Dynamic Network Slicing and Resource Allocation in Mobile Edge Computing Systems

Jie Feng, Qingqi Pei, F. Richard Yu, *Fellow*, IEEE, Xiaoli Chu, *Member*, IEEE, Jianbo Du, and Li Zhu

***Abstract*—The application of network slicing to mobile edge computing (MEC) systems has aroused great interests from both academia and industry. However, the optimization of network slicing and MEC in most existing research works only focuses on resource slicing, energy scheduling, and power allocation from the perspective of mobile devices, without considering the operator's revenue. In this paper, we propose a novel framework for network slicing in MEC systems, including slice request admission and a revenue model, to investigate the operator's revenue escalation problem while considering traffic variations. The revenue model is mainly composed of the longer-term revenue and short-term revenue. Particularly, we jointly optimize slice request admission in the long-term and resource allocation in the short-term to maximize the operator's average revenue. By employing the Lyapunov optimization technique, we develop an algorithm without requiring any prior-knowledge of traffic distributions, referred to as the DNSRA, to solve the problem. To reduce the computational complexity of directly solving the DNSRA, we decouple the optimization variables for efficient algorithm design. By this, the strategies for user association and CPU-cycle frequency are obtained in closed forms, respectively. Power allocation and subcarriers assignment are obtained by employing the successive convex approximation approach. Meanwhile, we develop a heuristic algorithm to obtain the dynamic slice request admission decision. Simulation results show that the proposed DNSRA can strike a flexible balance between the average revenue and the average delay, and can significantly increase the operator's revenue against existing schemes.**

***Index Terms*—Mobile edge computing (MEC), network slicing, traffic variations, operator's revenue, resource allocation.**

I. Introduction

In recent years, it is expected that the next generation of wireless communication networks can support various application scenarios with different requirements. However, it is not practical to deploy a separate communication network for each application scenario. *Network slicing* has been envisioned as a promising approach to solve this issue [1]–[3]. The main feature of network slicing is the running of multiple logically separate networks, which are independent of each other, on top of a common shared physical infrastructure [4]. Through network slicing, network resources can be flexibly and dynamically allocated to logical network slices according to customized service requirements on-demand [5]–[8]. With the rapid development of network edge applications such as the Internet of Things (IoT) and cyber-physical systems, the service requirements of these applications have also been changing [9]. However, the existing works have not been able to meet the various changing performance requirements of these applications [10]. Therefore, how to efficiently and simultaneously support these applications in a shared physical infrastructure is still an unaddressed problem [11].

Mobile edge computing (MEC), which supports not only computing but also communications and storage, is a promising technology for supporting edge services with specific requirements [12]–[15]. Different from conventional cloud computing systems, MEC is close to the edge of networks, which can improve the quality of user experience, including decreased latency and reduced energy consumption [16]. The conjunction of network slicing and MEC has aroused great interests in academia and industry. In [11], the authors proposed a new dynamic network slicing architecture for large-scale energy-harvesting fog computing networks to maximize the utilization efficiency of available resources while balancing the workloads among fog nodes. Sun *et al.* [17] studied a hierarchical radio resource allocation for network slicing in fog radio access networks (F-RAN). Zhao *et al.* [18] studied the problem of network slicing resource allocation in MEC systems but without considering the computational resource. Sanguanpuak *et al.* [19] solved the slice allocation problem of multiple mobile network operators (MNOs) with the goal of maximizing the social welfare of the network.

Most existing works on the application of network slicing to MEC systems have not considered the dynamic demand of services, e.g., the frequency of request [20]. Since network resources are limited, it is necessary to properly determine slice request admission, which has a significant impact on the quality of service (QoS) of users. In addition, most existing works centre around resource slicing, energy scheduling, and power allocation from the perspective of mobile users, regardless of the operator's revenue. Meanwhile, spatial and temporal traffic variations, both of which would dramatically affect resource allocation and thus the operators revenue, have

*This work is supported by the National Key Research and Development Program of China under Grant 2018YFE0126000, the Key Program of NSFC-Tongyong Union Foundation under Grant U1636209, the National Natural Science Foundation of China under Grant 61902292, and the Key Research and Development Programs of Shaanxi under Grant 2019ZDLGY13-07 and 2019ZDLGY13-04, the Natural Science Foundation of China under Grant 61901367, and Beijing Wuzi University Youth Research Fund Project (No. 2017XJQN07). (*Corresponding authors: Qingqi Pei*)

J. Feng, and Q. Pei are with State Key Laboratory of ISN, Xidian University, No.2 Taibainan-lu, Xi'an, 710071, Shaanxi, China (e-mail: jiefengcl@163.com; qqpei@mail.xidian.edu.cn).

F. R. Yu is with the Department of Systems and Computer Engineering, Carleton University, Ottawa, ON, K1S 5B6, Canada (e-mail: RichardYu@cunet.carleton.ca).

X. Chu is with the Department of Electronic and Electrical Engineering, University of Sheffield, Sheffield S1 3JD, U.K. (e-mail: x.chu@sheffield.ac.uk).

J. Du is with Shaanxi Key Laboratory of Information Communication Network and Security, Xian University of Posts and Telecommunications, Xian 710121, China (e-mail: dujianboo@163.com).

Li Zhu is with Beijing Jiaotong University, Beijing, 100044,China (e-mail: zhulibjtu@gmail.com).

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2020.2992607, IEEE Transactions on Vehicular Technology

not been explicitly incorporated into the formulation.

Motivated by the above, in this paper, we study the problem of dynamic network slicing and resource allocation in MEC systems to maximize the operator's average revenue while considering traffic variations. More specifically, we consider a scenario where there are two slice requests, i.e., low-latency computation offloading slice request and high-rate data sharing slice request.

The main contributions of this work are summarized as follows.

- We develop a neoteric framework for network slicing in MEC systems to investigate the average revenue maximization problem, which consists of slice request admission and a revenue model. Since the resources of an MEC system are limited, a binary integer variable is introduced to decide whether to accept or reject a slice request. The operator's revenue consists of long term revenue and short term revenue.
- We formulate a stochastic optimization problem to jointly optimize the dynamic slice request admission, user association, CPU-cycle frequency, subcarrier assignment and power allocation while taking into account spatial and temporal variations of traffic. With the assistance of the Lyapunov optimization technique, we develop a **D**ynamic **N**etwork **S**licing and **R**esource **A**llocation algorithm (DNSRA) to solve the problem. The proposed DNSRA does not require any prior knowledge of channel information or traffic distributions.
- To tackle the highly coupled and mixed combinational subproblem in the DNSRA, we develop an efficient algorithm by decoupling optimization variables. By doing this, we develop exceedingly simple policies for user association and CPU-cycle frequency allocation, where both of them are obtained in closed forms. By using the successive convex approximation approach, we obtain power allocation and subcarrier assignment. Meanwhile, a heuristic algorithm is developed to acquire the dynamic slice request admission decision.
- Simulation results exhibit that the proposed algorithm converges fast, can flexibly achieve the revenue-delay tradeoff, and, more importantly, can improve the operator's revenue against existing schemes.

The rest of this paper is organized as follows. In Section II, we describe system model and problem formulation. In Section III, the DNSRA is devised, and we design algorithms for user association, subcarrier assignment, CPU-cycle frequency, and power allocation in the short-time slot, respectively. In addition, slice request admission and channel allocation are obtained in the long-time slot. The performance of the proposed algorithm is evaluated by the simulation in Section IV. Finally, we conclude this paper in Section V.

II. System Model

We consider a scenario where RAN slicing is built upon an MEC system, as shown in Fig. 1. There are B base stations (BS) in the system, in which each BS is integrated with an MEC server [21], [22]. Let $\mathcal{B} = \{1, 2, ...B\}$ denote the set

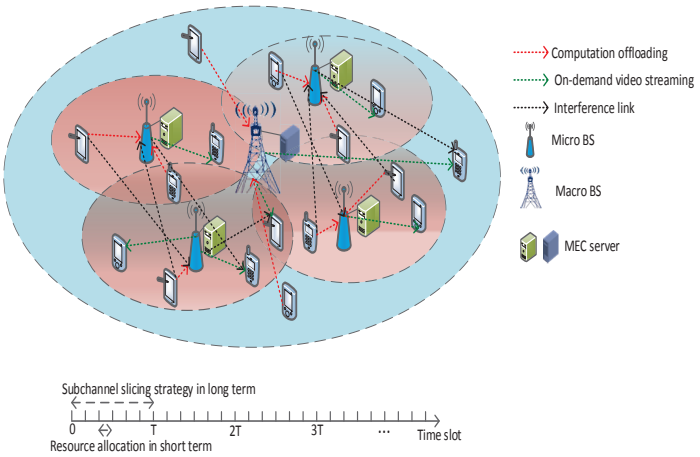


Fig. 1. The system scenario.

of BSs. The system is assumed to be operated in slots, and the length of the time slot is τ. Based on the time-slotted system, we consider two types of time slots. One is a long time slot (LTS), and the other is a short time slot (STS). In this paper, we consider that the system contains multiple LTSs. The length of the LTS is T, and we assume that each LTS contains p STSs, i.e., $T = p\tau$. At the beginning of each LTS, the network operator can decide whether to accept or reject the arrived network slice requests. Nevertheless, at the beginning of each STS, resource allocation policies are obtained. The two timescales system will be discussed in detail in the following sections of this section. Similar to [17], we consider two types of network slice requests in the MEC system, i.e., low-latency computation offloading slice (*SI*) and high-rate data sharing slice (*SII*). Our proposed algorithms can be adapted for multiple slice service requests with minor modifications.

A. Slicing Model

Based on [7], each network slice request is composed of two components, namely, the number of mobile devices and the QoS requirements. Since the number of mobile devices and QoS requirements usually vary with time and space, we describe the number of mobile devices and QoS requirements in each LTS, as follows.

1) **The number of mobile devices**. We denote the number of mobile devices in LTS t_k by $I^{SI}(t_k)$ in the computation offloading slice, and use $I^{SII}(t_k)$ to denote the number of mobile devices in LTS t_k in the data sharing slice.

2) **The QoS requirements**. The QoS requirement metrics for different network slices are different. For the computation offloading slice, we utilize the maximum latency as the QoS metric, which is denoted as $d^{SI}(t_k)$ in LTS t_k. For the data sharing slice, we are interested in the minimum data rate, which is denoted as $R^{SII}(t_k)$ in LTS t_k.

Based on the above, we utilize the two-tuple $\{I^{SI}(t_k), d^{SI}(t_k)\}$ and $\{I^{SII}(t_k), R^{SII}(t_k)\}$ to represent the slice request for the computation offloading slice and the data sharing slice, respectively. At the beginning of each LTS,

we update a binary variable $a_s(t_k)(s \in \{SI, SII\})$, where $a_s(t_k) = 1$ if a slice request is accepted by the operator and 0 otherwise. There are N subcarriers in the system. Assume that the system adopts the universal frequency reuse scheme, i.e., all BSs can use all the subcarriers for uplink and downlink data transmission, to improve spectral efficiency. We assume that the subchannels assigned to different users are orthogonal within a cell. To avoid the inter-slice interference and without loss of generality, the set of subcarriers allocated to the two slices in LTS t_k is denoted by $N^{SI}(t_k) = \{1, 2, ..., N_n\}$ and $N^{SII}(t_k) = \{N_{n+1}, N_{n+2}, ...N\}$ in LTS t_k, respectively. The bandwidth of each subcarrier is W.

It is worth noting that the time-slot network slicing system we consider involves two timescales, i.e., the LTS and the STS. We investigate a long-term average revenue of the operator, so multiple LTSs are considered. At the beginning of each LTS, the operator needs to decide whether to accept a slice request or not and also allocate the corresponding number of subcarriers for each slice in the LTS. In the whole LTS, variables $a_s(t_k)$, $N^{SI}(t_k)$, and $N^{SII}(t_k)$ remain unchanged till the next LTS. At the beginning of each STS, the resource allocation scheme is determined, and the short-term variables keep unchanged, which may change between adjacent STSs.

B. Computation Offloading Slice SI Model

In this work, we assume that the mobile devices in slice *SI* have relatively weak computing capability and the computing tasks need to be offloaded to an MEC server through uplink communication. We denote the set of mobile devices in slice *SI* by $I^{SI}(t_k) = \{1, 2, ..., M_1\}$. Suppose that the mobile devices are running independent and fine-grained tasks [23], [24], [25], and let $D_{i,SI}(t)$ denote the offloaded data size of mobile device i in STS t. We let $\boldsymbol{P}_{SI} = (P^n_{ib,SI})$ and $\boldsymbol{G}_{SI}(t) = (g^n_{ib,SI}(t))$ denote the transmit power and the channel gain that accounts for path loss, shadowing, and fading from mobile device i to BS b on subcarrier n in STS t, respectively. We assume that the transmit power of mobile devices is fixed for slice *SI*. Then, the transmit rate from mobile device i to BS b on subcarrier n in STS t is give by

$$R^n_{ib,SI}(t) = W \log_2(1 + \gamma^n_{ib,SI}(t)), \tag{1}$$

where $\gamma^n_{ib,SI}(t)$ denotes the signal-to-interference-plus-noise-ratio (SINR), given by

$$\gamma^n_{ib,SI}(t) = \frac{P^n_{ib,SI} g^n_{ib,SI}(t)}{\sum\limits_{c \in \mathcal{B}, c \neq b} \sum\limits_{j=1}^{M_1} \rho^n_{jc,SI}(t) P^n_{jc,SI} g^n_{j(i)c,SI}(t) + \sigma^2_i(t)}, \tag{2}$$

where $\sigma^2_i(t)$ is the noise power, $\boldsymbol{\rho}_1(t) = (\rho^n_{ib,SI}(t))$, and $\rho^n_{ib,SI}(t)$ is an indicator variable that is 1 if subcarrier n is allocated to mobile device i associated with BS b and 0 otherwise.

