVNOs respect to an increasing computation capability, where the total users is 40. The average data rate requirements is 1Mbps, and the average size of compuation tasks is 1Mb. The SIC factor of FD transmission is 10^{-11}. The total utility obtained by different framework increases when the computation capability rising from 10GHz to 20GHz. Our proposed algorithm achieves better performance than conventional cellular network due to the communication capability of our framework is higher.

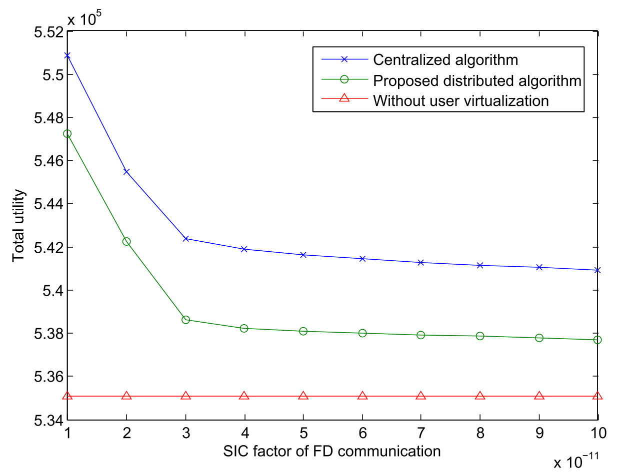

FIGURE 7. Performance evaluation under different parameter setting of.

Fig. 7 shows the impact of the SIC factor on the performance of different algorithms. In this scenario, the user number is set to 40, and the computation capability of each MEC server is 10GHz. The average size of data rate requirement is set 1Mbps. The average size of compuation tasks is 1Mb, and the SIC factor of FD transmission is 10^{-11}. In Fig. 7, it can be observed that the total utility decrease when the SIC factor of FD transmission increases. This is because the

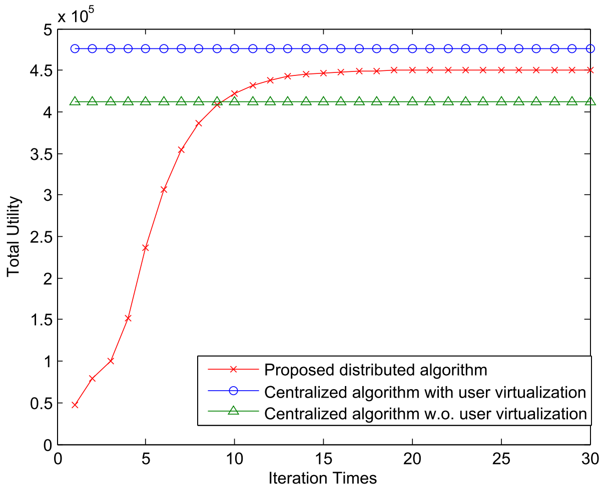

FIGURE 8. Convergence performance.

incremental of SIC factor has weakened the cancellation of the self-interference, more self-interference will impact the communication link.

We study the convergence behavior of the the proposed distributed resource allocation algorithm in Fig. 8. the computation capability of each MEC server is 10GHz,and the average size of data rate requirement is set 1Mbps. The average size of compuation task is 1Mb, and the SIC factor of FD transmission is 10^{-11}. We initialize the start point with zero. It can be observed that our algorithm converges almost the same value as the centralized algorithm, which gather the global CSI with a huge amount signaling overhead. Furthermore, our proposed distributed algorithm performs better compared to the centralized algorithm without user virtualization.

VII. CONCLUSION

In this paper, we proposed a novel MEC framework with user virtualization scheme in SDNV cellular network. The user virtualization scheme assisted by FD communication was introduced which could extend the virtual resource of virtualized networks without extra memory occupation and additional delay augment. We formulated the virtual resource allocation for MEC as a joint optimization problem. An ADMM-based distributed resource allocation algorithm was proposed which can significantly reduce the computational complexity and signaling overhead. Simulation results indicated that the total utility of MVNOs was significantly improved which benefited from the user virtualization, and the proposed distributed algorithm achieved almost same results compared to the centralized algorithm.