Let $\boldsymbol{y}_1(t) = (y_{ib,SI}(t))$ denote the user association variable, where $y_{ib,SI}(t) = 1$ indicates that mobile device i is associated with BS b in STS t, otherwise $y_{ib,SI}(t) = 0$. The sum transmit rate and the total transmit power from mobile device i to BS b are respectively given by

$$R_{ib,SI}(t) = y_{ib,SI}(t) \sum_{n=1}^{|N^{SI}(t_k)|} \rho^n_{ib,SI}(t) R^n_{ib,SI}(t), \tag{3}$$

$$P_{ib,SI}(t) = y_{ib,SI}(t) \sum_{n=1}^{|N^{SI}(t_k)|} \rho^n_{ib,SI}(t) P^n_{ib,SI}. \tag{4}$$

Throughout the paper, we assume that each mobile device in STS t is served by only one BS. Thus, the sum transmit rate and total power consumption of the uplink communication are respectively given by

$$R_{i,SI}(t) = \sum_{b \in \mathcal{B}} y_{ib,SI}(t) \sum_{n=1}^{|N^{SI}(t_k)|} \rho^n_{ib,SI}(t) R^n_{ib,SI}(t), \tag{5}$$

$$P_{i,SI}(t) = \sum_{b \in \mathcal{B}} y_{ib,SI}(t) \sum_{n=1}^{|N^{SI}(t_k)|} \rho^n_{ib,SI}(t) P^n_{ib,SI}. \tag{6}$$

Accordingly, the transmission delay is expressed as

$$\delta_{i,SI}(t) = \frac{D_{i,SI}(t)}{R_{i,SI}(t)}. \tag{7}$$

Network slicing integrates resource of multiple BSs, including radio, computational, and storage resources, to provide customized services for mobile users. Computing tasks arrive randomly at every STS and are queued in their respective network slices. Let $\boldsymbol{A}_{SI}(t) = (A_{i,SI}(t))$ be the process of random tasks arrivals, where $A_{i,SI}(t)$ is the number of computing tasks of mobile device i that arrive to slice *SI* on STS t. For simplicity, we assume that $\boldsymbol{A}_{SI}(t)$ is independent and identically distributed (i.i.d.), over slots with arrival rate $\boldsymbol{\lambda}_1$, so that $E\{\boldsymbol{A}_{SI}(t)\} = \boldsymbol{\lambda}_1$ for all t. Furthermore, let $Q_{SI}(t)$ denote the number of computational tasks currently stored in slice *SI*. Then, the dynamics of the offloading queue for slice *SI* can be expressed as

$$Q_{SI}(t+1) = \max\Big\{Q_{SI}(t) - a_s(t) \sum_{i \in I^{SI}(t)} R_{i,SI}(t)\tau, 0\Big\} + a_s(t) \sum_{i \in I^{SI}(t)} A_{i,SI}(t). \tag{8}$$

When the computational tasks are offloaded to the MEC server, we assume that the clock speed of the MEC server b that executes the computational task of the mobile device i in STS t is denoted as $f_{ib,SI}(t)$ (in CPU cycles/s) using the Dynamic Voltage and Frequency Scaling (DVFS) scheme. Therefore, the time when the MEC server b processes the offloaded data of mobile device i is given by

$$\upsilon_{i,SI}(t) = \frac{D_{i,SI}(t) L_i}{\sum\limits_{b \in \mathcal{B}} y_{ib,SI}(t) f_{ib,SI}(t)}, \tag{9}$$

where L_i denotes the processing density (in CPU cycles/bit). The total delay of mobile device i is given by

$$T^{SI}_i(t) = \delta_{i,SI}(t) + \upsilon_{i,SI}(t). \tag{10}$$

To meet the performance requirement of mobile devices, the delay constraint of mobile devices in the slice *SI* in slot t is given by

$$T_i^{SI}(t) \leq d^{SI}(t_k),\ \forall k, t \in [t_k, t_k + T], i \in I^{SI}(t_k). \quad (11)$$

Accordingly, the power consumption of the MEC server b processing the offloaded tasks of mobile device i can be expressed as

$$P_{ib,SI}(t) = k_{ser,b} f_{ib,SI}^3(t), \quad (12)$$

where $k_{ser,b}$ is the effective switched capacitance of the MEC server. The total power consumption of MEC servers in slice *SI* is given by

$$P_{SI}(t) = \sum_{i \in I^{SI}(t)} \sum_{b \in \mathcal{B}} y_{ib,SI}(t) P_{ib,SI}(t). \quad (13)$$

The computing resources of all MEC servers are integrated, and the operator manages resources in a unified manner and allocates resources to the slice *SI*. Therefore, the slice *SI* maintains a task buffer for each mobile device to store the computational tasks that have been offloaded but not yet executed, which is assumed to with sufficiently large capacity. Then, the dynamics of processing queue in slice *SI* can be expressed as

$$F_{SI}(t+1) = \max\Big\{F_{SI}(t) - a_s(t) \sum_{b \in \mathcal{B}} \sum_{i \in I^{SI}(t)} f_{ib,SI}(t) y_{ib,SI}(t), 0\Big\} + \min L_i\{a_s(t) \sum_{i \in I^{SI}(t_k)} R_{i,SI}(t), Q_{SI}(t)\}. \quad (14)$$

We assume that the size of the output data is small enough so that the return time can be ignored [26].

C. Data sharing Slice SII Model

We consider a data sharing slice, in which the data desired for each mobile device can be shared among all MEC servers, and the desired data required by mobile devices can be downloaded from any MEC server through downlink communication. We assume that mobile devices in slice *SII* utilize channels assigned to the slice to receive data, and each subchannel can be reused by mobile devices in different BSs. We denote the set of mobile devices in slice *SII* by $I^{SII}(t_k) = \{1, 2, ..., M_2\}$. Let $\boldsymbol{P}'_{SII}(t) = (P^n_{bi,SII}(t))$ and $\boldsymbol{G}_{SII}(t) = (g^n_{bi,SII}(t))$ denote the transmit power and the channel gain from BS b to mobile device i on subcarrier n in STS t, respectively. Then, the transmit rate from BS b to mobile device i on subcarrier n in STS t is given by

$$R^n_{bi,SII}(t) = W \log_2(1 + \gamma^n_{bi,SII}(t)), \quad (15)$$

where $\gamma^n_{bi,SII}(t)$ is given by

$$\gamma^n_{bi,SII}(t) = \frac{P^n_{bi,SII}(t) g^n_{bi,SII}(t)}{\sum_{c \in \mathcal{B}, c \neq b} \sum_{j=1}^{M_2} \rho^n_{cj,SII}(t) P^n_{cj,SII}(t) g^n_{cj(i),SII}(t) + \sigma_i^2(t)}, \quad (16)$$

where $\boldsymbol{\rho}_2(t) = (\rho^n_{cj,SII}(t))$ is the subcarrier variable in slice *SII*.

Accordingly, the sum transmit rate and the total transmit power are respectively given by

$$R_{i,SII}(t) = \sum_{b \in \mathcal{B}} y_{bi,SII}(t) \sum_{n=1}^{|N^{SII}(t_k)|} \rho^n_{bi,SII}(t) R^n_{bi,SII}(t), \quad (17)$$

$$P_{i,SII}(t) = \sum_{b \in \mathcal{B}} y_{bi,SII}(t) \sum_{n=1}^{|N^{SII}(t_k)|} \rho^n_{bi,SII}(t) P^n_{bi,SII}(t), \quad (18)$$

where $\boldsymbol{y}_2(t) = (y_{bi,SII}(t))$ is the user association variable in slice *SII*.

For slice *SII*, the rate requirement of mobile devices in STS t needs to meet the following condition.

$$R_{i,SII}(t) \geq R^{SII}(t_k),\ \forall k, t \in [t_k, t_k + T - 1], i \in I^{SII}(t_k). \quad (19)$$

The total power consumption of the slice *SII* in STS t is given by

$$P_{SII}(t) = \sum_{i \in I^{SII}(t)} P_{i,SII}(t_k). \quad (20)$$

Let $\boldsymbol{A}_{SII}(t) = (A_{i,SII}(t))$ denote the process of random data arrivals in slice *SII*, where $A_{i,SII}(t)$ is the amount of new data that arrives in mobile device i on STS t. Similarly, we assume that $\boldsymbol{A}_{SII}(t)$ is i.i.d. over slots with arrival rate $\boldsymbol{\lambda}_2$, so that $E\{\boldsymbol{A}_{SII}(t)\} = \boldsymbol{\lambda}_2$ for all t. Besides, $Q_{SII}(t)$ denotes the number of data current stored in slice *SII*. Then, the dynamics of the data queue for slice *SII* can be expressed as.

$$Q_{SII}(t+1) = \max\Big\{Q_{SII}(t) - a_s(t) \sum_{i \in I^{SII}(t_k)} R_{i,SII}(t)\tau, 0\Big\} + a_s(t) \sum_{i \in I^{SII}(t_k)} A_{i,SII}(t). \quad (21)$$

D. Revenue Model

The main purpose of this paper is to maximize the average revenue of the operator. The revenue of the operator is mainly composed of the longer-term revenue and short-term revenue.

1) The long-term revenue is related to the parameters in the network slicing request. Let $G^{SI}(I^{SI}(t_k), d^{SI}(t_k))$ and $G^{SII}(I^{SII}(t_k), R^{SII}(t_k))$ denote the long-term revenue for the computation offloading slice and the data sharing slice, respectively. Similar to [7], the long-term revenue of the two slices are respectively given by

$$G^{SI}(I^{SI}(t_k), d^{SI}(t_k)) = \frac{\tilde{a} I^{SI}(t_k)}{1 - e^{(-d^{SI}(t_k))}}, \quad (22)$$

$$G^{SII}(I^{SII}(t_k), R^{SII}(t_k)) = \tilde{b} I^{SII}(t_k) \ln(1 + R^{SII}(t_k)), \quad (23)$$

where $\tilde{a}$ and $\tilde{b}$ are constants. At the beginning of each LTS, the long-term revenue is obtained when a slice is admitted by the operator.

2) The short-term revenue is obtained by saving the power consumption of the system in each slot. The short-term revenue of slice *SI* at slot t is given by

$$r_{SI}^{revn}(t) = P_{SI}^{max} - P_{SI}(t), \tag{24}$$

where P_{SI}^{max} is the maximum power consumption of slice *SI*.

Similarly, the short-term revenue of slice *SII* at slot t is given by

$$r_{SII}^{revn}(t) = P_{SII}^{max} - P_{SII}(t), \tag{25}$$

where P_{SII}^{max} is the maximum power consumption of slice *SII*.

Therefore, during one LTS, the revenue for slice *SI* is given by

$$U_{SI}(t_k) = a_s(t_k)[G^{SI}(I^{SI}(t_k), d^{SI}(t_k)) + \eta^{-1} \sum_{t=t_k}^{t_k+T-1} r_{SI}^{revn}(t)], \tag{26}$$

and the revenue for slice *SII* is given by

$$U_{SII}(t_k) = a_s(t_k)[G^{SII}(I^{SII}(t_k), R^{SII}(t_k)) + \eta^{-1} \sum_{t=t_k}^{t_k+T-1} r_{SII}^{revn}(t)], \tag{27}$$

where η is a factor used to strike the tradeoff between long-term revenue and short-term revenue, and $U_{SI}(t_k)$ and $U_{SII}(t_k)$ can be interpreted as the profit for slice SI and slice SII, respectively.

Then, the overall revenue gain from the accepted slice requests is

$$U(t_k) = U_{SI}(t_k) + U_{SII}(t_k). \tag{28}$$

Furthermore, we define the average revenue of the operator as

$$\bar{U} = \lim_{Z\to\infty} \frac{1}{Z} \sum_{t=0}^{Z-1} E\{U(t)\}. \tag{29}$$

E. Problem Formulation

In this paper, we investigate the operator's revenue maximization problem in MEC systems by jointly controlling slice request admission $a_s(t_k)$, the number of channels $\boldsymbol{N}(t_k) = (\mid N^{SI}(t_k) \mid, \mid N^{SII}(t_k) \mid)$, subcarrier assignment $\boldsymbol{\rho}(t) = (\boldsymbol{\rho}_1(t), \boldsymbol{\rho}_2(t))$, user association $\boldsymbol{y}(t) = (\boldsymbol{y}_1(t), \boldsymbol{y}_2(t))$, clock speed of the CPU $\boldsymbol{f}(t) = (f_{ib,SI}(t))$, and power allocation $\boldsymbol{P}_{SII}^{'}(t) = (P_{bi,SII}^{n}(t))$. In particular, we formulate it as the following stochastic optimization problem.

$$\begin{aligned}
&\max_{a_s(t),\boldsymbol{P}_{SII}^{'}(t),\boldsymbol{\rho}(t),\boldsymbol{y}(t),\boldsymbol{f}(t),\boldsymbol{N}(t_k)} \quad \lim_{Z\to\infty} \frac{1}{Z} \sum_{t=0}^{Z-1} \mathbb{E}\{U(t)\} \\
&\text{s.t.} \quad (\text{C1}) : \mid N^{SI}(t_k) \mid \geq N_{SI,min}, \mid N^{SII}(t_k) \mid \geq N_{SII,min}, \forall t_k, \\
&\quad (\text{C2}) : \mid N^{SI}(t_k) \mid + \mid N^{SII}(t_k) \mid = N, \forall t_k, \\
&\quad (\text{C3}) : a_s(t_k) \in \{0,1\}, \ \forall s \in \{SI, SII\}, \\
&\quad (\text{C4}) : \rho_{ib,SI}^{n}(t), \rho_{bi,SII}^{n}(t) \in \{0,1\}, \forall n, b, t, \\
&\quad (\text{C5}) : \sum_{b\in\mathcal{B}} \sum_{i\in I^{SII}(t_k)} \rho_{bi,SII}^{n}(t) \leq 1, \forall, n, b, t, \\
&\quad (\text{C6}) : \sum_{i\in I^{SII}(t_k)} \sum_{n=1}^{N^{SII}(t_k)} \rho_{bi,SII}^{n}(t) P_{bi,SII}^{n}(t) \leq P_{b,SII}^{max}, \forall b, t, \\
&\quad (\text{C7}) : P_{bi,SII}^{n}(t) \geq 0, \forall i \in I^{SII}(t), n, b, t, \\
&\quad (\text{C8}) : 0 \leq f_{ib,SI}(t) \leq f_{ib,SI}^{max}, \forall i \in I^{SI}(t_k), t, \\
&\quad (\text{C9}) : y_{ib,SI}(t), y_{bi,SII}(t) \in \{0,1\}, \forall i, b, t, \\
&\quad (\text{C10}) : \sum_{i\in\Omega_B} \rho_{ib,SI}^{n}(t) \leq 1, \forall n, b, t, \\
&\quad (\text{C11}) : \sum_{b\in\mathcal{B}} y_{ib,SI}(t) = 1, \sum_{b\in\mathcal{B}} y_{bi,SII}(t) = 1, \forall i, b, t, \\
&\quad (\text{C12}) : \bar{Q}_{SI} < \infty, \bar{F}_{SI} < \infty, \bar{Q}_{SII} < \infty, \forall t, \\
&\quad (\text{C13}) : (11), (19).
\end{aligned} \tag{30}$$

In (30), $f_{ib,SI}^{max}$ and $P_{b,SII}^{max}$ denote the maximum CPU-cycle frequency of an MEC server and the maximum transmit power of BS b in slice *SII*, respectively. Meanwhile, $N_{SI,min}$ and $N_{SII,min}$ denote the minimum number of channels required for slice *SI* and slice *SII*, respectively. (C1) and (C2) are constraints on the allocation of channel numbers for two slices. (C4), (C5), and (C10) indicate that any subcarrier of a BS can be allocated at most to one mobile device. (C6) and (C7) are the peak and nonnegative transmit power constraints, respectively. (C8) is the peak CPU-cycle frequency constraint. (C9) and (C11) denote that all MDs must be and at most associated with one BS. (C12) represents that the data rate should be greater than or equal to the arrival rate for all data queues and a processing queue which is equivalent to mean rate stability[1] [27].

III. ALGORITHM FRAMEWORK AND DESIGN OF NETWORK POLICIES

In this section, we devise an algorithm, called the DSSRA, to solve (30) based on the Lyapunov optimization [27]. With the help of the Lyapunov optimization framework, we can solve this challenging stochastic optimization problem by decomposing the two timescales problem into much single time slot subproblem. Besides, a slice request admission and channel slicing allocation algorithm will be designed. Next, we will show the proposed algorithm is capable of achieving the revenue-delay tradeoff in MEC systems.

A. The Lyapunov Optimization-Based Algorithm

Let $\boldsymbol{\Theta}(t) \triangleq [Q_{SI}(t), Q_{SII}(t), F_{SI}(t)]$ be a concatenated vector. Then, we define the Lyapunov function as

$$L(\boldsymbol{\Theta}(t)) \triangleq \frac{1}{2}\left[Q_{SI}^2(t) + Q_{SII}^2(t) + F_{SI}^2(t)\right]. \tag{31}$$

Then the LTS conditional Lyapunov drift $\triangle_T(\boldsymbol{\Theta}(t_k))$ is given by [27]

$$\triangle_T(\boldsymbol{\Theta}(t_k)) \triangleq \mathbb{E}[L(\boldsymbol{\Theta}(t_k + T)) - L(\boldsymbol{\Theta}(t_k))|\Theta(t_k)], \tag{32}$$

where $\boldsymbol{\Theta}(t_k) = \{Q_{SI}(t), Q_{SII}(t), F_{SI}(t), t \in [t_k, t_k+T-1]\}$. Accordingly, the drift-plus-penalty expression of (30) can be expressed as

$$\triangle_T(\boldsymbol{\Theta}(t_k)) - V\mathbb{E}\{U(t_k)|\Theta(t_k)\}. \tag{33}$$

[1]Mean rate stable of queue means $\lim_{t\to\infty} \frac{E\{Q(t)\}}{t} = 0$.

The following theorem provides an upper bound on the above drift-plus-penalty expression.

Theorem 1. *Suppose $\boldsymbol{G}(t)=(\boldsymbol{G}_{SI}(t),\boldsymbol{G}_{SII}(t))$ is i.i.d. over slots. For arbitrary $a_s(t),\boldsymbol{P}_{SII}(t),\boldsymbol{\rho}(t),\boldsymbol{y}(t),\boldsymbol{f}(t),\boldsymbol{N}(t_k)$, all parameters $V>0$, and all possible values of $\boldsymbol{\Theta}(t_k)$, $\triangle_T(\boldsymbol{\Theta}(t_k))-V\mathbb{E}\{U(t_k)|\Theta(t_k)\}$ is upper bounded by*

$$\begin{aligned}
&\triangle_T(\boldsymbol{\Theta}(t_k))-V\mathbb{E}\{U(t_k)|\Theta(t_k)\}\\
&\le C-\sum_{t=t_k}^{t_k+T-1}Q_{SI}(t)\mathbb{E}\Bigg\{a_s(t_k)\Bigg[\sum_{i\in I^{SI}(t_k)}R_{i,SI}(t)\tau\\
&-\sum_{i\in I^{SI}(t_k)}A_{i,SI}(t)\Bigg]\Bigg|\boldsymbol{\Theta}(t_k)\Bigg\}-\sum_{t=t_k}^{t_k+T-1}Q_{SII}(t)\mathbb{E}\Bigg\{a_s(t_k)\\
&\Bigg[\sum_{i\in I^{SII}(t_k)}R_{i,SII}(t)\tau-\sum_{i\in I^{SII}(t_k)}A_{i,SII}(t)\Bigg]\Bigg|\boldsymbol{\Theta}(t_k)\Bigg\}\\
&-\sum_{t=t_k}^{t_k+T-1}F_{SI}(t)\mathbb{E}\Bigg\{a_s(t_k)\Bigg[\sum_{b\in\mathcal{B}}\sum_{i\in I^{SI}(t_k)}y_{ib,SI}(t)f_{ib,SI}(t)\\
&-\sum_{i\in I^{SI}(t_k)}L_iR_{i,SI}(t)\Bigg]\Bigg|\boldsymbol{\Theta}(t_k)\Bigg\}-V\mathbb{E}\Bigg\{a_s(t_k)\Big[\\
&G^{SI}(I^{SI}(t_k),d^{SI}(t_k))+\eta\sum_{t=t_k}^{t_k+T-1}r_{SI}^{revn}(t)+\eta^{-1}\sum_{t=t_k}^{t_k+T-1}\\
&r_{SII}^{revn}(t)+G^{SII}(I^{SII}(t_k),R^{SII}(t_k))\Big]|\boldsymbol{\Theta}(t_k)\Bigg\},
\end{aligned}\tag{34}$$

$$\begin{aligned}
C\ge\frac{a_s(t_k)^2}{2}\Bigg\{&\Bigg[\sum_{t=t_k}^{t_k+T-1}\sum_{i\in I^{SI}(t_k)}R_{i,SI}(t)\tau\Bigg]^2+\Bigg[\sum_{t=t_k}^{t_k+T-1}\\
&\sum_{i\in I^{SI}(t_k)}A_{i,SI}(t)\Bigg]^2+\Bigg[\sum_{t=t_k}^{t_k+T-1}\sum_{i\in I^{SII}(t_k)}R_{i,SII}(t)\tau\Bigg]^2\\
&+\Bigg[\sum_{t=t_k}^{t_k+T-1}\sum_{i\in I^{SII}(t_k)}A_{i,SII}(t)\Bigg]^2+\sum_{t=t_k}^{t_k+T-1}\sum_{b\in\mathcal{B}}\sum_{i\in I^{SI}(t_k)}\\
&y_{ib,SI}(t)f_{ib,SI}(t)+\max_{i\in I^{SI}(t)}\Bigg[\sum_{t=t_k}^{t_k+T-1}L_iR_{i,SI}(t)\Bigg]^2\Bigg\}.
\end{aligned}\tag{35}$$

where C is a non-negative constant that satisfies

Proof: Please refer to Appendix A.

In [27], the stochastic optimization theory indicates that a stochastic optimization problem can be solved by minimizing the upper bound of its drift-plus-penalty expression subject to the same constraints except the stability one. In this paper, we need to minimize the right-hand-side of (34) to solve (30) subject to (C1)-(C11) and (C13), because (C12) is a stability constraint. The detailed procedure to solve (30), called DNSRA, is summarized in Algorithm 1.

We need to explain that the objective function of (36) is obtained 1) by ignoring the constants C, $Q_{SI}(t)\sum_{i\in I^{SI}(t_k)}A_{i,SI}(t)$, $Q_{SII}(t)\sum_{i\in I^{SII}(t_k)}A_{i,SII}(t)$, $G^{SI}(I^{SI}(t_k),d^{SI}(t_k))$, and $G^{SII}(I^{SII}(t_k),R^{SII}(t_k)$ in (34), and 2) by removing the conditional expectations in (34), because minimizing $f(t)$ ensures that $\mathbb{E}\{f(t)|\boldsymbol{\Theta}(t)\}$ is minimized from the principle of opportunistically minimizing an expectation [27].

Algorithm 1 Dynamic Network Slicing and Resource Allocation Algorithm (**DNSRA**)

1: In each LTS $t_k=pT$, $p=0,1,...$, obtain the current queue state $Q_{SI}(t_k)$ and $Q_{SII}(t_k)$, and $\{I^{SI}(t_k),d^{SI}(t_k)\}$ and $\{I^{SII}(t_k),R^{SII}(t_k)\}$.
2: Determine slicing request admission $a_s(t_k)$ and channel assignment $N^{SI}(t_k)$ and $N^{SII}(t_k)$ by calling Algorithm 5.
3: In each STS $t\in[t_k,t_k+T-1]$, obtain user association $\boldsymbol{y}(t)$, subcarrier assignment $\boldsymbol{\rho}(t)$, CPU-cycle frequency $\boldsymbol{f}(t)$, and power allocation $\boldsymbol{P}_{SII}^{'}(t)$ according to

$$\begin{aligned}
\min_{a_s(t),\boldsymbol{P}_{SII}^{'}(t),\boldsymbol{\rho}(t),\boldsymbol{y}(t),\boldsymbol{f}(t),\boldsymbol{N}(t_k)}\quad&V\eta^{-1}\left[P_{SI}(t)+P_{SII}(t)\right]\\
&+\sum_{i\in I^{SI}(t_k)}R_{i,SI}(t)\left[F_{SI}(t)L_i-Q_{SI}(t)\tau\right]\\
&-Q_{SII}(t)\sum_{i\in I^{SII}(t_k)}R_{i,SII}(t)\tau\\
&-F_{SI}(t)\sum_{b\in\mathcal{B}}\sum_{i\in I^{SI}(t_k)}y_{ib,SI}(t)f_{ib,SI}(t)\\
\text{s.t.}\quad&(\text{C1})-(\text{C11}),\ (\text{C13}).
\end{aligned}\tag{36}$$

4: $t=t+1$.
5: Updata the current queue length $\boldsymbol{Q}_{SI}(t)$, $F_{SI}(t)$, and $\boldsymbol{Q}_{SII}(t)$ according to (8), (14), and (21).

In LTS t_k, by substituting (12), (13), and (18) into (36), we recast (36) to

$$\begin{aligned}
&\min_{a_s(t),\boldsymbol{P}_{SII}^{'}(t),\boldsymbol{\rho}(t),\boldsymbol{y}(t),\boldsymbol{f}(t),\boldsymbol{N}(t_k)}\Phi(a_s(t_k),\boldsymbol{N}(t_k))f_{ib,SI}^3(t)\\
&=V\eta^{-1}\sum_{b\in\mathcal{B}}\Bigg[\sum_{i\in I^{SI}(t_k)}y_{ib,SI}(t)k_{ser,b}f_{ib,SI}^3(t)+y_{bi,SII}(t)\\
&\sum_{i\in I_{SII}(t_k)}\sum_{n=1}^{|N^{SII}(t_k)|}\rho_{bi,SII}^n(t)P_{bi,SII}^n(t)\Bigg]+\sum_{i\in I^{SI}(t_k)}\Bigg\{[F_{SI}(t)L_i\\
&-Q_{SI}(t)\tau]\sum_{b\in\mathcal{B}}y_{ib,SI}(t)\sum_{n=1}^{|N^{SI}(t_k)|}\rho_{ib,SI}^n(t)R_{ib,SI}^n(t)\Bigg\}\\
&-Q_{SII}(t)\tau\sum_{i\in I^{SII}(t_k)}\sum_{b\in\mathcal{B}}y_{bi,SII}(t)\sum_{n=1}^{|N^{SII}(t_k)|}\rho_{bi,SII}^n(t)\\
&R_{bi,SII}^n(t)-F_{SI}(t)\sum_{b\in\mathcal{B}}\sum_{i\in I^{SI}(t_k)}y_{ib,SI}(t)f_{ib,SI}(t)\\
&\text{s.t.}\quad(\text{C1})-(\text{C11}),\ (\text{C13}).
\end{aligned}\tag{37}$$

In (37), we can observe that user association, subcarrier assignment, power allocation, and CPU-cycle frequency allocation are highly coupled with each other. We further decouple

these optimization variables to develop low-complexity algorithms.