REFERENCES

[1] L. Wei, J. Cai, C. H. Foh, and B. He, "QoS-aware resource allocation for video transcoding in clouds," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 1, pp. 49–61, Jan. 2017.

[2] H. T. Dinh, C. Lee, D. Niyato, and P. Wang, "A survey of mobile cloud computing: Architecture, applications, and approaches," *Wireless Commun. Mobile Comput.*, vol. 13, no. 18, pp. 1587–1611, Dec. 2013.

[3] S. Abolfazli, Z. Sanaei, E. Ahmed, A. Gani, and R. Buyya, "Cloud-based augmentation for mobile devices: Motivation, taxonomies, and open challenges," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 1, pp. 337–368, 1st Quart., 2014.

[4] ETSI, "Mobile-edge computing: Introductory technical white paper," ETSI, White Paper, Sep. 2014, no. 11.

[5] X. Sun and N. Ansari, "EdgeIoT: Mobile edge computing for the Internet of Things," *IEEE Commun. Mag.*, vol. 54, no. 12, pp. 22–29, Dec. 2016.

[6] T. X. Tran, A. Hajisami, P. Pandey, and D. Pompili, "Collaborative mobile edge computing in 5G networks: New paradigms, scenarios, and challenges," *IEEE Commun. Mag.*, vol. 55, no. 4, pp. 54–61, Apr. 2017.

[7] X. Chen, L. Jiao, W. Li, and X. Fu, "Efficient multi-user computation offloading for mobile-edge cloud computing," *IEEE/ACM Trans. Netw.*, vol. 24, no. 5, pp. 2795–2808, Oct. 2016.

[8] S. Deng, L. Huang, J. Taheri, and A. Y. Zomaya, "Computation offloading for service workflow in mobile cloud computing," *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 12, pp. 3317–3329, Dec. 2015.

[9] D. Kreutz, F. Ramos, P. E. Veríssimo, C. E. Rothenberg, S. Azodolmolky, and S. Uhlig, "Software-defined networking: A comprehensive survey," *Proc. IEEE*, vol. 103, no. 1, pp. 14–76, Jan. 2015.

[10] N. M. M. K. Chowdhury and R. Boutaba, "A survey of network virtualization," *Comput. Netw.*, vol. 54, no. 5, pp. 862–876, Apr. 2010.

[11] Q. Duan, N. Ansari, and M. Toy, "Software-defined network virtualization: An architectural framework for integrating SDN and NFV for service provisioning in future networks," *IEEE Netw.*, vol. 30, no. 5, pp. 10–16, Sep./Oct. 2016.

[12] D. Bharadia, E. McMilin, and S. Katti, "Full duplex radio," in *Proc. ACM SIGCOMM*, Aug. 2013, pp. 375–386.

[13] L. Jiménez Rodríguez, N. H. Tran, and T. Le-Ngoc, "Performance of full-duplex AF relaying in the presence of residual self-interference," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1752–1764, Sep. 2014.

[14] X. Zhang, W. Cheng, and H. Zhang, "Full-duplex transmission in phy and mac layers for 5G mobile wireless networks," *IEEE Wireless Commun.*, vol. 22, no. 5, pp. 112–121, Oct. 2015.

[15] Y. Liu, Y. Dai, X.-G. Xia, C. Kang, and H. Zhang, "SC-FDE based full-duplex relay communication robust to residual loop interference," *IEEE Wireless Commun. Lett.*, vol. 6, no. 4, pp. 538–541, Aug. 2017.

[16] L. Mamatas, S. Clayman, and A. Galis, "A service-aware virtualized software-defined infrastructure," *IEEE Commun. Mag.*, vol. 53, no. 4, pp. 166–174, Apr. 2015.