B. Algorithms Design

1) Solution of Computation Offloading Slice SI: Under given slice request admission $a_s(t_k)$ and channel allocation $N^{SI}(t_k)$ in LTS t_k, the resource allocation problem for slice *SI* can be obtained by solving the following problem.

$$\begin{aligned}
&\min_{\boldsymbol{\rho}_1(t),\boldsymbol{f}(t),\boldsymbol{y}(t)} V\eta^{-1}\sum_{b\in\mathcal{B}}\sum_{i\in I^{SI}(t_k)} y_{ib,SI}(t)k_{ser,b}f_{ib,SI}^3(t)\\
&+\sum_{i\in I^{SI}(t_k)}\Bigg\{[F_{SI}(t)L_i-Q_{SI}(t)\tau]\\
&\quad\sum_{b\in\mathcal{B}} y_{ib,SI}(t)\sum_{n=1}^{|N^{SI}(t_k)|}\rho_{ib,SI}^n(t)R_{ib,SI}^n(t)\Bigg\}\\
&\quad -F_{SI}(t)\sum_{b\in\mathcal{B}}\sum_{i\in I^{SI}(t_k)} y_{ib,SI}(t)f_{ib,SI}(t) \qquad (38)\\
&\text{s.t.}\quad (\text{C4})': \rho_{ib,SI}^n(t)\in\{0,1\},\forall n,b,t,\\
&\quad (\text{C8}): 0\le f_{ib,SI}(t)\le f_{ib,SI}^{max},\forall i\in I^{SI}(t_k),t,\\
&\quad (\text{C9})': y_{ib,SI}(t)\in\{0,1\},\forall i,b,t,\\
&\quad (\text{C10}): \sum_{i\in\Omega_B}\rho_{ib,SI}^n(t)\le 1,\forall n,b,t,\\
&\quad (\text{C11})': \sum_{b\in\mathcal{B}} y_{ib,SI}(t)=1,\forall i,b,t,\\
&\quad (\text{C13})': (11),
\end{aligned}$$

where $\sum_{b\in\mathcal{B}} y_{bi,SII}(t)\sum_{i\in I_{SII}(t_k)}\sum_{n=1}^{|N^{SII}(t_k)|}\frac{V}{\eta}\rho_{bi,SII}^n(t)P_{bi,SII}^n(t)-Q_{SII}(t)\tau\sum_{i\in I^{SII}(t_K)}\sum_{b\in\mathcal{B}} y_{bi,SII}(t)\sum_{n=1}^{|N^{SII}(t_k)|}\rho_{bi,SII}^n(t)R_{bi,SII}^n(t)$ in the objective function is omitted because it is a constant that doesn't affect the solution of the problem.

♦ *Optimal CPU-Cycle Frequencies of the MEC server*

The CPU-cycle frequencies problem for a given $\boldsymbol{y}_1(t)$ and $\boldsymbol{\rho}_1(t)$ from (38) becomes

$$\begin{aligned}
&\min_{\boldsymbol{f}(t)}\sum_{b\in\mathcal{B}}\sum_{i\in I^{SI}(t)}\left[V\eta^{-1}k_{ser,b}f_{ib,SI}^3(t)-F_{SI}(t)f_{ib,SI}(t)\right]\\
&\text{s.t.}\quad (\text{C8}): 0\le f_{ib,SI}(t)\le f_{ib,SI}^{max},\forall i\in I^{SI}(t_k),t, \qquad (39)\\
&\qquad (\text{C13})': f_{ib,SI}(t)\ge\frac{D_{i,SI}(t)}{d^{SI}(t_k)-\frac{D_{i,SI}(t)}{R_{i,SI}(t)}},\ \forall k,i.
\end{aligned}$$

Since the objective function of (39) is convex and its constraints are linear, (39) is a convex optimization problem. Furthermore, the optimization of $f_{ib,SI}(t)$ can be done separately for the computational tasks of each mobile device, because the objective function and constraints of (39) can be decomposed for individual $f_{ib,SI}(t)$. Then, the optimal CPU-cycle frequencies required by the MEC server b to process the computational tasks of mobile device i can be expressed as

$$f_{ib,SI}(t)= \qquad (40)$$
$$\begin{cases}
\frac{D_{i,SI}(t)}{d^{SI}(t_k)-\frac{D_{i,SI}(t)}{R_{i,SI}(t)}}, & \sqrt{\frac{F_{SI}(t)\eta}{3Vk_{ser,b}}}<\frac{D_{i,SI}(t)}{d^{SI}(t_k)-\frac{D_{i,SI}(t)}{R_{i,SI}(t)}},\\
\sqrt{\frac{F_{SI}(t)\eta}{3Vk_{ser,b}}}, & \frac{D_{i,SI}(t)}{d^{SI}(t_k)-\sqrt{\frac{D_{i,SI}(t)}{R_{i,SI}(t)}}}<\sqrt{\frac{F_{SI}(t)\eta}{3Vk_{ser,b}}}<f_{ib,SI}^{max},\\
f_{ib,SI}^{max}, & \sqrt{\frac{F_{SI}(t)\eta}{3Vk_{ser,b}}}>f_{ib,SI}^{max}.
\end{cases}$$

♦ *User Association*

For notional brevity, we let

$$R_{ib,SI}^Z(t)=\sum_{n=1}^{N^{SI}(t_k)}\rho_{ib,SI}^n(t)R_{ib,SI}^n(t). \qquad (41)$$

Therefore, for a given $\boldsymbol{\rho}_1(t)$ and $\boldsymbol{f}(t)$, the user association problem from (38) is given by

$$\begin{aligned}
&\min_{\boldsymbol{y}_1(t)}\sum_{i\in I^{SI}(t_k)}\sum_{b\in\mathcal{B}}[V\eta^{-1}k_{ser,b}f_{ib,SI}^3(t)+(F_{SI}(t)L_i\\
&\quad -Q_{SI}(t)\tau)R_{ib,SI}^Z(t)-F_{SI}(t)f_{ib,SI}(t)]y_{ib,SI}(t)\\
&\quad \text{s.t.}(\text{C9}'): y_{ib,SI}(t)\in\{0,1\},\forall i,b,t,\\
&\qquad (\text{C11}'): \sum_{b\in\mathcal{B}} y_{ib,SI}(t)=1,\forall i,b,t. \qquad (42)
\end{aligned}$$

Then, the optimal solution of (42) can be expressed as

$$y_{ib,SI}(t)=\begin{cases}1, & \text{if } b=b^*(t),\\ 0, & \text{if } b\ne b^*(t),\end{cases} \qquad (45)$$

where

$$\begin{aligned}
b^*(t)=\arg\min_{b\in\mathcal{B}}\Big\{&V\eta^{-1}k_{ser,b}f_{ib,SI}^3(t)+(F_{SI}(t)L_i\\
&-Q_{SI}(t)\tau)R_{ib,SI}^Z(t)-F_{SI}(t)f_{ib,SI}(t)\Big\}. \qquad (46)
\end{aligned}$$

♦ *Subcarrier Assignment*

If the user association $\boldsymbol{y}_1(t)$ is determined, then the set of mobile devices served by BS b is denoted by $I_b^{SI}(t_k)$, as follow.

$$I_b^{SI}(t_k)=\{i\mid y_{ib,SI}(t)=1, i\in I^{SI}(t_k)\},\forall b\in\mathcal{B}. \qquad (47)$$

For this given $\boldsymbol{y}_1(t)$, the subcarrier assignment problem is given by (48).

For all $b\in\mathcal{B}$ and $i\in I_b^{SI}(t_k)$, we let

$$\begin{aligned}
\varphi_{ib,SI}(t)=&[F_{SI}(t)L_i-Q_{SI}(t)\tau]\\
&W\log_2(1+\frac{P_{ib,SI}^n g_{ib,SI}^n(t)}{\sum_{c\in\mathcal{B},c\ne b}\sum_{j=1}^{M_1}P_{jc,SI}^n g_{j(i)c,SI}^n(t)+\sigma_i^2(t)}). \qquad (49)
\end{aligned}$$

For subcarrier n, the subcarrier assignment is given by

$$\rho_{ib,SI}^n(t)=$$
$$\begin{cases}
0, & \text{if } \varphi_{ib,SI}(t)\ge 0,\\
1, & \text{if } \varphi_{ib,SI}(t)\le 0 \text{ and } i=\arg\min_j\varphi_{jb,SI}(t),\\
0, & \text{if } \varphi_{ib,SI}(t)\le 0 \text{ and } i\ne\arg\min_j\varphi_{jb,SI}(t).
\end{cases} \qquad (50)$$

We can observe, from (50), that there is a case where some subcarriers may not be allocated to any mobile device in slot t. Let $\mathcal{A}$ denote the set of unassigned subcarrier. We introduce

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2020.2992607, IEEE Transactions on Vehicular Technology

$$\min_{\boldsymbol{\rho}_1(t)} \sum_{b\in\mathcal{B}}\sum_{i\in I_b^{SI}(t_k)}\sum_{n=1}^{N^{SI}(t_k)} \left\{ [F_{SI}(t)L_i - Q_{SI}(t)\tau]W\log_2(1+\frac{P^n_{ib,SI}g^n_{ib,SI}(t)}{\sum\limits_{c\in\mathcal{B},c\neq b}\sum\limits_{j=1}^{M_1} P^n_{jc,SI}g^n_{j(i)c,SI}(t)+\sigma_i^2(t)}) \right\} \rho^n_{ib,SI}(t)$$

$$\text{s.t.} \quad (\text{C4})' : \rho^n_{ib,SI}(t)\in\{0,1\},\forall n,b,t, \tag{48}$$

$$(\text{C10}) : \sum_{i\in I_b^{SI}(t_k)} \rho^n_{ib,SI}(t)\le 1,\forall n,b,t,$$

$$(\text{C13})' : \sum_{n=1}^{N^{SI}(t_k)} \rho^n_{ib,SI}(t)W\log_2(1+\frac{P^n_{ib,SI}g^n_{ib,SI}(t)}{\sum\limits_{c\in\mathcal{B},c\neq b}\sum\limits_{j=1}^{M_1} P^n_{jc,SI}g^n_{j(i)c,SI}(t)+\sigma_i^2(t)}) \ge \frac{D_{i,SI}(t)}{d^{SI}(t_k)-\frac{D_{i,SI}(t)}{f_{ib,SI}(t)}}.$$

$\boldsymbol{\Omega}_{ib}$ to denote the set of subcarriers that mobile devices have assigned. To meet the basic data rate of mobile devices, i.e., $(\text{C13})'$, we develop a subcarrier assignment algorithm to solve (48), which the detailed process is shown in Algorithm 2.

Algorithm 2 Subcarrier Assignment Algorithm

Initialization:

1: Obtain $\boldsymbol{\Omega}_{ib}$ for all $i\in I_b^{SI}(t_k)$ and $\mathcal{A}$ according to (50).

Iteration:

2: Find i^* with $R_{i^*,SI}(t) < \frac{D_{i^*,SI}(t)}{d^{SI}(t_k)-\frac{D_{i^*,SI}(t)}{f_{i^*b,SI}(t)}}$ and $R_{i^*,SI}(t) - \frac{D_{i^*,SI}(t)}{d^{SI}(t_k)-\frac{D_{i^*,SI}(t)}{f_{i^*b,SI}(t)}} \le R_{i,SI}(t) - \frac{D_{i,SI}(t)}{d^{SI}(t_k)-\frac{D_{i,SI}(t)}{f_{ib,SI}(t)}}$ for all $i\in I_b^{SI}(t_k)$.

3: Find n^* satisfying $g^{n^*}_{i^*b}(t)\ge g^n_{i^*b}(t)$ for $n\in\mathcal{A}$ for the find i^*.

4: Update $\Omega_{i^*b} = \Omega_{ib}\cup\{n^*\}$, $\mathcal{A}=\mathcal{A}-\{n^*\}$, and $R_{i^*,SI}(t)=R_{i^*,SI}(t)+\log_2(1+\gamma_{i^*b}(t))$.

5: **Until** $R_{i,SI}(t)\ge\frac{D_{i,SI}(t)}{d^{SI}(t_k)-\frac{D_{i,SI}(t)}{f_{ib,SI}(t)}}$ for all i.

2) Solution of Data Sharing Slice SII: Under given slice request admission $a_s(t_k)$ and channel allocation $N^{SII}(t_k)$ in LTS t_k, the resource allocation problem for slice *SII* can be obtained by solving the following problem.

$$\min_{\boldsymbol{\rho}_2(t),\boldsymbol{y}_2(t),\boldsymbol{P}'_{SII}(t)} \sum_{b\in\mathcal{B}}\sum_{i\in I_{SII}(t_k)}\sum_{n=1}^{N^{SII}(t_k)} y_{bi,SII}(t)$$
$$\left[V\eta^{-1}\rho^n_{bi,SII}(t)P^n_{bi,SII}(t) - Q_{SII}(t)\tau\rho^n_{bi,SII}(t)R^n_{bi,SII}(t)\right]$$
$$\text{s.t.} \quad (\text{C4})'' : \rho^n_{bi,SII}(t)\in\{0,1\},\forall n,b,t,$$
$$(\text{C5}) : \sum_{b\in\mathcal{B}}\sum_{i\in I^{SII}(t_k)} \rho^n_{bi,SII}(t)\le 1,\forall,n,b,t,$$
$$(\text{C6}) : \sum_{i\in I^{SII}(t_k)}\sum_{n=1}^{N^{SII}(t_k)} \rho^n_{bi,SII}(t)P^n_{bi,SII}(t)\le P^{max}_{b,SII},\forall b,t,$$
$$(\text{C7}) : P^n_{bi,SII}(t)\ge 0,\forall i\in I^{SII}(t),n,b,t, \tag{51}$$
$$(\text{C9})'' : y_{bi,SII}(t)\in\{0,1\},\forall i,b,t, \tag{52}$$
$$(\text{C11})'' : \sum_{b\in\mathcal{B}} y_{bi,SII}(t)=1,\forall i,b,t.$$
$$(\text{C13})' : \sum_{b\in\mathcal{B}}\sum_{n=1}^{N^{SII}(t_k)} \rho^n_{bi,SII}(t)R^n_{bi,SII}(t)\ge R^{SII}(t_k).$$

♦ *User Association*

The user association problem for a given $\boldsymbol{\rho}_2(t)$ and $\boldsymbol{P}_{SII}(t)$ from (51) becomes

$$\min_{\boldsymbol{y}_2(t)} \sum_{b\in\mathcal{B}}\sum_{i\in I_{SII}(t_k)}\sum_{n=1}^{N^{SII}(t_k)} y_{bi,SII}(t)\left[V\eta^{-1}\rho^n_{bi,SII}(t) P^n_{bi,SII}(t) - Q_{SII}(t)\tau\rho^n_{bi,SII}(t)R^n_{bi,SII}(t)\right]$$
$$(\text{C9})'' : y_{bi,SII}(t)\in\{0,1\},\forall i,b,t, \tag{53}$$
$$(\text{C11})'' : \sum_{b\in\mathcal{B}} y_{bi,SII}(t)=1,\forall i,b,t.$$

The optimal solution of (53) can be expressed as

$$y_{bi,SII}(t) = \begin{cases} 0, & b\neq b^*(t), \\ 1, & b=b^*(t), \end{cases} \tag{54}$$

where $b^*(t) = \arg\min_{b\in\mathcal{B}}\{V\eta^{-1}\rho^n_{bi,SII}(t)P^n_{bi,SII}(t) - Q_{SII}(t)\tau\rho^n_{bi,SII}(t)R^n_{bi,SII}(t)\}$.

♦ *Subcarrier Assignment and Power Allocation*

Similar to (47), the set of mobile devices served by BS b is denoted by $I_b^{SII}(t_k)$, as follow.

$$I_b^{SII}(t_k) = \{i \mid y_{bi,SII}(t)=1, i\in I^{SII}(t_k)\},\forall b\in\mathcal{B}. \tag{55}$$

After we obtain $\boldsymbol{y}_2(t)$, the subcarrier assignment and power allocation problem then becomes (56). It is difficult to solve the problem (56), because it is a mixed-integer nonlinear programming problem. To solve this problem, we first relax $\rho^n_{bi,SII}(t)$ to $[0,1]$, and then introduce a new variable $s^n_{bi}(t)=\rho^n_{bi,SII}(t)P^n_{bi,SII}(t)$.

$$\min_{\boldsymbol{\rho}_2(t),\boldsymbol{P}'_{SII}(t)} \sum_{b\in\mathcal{B}}\sum_{i\in I_{SII}(t_k)}\sum_{n=1}^{N^{SII}(t_k)} \rho^n_{bi,SII}(t)\Bigg[-Q_{SII}(t)\tau W\log_2(1+\frac{P^n_{bi,SII}(t)g^n_{bi,SII}(t)}{\sum\limits_{c\in\mathcal{B},c\neq b}\sum\limits_{j=1}^{M_2}\rho^n_{cj,SII}(t)P^n_{cj,SII}(t)g^n_{cj(i),SII}(t)+\sigma^2_i(t)})$$
$$+V\eta P^n_{bi,SII}(t)\Big]$$
$$\text{s.t.}\quad (\text{C4})'': \rho^n_{bi,SII}(t)\in\{0,1\},\forall n,b,t,$$
$$(\text{C5}): \sum_{b\in\mathcal{B}}\sum_{i\in I^{SII}(t_k)}\rho^n_{bi,SII}(t)\le 1,\forall,n,b,t,$$
$$(\text{C6}): \sum_{i\in I^{SII}(t_k)}\sum_{n=1}^{N^{SII}(t_k)}\rho^n_{bi,SII}(t)P^n_{bi,SII}(t)\le P^{max}_{b,SII},\forall b,t,$$
$$(\text{C7}): P^n_{bi,SII}(t)\ge 0,\forall i\in I^{SII}(t),n,b,t, \tag{56}$$
$$(\text{C13})': \sum_{n=1}^{N^{SII}(t_k)}\rho^n_{bi,SII}(t)W\log_2(1+\frac{P^n_{bi,SII}(t)g^n_{bi,SII}(t)}{\sum\limits_{c\in\mathcal{B},c\neq b}\sum\limits_{j=1}^{M_2}\rho^n_{cj,SII}(t)P^n_{cj,SII}(t)g^n_{cj(i),SII}(t)+\sigma^2_i(t)})\ge R^{SII}(t_k).$$

We rearrange problem (56) to

$$\min_{\boldsymbol{\rho}_2(t),\boldsymbol{P}_{SII}(t)} \sum_{b\in\mathcal{B}}\sum_{i\in I^{SII}(t_k)}\sum_{n_1=1}^{N^{SII}(t_k)}\Bigg[\frac{V}{\eta}s^n_{bi}(t)-Q_{SII}(t)\tau W\log_2(1+\frac{s^n_{bi}(t)g^n_{bi,SII}(t)}{B^n_{jc,SII}(t)})\Bigg]$$
$$\text{s.t.}\quad (\text{C4})'': \rho^n_{bi,SII}(t)\in\{0,1\},\forall n,b,t, \tag{57}$$
$$(\text{C5}): \sum_{b\in\mathcal{B}}\sum_{i\in I^{SII}(t_k)}\rho^n_{bi,SII}(t)\le 1,\forall,n,b,t,$$
$$(\text{C6}): \sum_{i\in I^{SII}(t_k)}\sum_{n=1}^{N^{SII}(t_k)}s^n_{bi}(t)\le P^{max}_{b,SII},\forall b,t,$$
$$(\text{C7}): s^n_{bi}(t)\ge 0,\forall i\in I^{SII}(t),n,b,t,$$
$$(\text{C13})'': \sum_{n=1}^{N^{SII}(t_k)}W\log_2(1+\frac{s^n_{bi}(t)g^n_{bi,SII}(t)}{B^n_{jc,SII}(t)})\ge R^{SII}(t_k),$$

where $B^n_{jc,SII}(t)=\sum\limits_{c\in\mathcal{B},c\neq b}\sum\limits_{j=1}^{M_2}s^n_{cj}(t)g^n_{cj(i),SII}(t)+\sigma^2_i(t)$.

We develop an algorithm by approximating (57) in the exponential domain of power to solve this problem. Specially, let $v^n_{bi}(t)=\log(s^n_{bi}(t))$ when $s^n_{bi}(t)>0$ and $v^n_{bi}(t)=$ otherwise, and $\boldsymbol{v}(t)=(v^n_{bi}(t))$. Then $s^n_{bi}(t)=e^{v^n_{bi}(t)}$ when $v^n_{bi}(t)\neq 0$ and $s^n_{bi}(t)=0$ otherwise. We denote $f_i(\boldsymbol{v}(t))$ and $h_i(\boldsymbol{v}(t))$ by

$$f_i(\boldsymbol{v}(t))=\sum_{b\in\mathcal{B}}\sum_{n=1}^{N^{SI}(t_k)}\Bigg[\frac{V}{\eta}e^{v^n_{bi}(t)}+Q_{SII}(t)\tau W\log_2(\sum_{c\in\mathcal{B},c\neq b}\sum_{j\in I^{SII}_b(t_k)}e^{v^n_{cj}(t)}g^n_{j(i)c,SII}(t)+\sigma^2_i(t))\Bigg], \tag{58}$$

$$h_i(\boldsymbol{v}(t))=Q_{SII}(t)\tau W\log_2(\sum_{c\in\mathcal{B}}\sum_{j\in I^{SII}_b(t_k)}e^{v^n_{cj}(t)}g^n_{j(i)c,SII}(t)+\sigma^2_i(t)). \tag{59}$$

Then, (57) can be equivalently recast to

$$\max_{\boldsymbol{v}_i(t)}\sum_{i\in I^{SII}_b(t_k)}h_i(\boldsymbol{v}(t))-f_i(\boldsymbol{v}(t))$$
$$\text{s.t.}\ \ (\text{C4})'',(\text{C5}),$$
$$(\text{C6}): \sum_{i\in I^{SII}(t_k)}\sum_{n=1}^{N^{SII}(t_k)}e^{v^n_{bi}(t)}\le P^{max}_{b,SII},\forall b,t, \tag{60}$$
$$(\text{C7}): e^{v^n_{bi}(t)}\ge 0,\forall i\in I^{SII}(t),n,b,t,$$
$$(\text{C13})'': h'_i(t)-f'_i(t)\ge R^{SII}(t_k),$$

where

$$f'_i(t)=\sum_{b\in\mathcal{B}}\sum_{n=1}^{N^{SI}(t_k)}W\log_2(\sum_{c\in\mathcal{B},c\neq b}\sum_{j\in I^{SII}_b(t_k)}e^{v^n_{cj}(t)}g^n_{j(i)c,SII}(t)+\sigma^2_i(t)), \tag{61}$$

$$h'_i(t)=\sum_{b\in\mathcal{B}}\sum_{n=1}^{N^{SI}(t_k)}W\log_2(\sum_{c\in\mathcal{B}}\sum_{j\in I^{SII}_b(t_k)}e^{v^n_{cj}(t)}g^n_{j(i)c,SII}(t)+\sigma^2_i(t)). \tag{62}$$

It is observed that the objective function of (60) is the difference of two convex functions, which it has a D.C. structure in the $\boldsymbol{v}(t)$ domain. Therefore, we can apply the sequential convex approximation [28] to solve (60) by approximating $h_i(\boldsymbol{v}(t))$ and $h'_i(\boldsymbol{v}(t))$ in the $\boldsymbol{v}(t)$ domain. Then, we obtain an approximate $h_i(\boldsymbol{v}(t))$ and $h'_i(\boldsymbol{v}(t))$ by its first-order Taylor expansion at $\boldsymbol{v}^m(t)$, where $\boldsymbol{v}^m(t)$ is an approximation solution during each iteration, and m is the iteration index.

$$h_i(\boldsymbol{v}(t))\approx h_i(\boldsymbol{v}^m(t))+\bigtriangledown h^T_i(\boldsymbol{v}^m(t))(\boldsymbol{v}(t)-\boldsymbol{v}^m(t)), \tag{63}$$

$$h_i'(\boldsymbol{v}(t)) \approx h_i'(\boldsymbol{v}^m(t)) + \bigtriangledown h_i^{'T}(\boldsymbol{v}^m(t))(\boldsymbol{v}(t) - \boldsymbol{v}^m(t)), \quad (64)$$

where $\bigtriangledown h_i^T(\boldsymbol{v}^m(t))$ and $\bigtriangledown h_i^{'T}(\boldsymbol{v}^m(t))$ are the gradient of $h_i(\boldsymbol{v}(t))$ and $h_i'(\boldsymbol{v}(t))$, and are given by

$$\bigtriangledown h_i^T(\boldsymbol{v}^m(t)) = \frac{\partial h_i(\boldsymbol{v}(t))}{\partial v_{bi}^n(t)}, \ \forall i \in I_b^{SII}(t_k), n \in N^{SII}(t_k), \quad (65)$$

$$\bigtriangledown h_i^{'T}(\boldsymbol{v}^m(t)) = \frac{\partial h_i'(\boldsymbol{v}(t))}{\partial v_{bi}^n(t)}, \ \forall i \in I_b^{SII}(t_k), n \in N^{SII}(t_k). \quad (66)$$

Substituting (62) and (63) into (60) yields

$$\begin{aligned}
&\max_{\boldsymbol{v}_i(t)} \sum_{i \in I_b^{SII}(t_k)} \left[h_i(\boldsymbol{v}^m(t)) + \bigtriangledown h_i^T(\boldsymbol{v}^m(t))(\boldsymbol{v}(t) - \boldsymbol{v}^m(t))\right] \\
&\qquad - f_i(\boldsymbol{v}(t)) \\
&\text{s.t.} \quad \text{(C4)} :'' \ 0 \le \rho_{bi,SII}^n(t) \le 1, \ \forall n, b, t, \quad \text{(C5)}, \\
&\text{(C6)} : \sum_{i \in I^{SII}(t_k)} \sum_{n=1}^{N^{SII}(t_k)} e^{v_{bi}^n(t)} \le P_{b,SII}^{max}, \forall b, t, \\
&\text{(C7)} : e^{v_{bi}^n(t)} \ge 0, \forall i \in I^{SII}(t), n, b, t, \quad (67) \\
&\text{(C13)}'' : \left[h_i'(\boldsymbol{v}^m(t)) + \bigtriangledown h_i^{'T}(\boldsymbol{v}^m(t))(\boldsymbol{v}(t) - \boldsymbol{v}^m(t))\right] \\
&\qquad - f_i'(t) \ge R^{SII}(t_k).
\end{aligned}$$

Now, we have transformed the original problem (60) into a standard convex problem (67), which could be solved in polynominal time using standard CVX tools such as SeDuMi [29].