[17] F. Hu, Q. Hao, and K. Bao, "A survey on software-defined network and OpenFlow: From concept to implementation," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 4, pp. 2181–2206, 4th Quart., 2014.

[18] M. Casado *et al.*, "Software-defined Internet architecture: Decoupling architecture from infrastructure," in *Proc. 11th ACM Workshop Hot Topics Netw.*, Oct. 2012, pp. 43–48.

[19] M. Casado, T. Koponen, S. Shenker, and A. Tootoonchian, "Fabric: A retrospective on evolving SDN," in *Proc. 1st Workshop Hop Topics Softw.-Defined Netw.*, Aug. 2012, pp. 85–90.

[20] "Network functions virtualisation introductory white paper," ETSI, Sophia Antipolis, France, Tech. Rep., Oct. 2012.

[21] C. Liang and F. R. Yu, "Wireless network virtualization: A survey, some research issues and challenges," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 358–380, 1st Quart., 2014.

[22] K. Zhu and E. Hossain, "Virtualization of 5G cellular networks as a hierarchical combinatorial auction," *IEEE Trans. Mobile Comput.*, vol. 15, no. 10, pp. 2640–2654, Oct. 2016.

[23] X. Costa-Perez, J. Swetina, T. Guo, R. Mahindra, and S. Rangarajan, "Radio access network virtualization for future mobile carrier networks," *IEEE Commun. Mag.*, vol. 51, no. 7, pp. 27–35, Jul. 2013.

[24] D. Drutskoy, E. Keller, and J. Rexford, "Scalable network virtualization in software-defined networks," *IEEE Internet Comput.*, vol. 17, no. 2, pp. 20–27, Mar./Apr. 2013.

[25] J. Matias, J. Garay, N. Toledo, J. Unzilla, and E. Jacob, "Toward an SDN-enabled NFV architecture," *IEEE Commun. Mag.*, vol. 53, no. 4, pp. 187–193, Apr. 2015.

[26] J.-L. Chen, Y.-W. Ma, H.-Y. Kuo, C.-S. Yang, and W.-C. Hung, "Software-defined network virtualization platform for enterprise network resource management," *IEEE Trans. Emerg. Topics Comput.*, vol. 4, no. 2, pp. 179–186, Apr./Jun. 2016.

[27] "ETSI GS NFV-EVE 005 V1.1.1 (2015-12) group specification network functions virtualisation (NFV); Ecosystem; report on SDN usage in NFV architectural framework," ETSI, Sophia Antipolis, France, Tech. Rep., Oct. 2015