Based on the sequential convex approximation and the solutions of problem (67), we develop an algorithm to solve (56). Then, the procedure of the proposed **s**equential **c**onvex **a**pproximation based **p**ower **a**llocation and **s**ubcarrier **a**ssignment algorithm (SCA-PASA) is described in Algorithm 3.

Algorithm 3 Sequential Convex Approximation based Power Allocation and Subcarrier Assignment Algorithm (SCA-PASA)

Initialization:

1: Set $m = 0$,$\iota = 0$ and the maximum tolerance $\delta > 0$ and $\epsilon > 0$.
2: Set the initial $\boldsymbol{v}^0(t)$.
3: Compute $I^0(t) = \sum_{i \in I_b^{SII}(t_k)} h_i(\boldsymbol{v}^0(t)) - f_i(\boldsymbol{v}^0(t))$.

Iteration:

4: Obtain power allocation $\boldsymbol{P}_{SII}^{\iota}(t)$ and subcarrier assignment $\rho_2^{\iota}(t)$ by solving (67).
5: $\boldsymbol{P}_{SII}^m(t) = \boldsymbol{P}_{SII}^{\iota+1}(t)$, $\boldsymbol{\rho}_2^m(t) = \boldsymbol{\rho}_2^{\iota+1}(t)$.
6: Set $m = m+1$, $\boldsymbol{v}^m(t) = \boldsymbol{\rho}_2^m(t) \log_2(\boldsymbol{P}_{SII}^m(t))$, and $\iota = 0$.
7: Compute $I^m(t) = \sum_{i \in I_b^{SII}(t_k)} h_i(\boldsymbol{v}^m(t)) - f_i(\boldsymbol{v}^m(t))$.
8: **Until** $\mid I^m(t) - I^{m-1}(t) \mid \le \delta$.

By combining user association, subcarrier assignment, CPU-cycle frequency, and power allocation, we propose Algorithm 4 to solve (37) for given slice request admission $a_s(t_k)$ and the number of channels for two slices $N^{SI}(t_k)$ and $N^{SII}(t_k)$.

Algorithm 4 Resource allocation for given slice request admission and channel slicing strategy

Initialization:

1: Set user association $\boldsymbol{y}^0(t)$ and CPU-cycle frequency $\boldsymbol{f}^0(t)$.
2: Set $l = 0$ and the maximum tolerance $\varepsilon > 0$.
3: Assign subcarriers $\boldsymbol{\rho}_1^0(t)$ by calling Algorithm 2 based on $\boldsymbol{y}_1^0(t)$ and $\boldsymbol{f}^0(t)$.
4: Obtain power allocation $\boldsymbol{P}_{SII}^0(t)$ and subcarrier assignment $\boldsymbol{\rho}_2^0(t)$ by calling Algorithm 3 based on $\boldsymbol{y}_2^0(t)$.
5: Compute $\Phi_0(t_k)$ from (37) based on $\boldsymbol{y}^0(t)$, $\boldsymbol{f}^0(t)$, $\boldsymbol{\rho}^0(t)$, and $\boldsymbol{P}_{SII}^0(t)$.

Iteration:

6: Update $l = l + 1$.
7: Associate mobile devices $\boldsymbol{y}^l(t)$ from (45) and (54) based on $\boldsymbol{\rho}^{l-1}(t)$ and $\boldsymbol{P}_{SII}^{l-1}(t)$.
8: Allocate CPU-cycle frequency $f^l(t)$ based on $\boldsymbol{y}_1^l(t)$ and $\rho_1^{l-1}(t)$.
9: Assign subcarrier $\boldsymbol{\rho}_1^l(t)$ by calling Algorithm 2 based on $\boldsymbol{y}_1^l(t)$.
10: Allocate power $\boldsymbol{P}_{SII}^l(t)$ and assign subcarriers $\boldsymbol{\rho}_2^l(t)$ by calling Algorithm 3 based on $\boldsymbol{y}_2^l(t)$.
11: Compute $\Phi_l(N(t_k))$ from (37) based on $\boldsymbol{f}^l(t)$, $\boldsymbol{y}^l(t)$, $\boldsymbol{P}_{SII}^l(t)$, and $\boldsymbol{\rho}^l(t)$.
12: **Until** $\mid \Phi_l(N(t_k)) - \Phi_{l-1}(N(t_k)) \mid \le \varepsilon$.

3) Slice request admission and channel allocation for two slices: In 1) and 2), we depend on the assumption that $a_s(t_k)$, $N^{SI}(t_k)$, and $N^{SII}(t_k)$ are given. Next, we propose an algorithm to de determine the vale of $a_s(t_k)$, $N^{SI}(t_k)$, and $N^{SII}(t_k)$. Before describing the algorithm, we first introduce a concept, the *idealized revenue (IR)* R_s of a certain slice request s, which is the maximum revenue the operator obtains if the operator only accepts one slice request. Then, the whole procedure is shown in Algorithm 5.

IV. Simulation Results

In this section, we present simulation results to evaluate the performance of the proposed algorithms.

A. Simulation Parameters

We consider a network topology covering 1000m × 1000m area, which consists of one macro BS and 4 micro BSs. In the simulation, we adopt the frequency-selective channel as the wireless channel model, which consists of twelve independent Rayleigh multipaths. Each multipath component is modeled by the Clarkes flat fading model, and the relative power of the twelve multipath components are [0,-1,-4,-3,-3.5, -5, -7.0, -6.0, -7.5, -10.6, -12, -13] dB [30]. Path loss model between macro/micro BS and mobile device is 128.1+37.6log10(d) [dB]/140.7+37.6log10(d) [dB], where d is the distance between macro/micro BS and mobile devices [31]. Shadowing is 8 dB/10 dB. Other simulation parameters are summarized in

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2020.2992607, IEEE Transactions on Vehicular Technology

$$A_1(\boldsymbol{\rho}) = \sum_{t=t_k}^{t_k+T-1} \sum_{n\in N^{SI}(t_k)} \sum_{b\in\mathcal{B}} \sum_{i\in I^{SI}(t_k)} \rho^n_{ib,SI}(t) y_{ib,SI}(t) \left[\frac{V}{\eta} P^n_{ib,SI} + [F_{SI}(t)L_i - Q_{SI}(t)\tau] R^n_{ib,SI}(t)\right], \tag{68}$$

$$A_2(\boldsymbol{\rho}) = \sum_{t=t_k}^{t_k+T-1} \sum_{n\in N^{SII}(t_k)} \sum_{b\in\mathcal{B}} \sum_{i\in I^{SII}(t_k)} \rho^n_{bi,SII}(t) y_{bi,SII}(t) \left[Q_{SII}(t)\tau R^n_{bi,SII}(t) - \frac{V}{\eta} P^n_{bi,SII}(t)\right]. \tag{69}$$

Algorithm 5 Slice Request Admission and Channel Allocation Algorithm

Initialization:

In every LTS, N^{SI}_{min}, N^{SII}_{min}, $N^{SI}_{max} = N - N^{SII}_{min}$, and $N^{SII}_{max} = N - N^{SI}_{min}$ such that $N^{SI}_{min} \le N^{*SI} \le N^{SI}_{max}$ and $N^{SII}_{min} \le N^{*SII} \le N^{SII}_{max}$.
Compute R_s for all slice.
Set $S^+ = \emptyset$ and $S^- = \{SI, SII\}$.
$|N^{SI}(t_0)| = |N^{SII}(t_0)| = N/2$, $k = 0$.

Iteration:

$k = k + 1$.
while $|S^-| \ge 1$ **do**
  $s^* = \arg\max_{s\in S^-} R_s$.
  Check the feasibility of problem (37) if $S^+ = S^+ \cup s^*$.
  if feasible **then**
    Update $S^+ = S^+ \cup s^*$ and $S^- = S^- \setminus s^*$.
    if $|S^+| \ge 2$ **then**
      Assume that channel gain $\boldsymbol{G}(t_k)$ is the same with $\boldsymbol{G}(t_k - 1)$.
      Obtain $\boldsymbol{y}(t)$, $\boldsymbol{\rho}(t)$, and $\boldsymbol{P}_{SII}(t)$ by calling algorithm 4 based on $|N^{SI}(t_k)| = N^{SI}_{min}(t_k)$ and $|N^{SII}(t_k)| = N^{SII}_{max}(t_k)$.
      Calculate $q_1 = A_1(\boldsymbol{\rho}) - A_2(\boldsymbol{\rho})$.
      Obtain $\boldsymbol{y}(t)$, $\boldsymbol{\rho}(t)$, and $\boldsymbol{P}_{SII}(t)$ by calling algorithm 4 based on $|N^{SI}(t_k)| = N^{SI}_{max}(t_k)$ and $|N^{SII}(t_k)| = N^{SII}_{min}(t_k)$.
      if $q_1 > q_2$ **then**
        $|N^{*SI}(t_k)| = N^{SI}_{max}(t_k)$ and $|N^{*SII}(t_k)| = N^{SII}_{min}(t_k)$.
      else
        $|N^{*SI}(t_k)| = N^{SI}_{min}(t_k)$ and $|N^{*SII}(t_k)| = N^{SII}_{max}(t_k)$.
      end if
    end if
  else
    Update $S^- = S^- \setminus s^*$.
  end if
end while

Table I. To verify the performance of the DNSRA, we will consider the following schemes:

- NSD: The scheme does not have slice request admission [32].
- FCA: The scheme is that channel allocation is fixed [11].
- NSRA : The scheme that only optimizes slice request

TABLE I
SIMULATION PARAMETERS

Parameter	Value
N_0	-174 dBm/Hz
$P^{max}_{b,SII}$ (macro BS)	$1.5 \times \frac{P^{max}_{1,i}}{N}$ where $P^{max}_{b,SII}$ = 43 dBm
$P^{max}_{b,SII}$ (micro BS)	$1.5 \times \frac{P^{max}_{1,i}}{N}$ where $P^{max}_{b,SII}$ = 30 dBm
B	15 MHz
N	128
$f^{max}_{ib,SI}$	2 G cycle/s
$P^n_{ib,SI}$	0.1 W
L_i	2000 cycle/bit
τ	1 s
T	100 s
$\kappa_{ser,b}$	10^{-27}

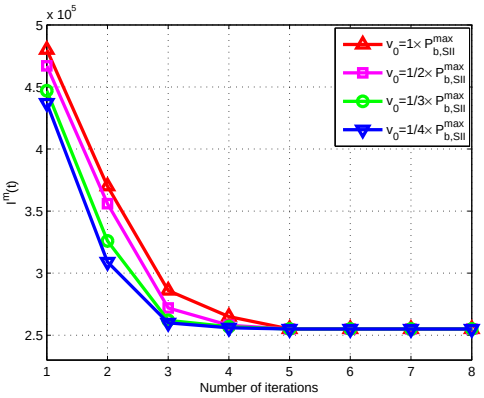


Fig. 2. Convergence of Algorithm 3.

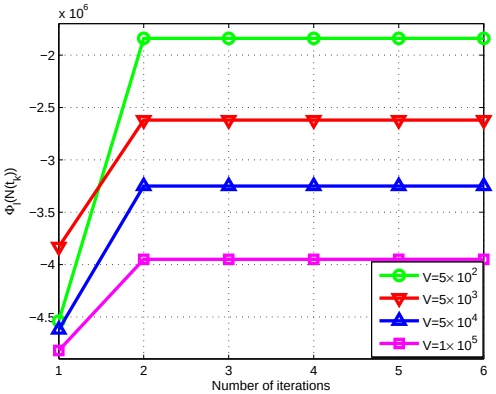


Fig. 3. Convergence of Algorithm 4.

admission [20].

B. Convergence of Algorithms 3 and 4

Fig. 2 plots $I^m(t)$ versus the number of iterations to show the convergence of Algorithm 3 under different parameter setting. As can be seen, $I^m(t)$ keeps decreasing after each iteration until convergence. Meanwhile, it is observed that it has a fast convergence rate insensitive to initial points. Fig. 3 displays the convergence of Algorithm 4 for different control parameter V. We can find that it converges very fast. Based on two figures, we can obtain that the proposed algorithms for solving the hard problem (37) are usually cost-efficient.

C. Revenue-Delay Tradeoff

Fig. 4 shows how different settings of traffic arrival rate λ and control parameter V affect the tradeoff between the average revenue and the average delay. For a given V, the average delay increases while the average revenue decreases as λ increases. For a given λ, both the average delay and the average revenue increase with V. This is because a larger V implies that the system puts less weight on the power consumption and more weight on the average delay from (36). When the system consumes less power, the short-term revenue increase, resulting in a increase of the average operator's revenue.

D. Performance of the Proposed Algorithm

In Fig. 5, we show the operator's average revenue vs. the latency requirement of the computation offloading slice $d^{SI}(t_k)$, while the data rate requirement of the data sharing slice $R^{SII}(t_k)$ is fixed at each LTS, i.e., $R^{SII}(t_k) = 3$Kb/s. We can observe that the operator's average revenue remains almost unchanged when $d^{SI}(t_k) < 0.3s$. However, the operator's average revenue rises sharply when $d^{SI}(t_k) > 0.3s$. This is because a looser QoS requirement will result in accepting more slice requests. When the number of allowed slice requests increases, the long-term revenue increases. Although the power consumption of the system increases, the long-term revenue dominates the operator's average revenue in this case. In addition, the operator's average revenue decreases as $d^{SI}(t_k)$ increases. That is because the long-term revenue decreases as $d^{SI}(t_k)$ increases.

Fig. 6 plots the operator's average revenue vs. the data rate requirement of the data sharing slice $R^{SII}(t_K)$, while the latency requirement of the computation offloading slice $d^{SI}(t_k)$ is fixed at each LTS, i.e., $d^{SI}(t_k) = 0.3$ s. From Fig. 6, we find that the higher the value of $R^{SII}(t_K)$, the less the operator's average revenue. That is because the power consumption of the system increases when the number of allowed slice requests increases, resulting in the reduction of short-term revenue. Although the long-term revenue increases with $R^{SII}(t_K)$, the short-term revenue dominates the operator's revenue in this case. In addition, we also find that the operator's average revenue drops sharply with the data rate, because the resource-limited systems can only accept one slice request.

In Figs. 7 and 8, we compare the average revenue obtained by applying different algorithms. In Fig. 7, the average revenue from all algorithms increases with the number of subcarriers because the power consumption of the system is reduced. It is observed that the proposed algorithm outperforms the others. Fig. 8 shows the average revenue vs. the number of mobile devices. From the figure, the average revenue decreases as the number of mobile devices increases. The reason is that the power consumption of the system increases with the number of mobile devices, which leads to a reduction in short-term revenue. Although the long-term revenue increases, its growth is slow due to the limited resources. Fig. 9 shows comparisons of the proposed algorithm with the NSD, FCA, and NSRA, respectively, under different values of η. We can observe that the average revenue of all algorithms decreases with the increasing η. That is because the system power consumption increases as η increases, i.e., the short-term revenue decreases.

V. Conclusions

In this paper, we studied the operator's average revenue maximization problem for network slicing in MEC systems. By utilizing the Lyapunov optimization technique, we developed a stochastic optimization framework, which formulates inhomogeneous traffic distributions. In particular, our problem is composed of slice request admission in the long-term slot and resource allocation in the short-term slot. Since network resources are limited, a binary integer variable is introduced to decided whether to accept or reject a slice request. Meanwhile, the operator's revenue is modeled as long-term revenue and short-term revenue. We designed an algorithm without requiring any prior-knowledge of traffic distributions, referred to as the DNSRA, to solve the problem. To decrease the computational complexity of directly solving the DNSRA, we decouple the optimization variables for efficient algorithm design. Simulation results show the proposed DNSRA can strike a flexible balance between the revenue and the average delay, and can significantly increase the operator's revenue against existing schemes.

Appendix

A. Proof of Theorem 1

The inequality

$$\{\max[Q-R,0]+A\}^2 \leq Q^2+R^2+A^2-2Q(R-A) \quad (70)$$

always hold. Based on this fact, squaring both sides of (8) yields

$$\begin{aligned} Q_{SI}(t_k+T-1)^2 \leq Q_{SI}(t_k)^2 &+ \left[a_s(t_k)\sum_{t=t_k}^{t_k+T-1}\sum_{i\in I^{SI}(t_k)} R_{i,SI}(t)\tau\right]^2 \\ &+ \left[a_s(t_k)\sum_{t=t_k}^{t_k+T-1}\sum_{i\in I^{SI}(t_k)} A_{i,SI}(t)\right]^2 - 2\sum_{t=t_k}^{t_k+T-1} Q_{SI}(t)\left[a_s(t_k)\right. \\ &\left.\sum_{i\in I^{SI}(t_k)} R_{i,SI}(t)\tau - a_s(t_k)\sum_{i\in I^{SI}(t_k)} A_{i,SI}(t)\right] \end{aligned} \quad (71)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2020.2992607, IEEE Transactions on Vehicular Technology

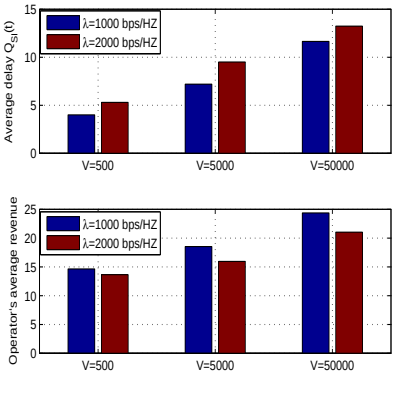


Fig. 4. Average operator's revenue and average delay $Q_{SI}(t_k)$ by the DNRSA under different parameter setting of traffic arrival rates λ and control parameter V.

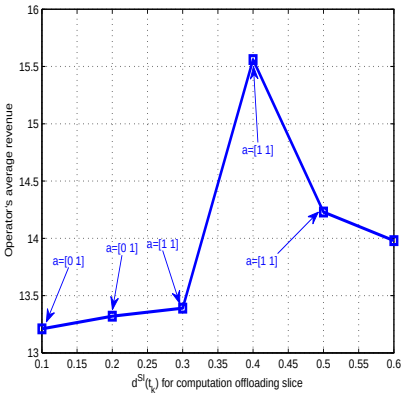


Fig. 5. The operator's average revenue vs. $d^{SI}(t_k)$..

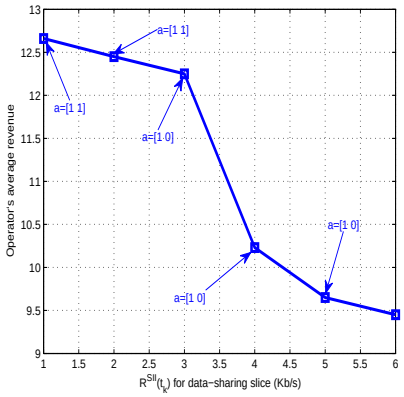


Fig. 6. The operator's average revenue vs. $R^{SII}(t_k)$.

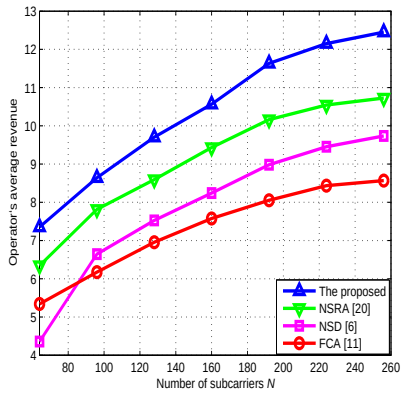


Fig. 7. The operator's average revenue vs. N.

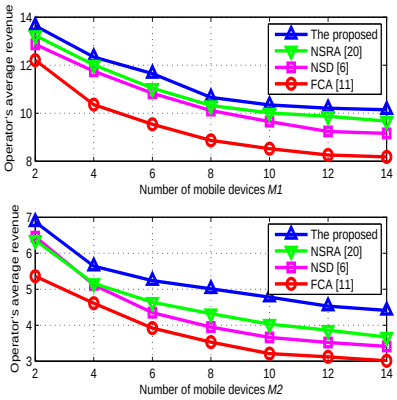


Fig. 8. The operator's average revenue vs. the number of mobile devices.

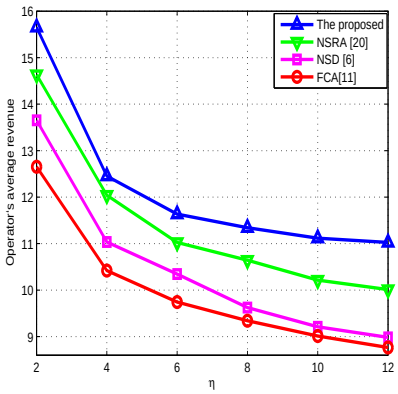


Fig. 9. The operator's average revenue vs. η.

For $Q_{SII}(t_k)$, similarly, we have

$$Q_{SII}(t_k+T-1)^2 \le Q_{SII}(t_k)^2 + \left[a_s(t_k)\sum_{t=t_k}^{t_k+T-1}\sum_{i\in I^{SII}(t_k)} R_{i,SII}(t)\tau\right]^2 + \left[a_s(t_k)\sum_{t=t_k}^{t_k+T-1}\sum_{i\in I^{SII}(t_k)} A_{i,SI}(t)\right]^2 - 2\sum_{t=t_k}^{t_k+T-1} Q_{SII}(t)\left[a_s(t_k)\sum_{i\in I^{SII}(t_k)} R_{i,SII}(t)\tau - a_s(t_k)\sum_{i\in I^{SII}(t_k)} A_{i,SII}(t)\right] \quad (72)$$

For $F_{SI}(t_k)$, similarly, we also have

$$F_{SI}(t_k+T-1)^2 \le F_{SI}(t_k)^2 + \left[a_s(t_k)\sum_{t=t_k}^{t_k+T-1}\sum_{b\in\mathcal{B}}\sum_{i\in I^{SI}(t_k)} y_{ib,SI}(t)f_{ib,SI}(t)\right]^2 + \max_{i\in I^{SI}(t)}\left[\sum_{t=t_k}^{t_k+T-1} L_i R_{i,SI}(t)\right]^2 - 2\sum_{t=t_k}^{t_k+T-1} F_{SI}(t)a_s(t_k)\left[\sum_{b\in\mathcal{B}}\sum_{i\in I^{SI}(t_k)} y_{ib,SI}(t)f_{ib,SI}(t) - \sum_{i\in I^{SI}(t_k)} L_i R_{i,SI}(t)\right] \quad (73)$$

By rearranging the above inequalities, we can obtain

$$\frac{Q_{SI}(t_k+T-1)^2 - Q_{SI}(t_k)^2}{2} \le \frac{1}{2}\left\{\left[a_s(t_k)\sum_{t=t_k}^{t_k+T-1}\sum_{i\in I^{SI}(t_k)} R_{i,SI}(t)\tau\right]^2 + \left[a_s(t_k)\sum_{t=t_k}^{t_k+T-1}\sum_{i\in I^{SI}(t_k)} A_{i,SI}(t)\right]^2\right\} - \sum_{t=t_k}^{t_k+T-1} Q_{SI}(t)a_s(t_k)\left[\sum_{i\in I^{SI}(t_k)}\right.$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2020.2992607, IEEE Transactions on Vehicular Technology

$$R_{i,SI}(t)\tau - \sum_{i\in I^{SI}(t_k)} A_{i,SI}(t)\Bigg] \tag{74}$$

$$\frac{Q_{SII}(t_k+T-1)^2 - Q_{SII}(t_k)^2}{2} \leq \frac{1}{2}\Bigg\{\Bigg[a_s(t_k)\sum_{t=t_k}^{t_k+T-1}\sum_{i\in I^{SII}(t_k)} R_{i,SII}(t)\tau\Bigg]^2 + \Bigg[a_s(t_k)\sum_{t=t_k}^{t_k+T-1} \sum_{i\in I^{SII}(t_k)} A_{i,SII}(t)\Bigg]^2\Bigg\} - \sum_{t=t_k}^{t_k+T-1} Q_{SII}(t)a_s(t_k)\Bigg[\sum_{i\in I^{SII}(t_k)} R_{i,SII}(t)\tau - \sum_{i\in I^{SII}(t_k)} A_{i,SII}(t)\Bigg] \tag{75}$$

$$\frac{F_{SI}(t_k+T-1)^2 + F_{SI}(t_k)^2}{2} \leq \frac{1}{2}\Bigg\{\Bigg[a_s(t_k)\sum_{t=t_k}^{t_k+T-1}\sum_{b\in\mathcal{B}}\sum_{i\in I^{SI}(t_k)} y_{ib,SI}(t)f_{ib,SI}(t)\Bigg]^2 + \max_{i\in I^{SI}(t)}\Bigg[\sum_{t=t_k}^{t_k+T-1} L_i R_{i,SI}(t)\Bigg]^2\Bigg\} - \sum_{t=t_k}^{t_k+T-1} F_{SI}(t)a_s(t_k)\Bigg[\sum_{b\in\mathcal{B}}\sum_{i\in I^{SI}(t_k)} y_{ib,SI}(t)f_{ib,SI}(t) - \sum_{i\in I^{SI}(t_k)} L_i R_{i,SI}(t)\Bigg] \tag{76}$$