[28] N. Abbas, Y. Zhang, A. Taherkordi, and T. Skeie, "Mobile edge computing: A survey," *IEEE Internet Things J.*, to be published.
[29] Y. Mao, C. You, J. Zhang, K. Huang, and K. B. Letaief, "A survey on mobile edge computing: The communication perspective," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 4, pp. 2322–2358, 4th Quart., 2017.
[30] J. O. Fajardo, I. Taboada, and F. Liberal, "Improving content delivery efficiency through multi-layer mobile edge adaptation," *IEEE Netw.*, vol. 29, no. 6, pp. 40–46, Nov./Dec. 2015.
[31] X. Chen, "Decentralized computation offloading game for mobile cloud computing," *IEEE Trans. Parallel Distrib. Syst.*, vol. 26, no. 4, pp. 974–983, Apr. 2015.
[32] X. Wang, S. Leng, and K. Yang, "Social-aware edge caching in fog radio access networks," *IEEE Access*, vol. 5, pp. 8492–8501, 2017.
[33] C. Wang, F. R. Yu, Q. Chen, and L. Tang, "Joint computation and radio resource management for cellular networks with mobile edge computing," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–6.
[34] G. Qiao, S. Leng, K. Zhang, and K. Yang, "Joint deployment and mobility management of energy harvesting small cells in heterogeneous networks," *IEEE Access*, vol. 5, pp. 183–196, 2017.
[35] K. Zhang *et al.*, "Energy-efficient offloading for mobile edge computing in 5G heterogeneous networks," *IEEE Access*, vol. 4, pp. 5896–5907, 2016.
[36] J. Marašević, J. Zhou, H. Krishnaswamy, Y. Zhong, and G. Zussman, "Resource allocation and rate gains in practical full-duplex systems," *IEEE/ACM Trans. Netw.*, vol. 25, no. 1, pp. 292–305, Feb. 2017.
[37] W. Yuan and K. Nahrstedt, "Energy-efficient CPU scheduling for multimedia applications," *ACM Trans. Comput. Syst.*, vol. 24, no. 3, pp. 292–331, Aug. 2006.
[38] T. D. Burd and R. W. Brodersen, "Processor design for portable systems," *J. VLSI Signal Process. Syst.*, vol. 13, nos. 2–3, pp. 203–221, Aug. 1996.
[39] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[40] A. Schrijver, *Theory of Linear and Integer Programming*. Hoboken, NJ, USA: Wiley, 1998.
[41] D. Fooladivanda and C. Rosenberg, "Joint resource allocation and user association for heterogeneous wireless cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 1, pp. 248–257, Jan. 2013.
[42] P. Xue, P. Gong, J. H. Park, D. Park, and D. K. Kim, "Radio resource management with proportional rate constraint in the heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 11, no. 3, pp. 1066–1075, Mar. 2012.
[43] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.
[44] M. Leinonen, M. Codreanu, and M. Juntti, "Distributed joint resource and routing optimization in wireless sensor networks via alternating direction method of multipliers," *IEEE Trans. Wireless Commun.*, vol. 12, no. 11, pp. 5454–5467, Nov. 2013.
[45] G. Liu, H. Ji, F. R. Yu, Y. Li, and R. Xie, "Energy-efficient resource allocation in full-duplex relaying networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2014, pp. 2400–2405.
[46] *Evolved Universal Terrestrial Radio Access (E-UTRA); Further Advancements for E-UTRA Physical Layer Aspects (Release 9)*, document 3GPP TS 36.814, 3rd Generation Partnership Project, 2012. [Online]. Available: http://www.3gpp.org/ftp/

MING LIU received the B.Eng. degree from Zhengzhou University. He is currently pursuing the Ph.D. degree with the School of Information and Communication Engineering, University of Electronic Science and Technology of China. His research interests include full-duplex communications, radio resource allocation, next generation mobile networks, and mobile edge computing.

YUMING MAO is currently a Professor with the School of Information and Communication Engineering, University of Electronic Science and Technology of China, Chengdu, where he is a Chair of the Department of Network Engineering. His main research interests include broadband communication networks, network organization and protocol analysis, transmission control protocol/Internet protocol technologies, network management and protocol, routing protocol, and network engineering. Prof. Mao received several awards, including the First-Grade, Second-Grade, and Third-Grade Awards of the Ministry of Electronic Industry for science and technology progress and the Second-Grade National Award for science and technology progress.

SUPENG LENG received the Ph.D. degree from Nanyang Technological University (NTU), Singapore. He was a Research Fellow with the Network Technology Research Center, NTU. He is currently a Professor with the School of Information and Communication Engineering, University of Electronic Science and Technology of China. He has authored or co-authored over 150 research papers in recent years. His research focuses on resource, spectrum, energy, routing, and networking in Internet of things, vehicular networks, broadband wireless access networks, smart grid, and the next generation mobile networks. Prof. Leng serves as an Organizing Committee Chair and a TPC Member for many international conferences, as well as a reviewer for over ten international research journals.

SUN MAO is currently pursuing the Ph.D. degree with the School of Information and Communication Engineering, University of Electronic Science and Technology of China. His research interests include simultaneous wireless information and power transfer, convex optimization, next generation mobile networks, and mobile edge computing.