Summing (74), (75) and (76), there is

$$\begin{aligned} &L(\mathbf{\Theta}(t_k+T)) - L(\mathbf{\Theta}(t_k)) \\ &\leq \frac{a_s(t_k)^2}{2}\Bigg\{\Bigg[\sum_{t=t_k}^{t_k+T-1}\sum_{i\in I^{SI}(t_k)} R_{i,SI}(t)\tau\Bigg]^2 + \Bigg[\sum_{t=t_k}^{t_k+T-1}\sum_{i\in I^{SI}(t_k)} A_{i,SI}(t)\Bigg]^2 + \Bigg[\sum_{t=t_k}^{t_k+T-1}\sum_{i\in I^{SII}(t_k)} R_{i,SII}(t)\tau\Bigg]^2 \\ &+ \Bigg[\sum_{t=t_k}^{t_k+T-1}\sum_{i\in I^{SII}(t_k)} A_{i,SII}(t)\Bigg]^2 + \sum_{t=t_k}^{t_k+T-1}\sum_{b\in\mathcal{B}}\sum_{i\in I^{SI}(t_k)} y_{ib,SI}(t)f_{ib,SI}(t) + \max_{i\in I^{SI}(t)}\Bigg[\sum_{t=t_k}^{t_k+T-1} L_i R_{i,SI}(t)\Bigg]^2\Bigg\} \\ &- \sum_{t=t_k}^{t_k+T-1} Q_{SI}(t)a_s(t_k)\Bigg[\sum_{i\in I^{SI}(t_k)} R_{i,SI}(t)\tau - \sum_{i\in I^{SI}(t_k)} A_{i,SI}(t)\Bigg] - \sum_{t=t_k}^{t_k+T-1} Q_{SII}(t)a_s(t_k)\Bigg[\sum_{i\in I^{SII}(t_k)} R_{i,SII}(t)\tau \\ &- \sum_{i\in I^{SII}(t_k)} A_{i,SII}(t)\Bigg] - \sum_{t=t_k}^{t_k+T-1} F_{SI}(t)a_s(t_k)\Bigg[\sum_{b\in\mathcal{B}}\sum_{i\in I^{SI}(t_k)} y_{ib,SI}(t)f_{ib,SI}(t) - \sum_{i\in I^{SI}(t_k)} L_i R_{i,SI}(t)\Bigg] \end{aligned} \tag{77}$$

Taking conditional expectation to the above inequality yields

$$\begin{aligned} &\triangle_T(\mathbf{\Theta}(t_k)) - V\mathbb{E}\{U(t_k)|\Theta(t_k)\} \\ &\leq C - \sum_{t=t_k}^{t_k+T-1} Q_{SI}(t)\mathbb{E}\Bigg\{a_s(t_k)\Bigg[\sum_{i\in I^{SI}(t_k)} R_{i,SI}(t)\tau - \sum_{i\in I^{SI}(t_k)} A_{i,SI}(t)\Bigg]\Bigg|\mathbf{\Theta}(t_k)\Bigg\} - \sum_{t=t_k}^{t_k+T-1} Q_{SII}(t)\mathbb{E}\Bigg\{a_s(t_k) \\ &\Bigg[\sum_{i\in I^{SII}(t_k)} R_{i,SII}(t)\tau - \sum_{i\in I^{SII}(t_k)} A_{i,SII}(t)\Bigg]\Bigg|\mathbf{\Theta}(t_k)\Bigg\} \\ &- \sum_{t=t_k}^{t_k+T-1} F_{SI}(t)\mathbb{E}\Bigg\{a_s(t_k)\Bigg[\sum_{b\in\mathcal{B}}\sum_{i\in I^{SI}(t_k)} y_{ib,SI}(t)f_{ib,SI}(t) - \sum_{i\in I^{SI}(t_k)} L_i R_{i,SI}(t)\Bigg]\Bigg|\mathbf{\Theta}(t_k)\Bigg\} - V\mathbb{E}\Bigg\{a_s(t_k)\,[\\ &G^{SI}(I^{SI}(t_k), d^{SI}(t_k)) + \eta^{-1}\sum_{t=t_k}^{t_k+T-1} r_{SI}^{revn}(t) + \eta^{-1}\sum_{t=t_k}^{t_k+T-1} r_{SII}^{revn}(t) + G^{SII}(I^{SII}(t_k), R^{SII}(t_k))|\mathbf{\Theta}(t_k)\Bigg\} \end{aligned} \tag{78}$$

where

$$\begin{aligned} C \geq \frac{a_s(t_k)^2}{2}\Bigg\{&\Bigg[\sum_{t=t_k}^{t_k+T-1}\sum_{i\in I^{SI}(t_k)} R_{i,SI}(t)\tau\Bigg]^2 + \Bigg[\sum_{t=t_k}^{t_k+T-1}\sum_{i\in I^{SI}(t_k)} A_{i,SI}(t)\Bigg]^2 + \Bigg[\sum_{t=t_k}^{t_k+T-1}\sum_{i\in I^{SII}(t_k)} R_{i,SII}(t)\tau\Bigg]^2 \\ &+ \Bigg[\sum_{t=t_k}^{t_k+T-1}\sum_{i\in I^{SII}(t_k)} A_{i,SI}(t)\Bigg]^2 + \sum_{t=t_k}^{t_k+T-1}\sum_{b\in\mathcal{B}}\sum_{i\in I^{SI}(t_k)} \\ &y_{ib,SI}(t)f_{ib,SI}(t) + \max_{i\in I^{SI}(t)}\Bigg[\sum_{t=t_k}^{t_k+T-1} L_i R_{i,SI}(t)\Bigg]^2\Bigg\} \end{aligned} \tag{79}$$

This completes the proof of Theorem 1.

REFERENCES

[1] H. Zhang, N. Liu, X. Chu, K. Long, A. Aghvami, and V. C. M. Leung, "Network slicing based 5G and future mobile networks: Mobility, resource management, and challenges," *IEEE Commun. Mag.*, vol. 55, no. 8, pp. 138–145, Aug. 2017.

[2] K. Samdanis, X. Costa-Perez, and V. Sciancalepore, "From network sharing to multi-tenancy: The 5G network slice broker," *IEEE Commun. Mag.*, vol. 54, no. 7, pp. 32–39, Jul. 2016.

[3] W. Wei, H. Gu, K. Wang, X. Yu, and X. Liu, "Improving cloud-based iot services through virtual network embedding in elastic optical inter-dc networks," *IEEE Internet of Things Journal*, vol. 6, no. 1, pp. 986–996, Feb 2019.

[4] *Description of Network Slicing Concept Version 1.0*, NGMN, Frankfurt, Germany, Jan. 2016.

[5] Z. Dawy, W. Saad, A. Ghosh, J. G. Andrews, and E. Yaacoub, "Toward massive machine type cellular communications," *IEEE Wireless Commun.*, vol. 24, no. 1, pp. 120–128, Feb. 2017.

[6] Y. Guo, F. R. Yu, J. An, K. Yang, C. Yu, and V. C. M. Leung, "Adaptive bitrate streaming in wireless networks with transcoding at network edge using deep reinforcement learning," *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 3879–3892, 2020.

[7] J. Tang, B. Shim, and T. Q. S. Quek, "Service multiplexing and revenue maximization in sliced C-RAN incorporated with URLLC and multicast embb," *IEEE J. Sel. Areas in Commun.*, vol. 37, no. 4, pp. 881–895, Apr. 2019.

[8] H. Halabian, "Distributed resource allocation optimization in 5G virtualized networks," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 3, pp. 627–642, Mar. 2019.

[9] F. Guo, F. R. Yu, H. Zhang, H. Ji, M. Liu, and V. C. M. Leung, "Adaptive resource allocation in future wireless networks with blockchain and mobile edge computing," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1689–1703, 2020.

[10] J. Du, L. Zhao, J. Feng, and X. Chu, "Computation offloading and resource allocation in mixed fog/cloud computing systems with min-max fairness guarantee," *IEEE Trans. Commun.*, vol. 66, no. 4, Apr. 2018.

[11] Y. Xiao and M. Krunz, "Dynamic network slicing for scalable fog computing systems with energy harvesting," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 12, Dec. 2018.

[12] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," *IEEE Internet Things J.*, vol. 3, no. 5, pp. 637–646, Oct. 2016.

[13] J. Du, F. R. Yu, G. Lu, J. Wang, J. Jiang, and X. Chu, "Mec-assisted immersive vr video streaming over terahertz wireless networks: A deep reinforcement learning approach," *IEEE Internet of Things Journal*, Accepted, 2020.

[14] J. Feng, F. R. Yu, Q. Pei, J. Du, and L. Zhu, "Joint optimization of radio and computational resources allocation in blockchain-enabled mobile edge computing systems," *IEEE Transactions on Wireless Communication*, be accepted, 2020, doi: 10.1109/TWC.2020.2982627.

[15] J. Feng, F. R. Yu, Q. Pei, X. Chu, J. Du, and L. Zhu, "Cooperative computation offloading and resource allocation for blockchain-enabled mobile edge computing: A deep reinforcement learning approach," *IEEE Internet of Things Journal*, be accepted, 2019, doi: 10.1109/JIOT.2019.2961707.

[16] L. Liu, C. Chen, Q. Pei, S. Maharjan, and Y. Zhang, "Vehicular edge computing and networking: A survey," *arXiv preprint arXiv:1908.06849*, 2019.

[17] Y. Sun, M. Peng, S. Mao, and S. Yan, "Hierarchical radio resource allocation for network slicing in fog radio access networks," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 3866–3881, Apr. 2019.

[18] P. Zhao, H. Tian, S. Fan, and A. Paulraj, "Information prediction and dynamic programming-based ran slicing for mobile edge computing," *IEEE Wireless Commun. Letters*, vol. 7, no. 4, pp. 614–617, Aug. 2018.

[19] T. Sanguanpuak, N. Rajatheva, D. Niyato, and M. Latva-aho, "Network slicing with mobile edge computing for micro-operator networks in beyond 5G," in *2018 21st International Symposium on Wireless Personal Multimedia Communications (WPMC)*, Nov. 2018, pp. 352–357.

[20] N. Van Huynh, D. Thai Hoang, D. N. Nguyen, and E. Dutkiewicz, "Optimal and fast real-time resource slicing with deep dueling neural networks," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 6, pp. 1455–1470, Jun. 2019.

[21] Y. Wang, Y. Zhang, M. Sheng, and K. Guo, "On the interaction of video caching and retrieving in multi-server mobile-edge computing systems," *IEEE Wireless Communications Letters*, vol. 8, no. 5, pp. 1444–1447, Oct 2019.

[22] Y. Song, Y. Fu, F. R. Yu, and L. Zhou, "Blockchain-enabled internet of vehicles with cooperative positioning: A deep neural network approach," *IEEE Internet of Things J.*, to be published, 2020.

[23] J. Kwak, Y. Kim, J. Lee, and S. Chong, "Dream: Dynamic resource and task allocation for energy minimization in mobile cloud systems," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 12, pp. 2510–2523, Dec. 2015.

[24] R. Zhang, F. R. Yu, J. Liu, R. Xie, and T. Huang, "Blockchain-incentivized D2D and mobile edge caching: A deep reinforcement learningapproach," *IEEE Network*, accepted, 2019.

[25] Y. Mao, J. Zhang, S. H. Song, and K. B. Letaief, "Stochastic joint radio and computational resource management for multi-user mobile-edge computing systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5994–6009, Sep. 2017.

[26] Y. Kim, H.-W. Lee, and S. Chong, "Mobile computation offloading for application throughput fairness and energy efficiency," *IEEE Trans. Wireless Commun.*, vol. 18, no. 1, pp. 3–19, Jan. 2019.

[27] M. J. Neely, "Stochastic network optimization with application to communication and queueing systems," *Synthesis Lectures on Communication Networks*, CA, UAS: Morgan & Claypool Publishers, 2010.

[28] S. Boyd, "Sequential convex programming," Lecture Slides and Notes. [Online]. Available: http://www.stanford.edu/class/ee364b/lectures.html.

[29] M. Grant, S. Boyd, and Y. Ye, *CVX: MATLAB Software for Disciplined Convex Programming*, [Online]. Available: http://cvxr.com/cvx/.

[30] Z. Shen, J. G. Andrews, and B. L. Evans, "Adaptive resource allocation in multiuser ofdm systems with proportional rate constraints," *IEEE Trans. Wireless Commun.*, vol. 4, no. 6, pp. 2726–2737, Nov. 2005.

[31] Y. Li, M. Sheng, Y. Sun, and Y. Shi, "Joint optimization of bs operation, user association, subcarrier assignment, and power allocation for energy-efficient hetnets," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 12, pp. 3339–3353, Dec. 2016.

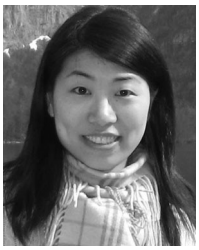
[32] Q. Ye, W. Zhuang, S. Zhang, A. Jin, X. Shen, and X. Li, "Dynamic radio resource slicing for a two-tier heterogeneous wireless network," *IEEE Trans. Veh. Technol.*, vol. 67, no. 10, pp. 9896–9910, Oct. 2018.

Jie Feng is currently pursuing the Ph.D. degree in Communication and Information System at Xidian University, Xian, China. She is also with Carleton University as Visiting Ph.D Student since January 2019. Her current research interests include mobile edge computing, Blockchain, deep reinforcement learning, Device to Device communication, resource allocation and convex optimization and stochastic network optimization.

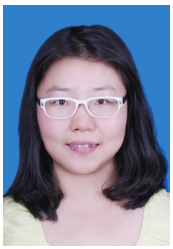
Qingqi Pei received his B.S., M.S. and Ph.D. degrees in Computer Science and Cryptography from Xidian University, in 1998, 2005 and 2008, respectively. He is now a Professor and member of the State Key Laboratory of Integrated Services Networks, also a Professional Member of ACM and Senior Member of IEEE, Senior Member of Chinese Institute of Electronics and China Computer Federation. His research interests focus on privacy preserving, blockchain and edge computing security.

F. Richard Yu (S00-M04-SM08-F18) received the PhD degree in electrical engineering from the University of British Columbia (UBC) in 2003. From 2002 to 2006, he was with Ericsson (in Lund, Sweden) and a start-up in California, USA. He joined Carleton University in 2007, where he is currently a Professor. He received the IEEE Outstanding Service Award in 2016, IEEE Outstanding Leadership Award in 2013, Carleton Research Achievement Award in 2012, the Ontario Early Researcher Award (formerly Premiers Research Excellence Award) in 2011, the Excellent Contribution Award at IEEE/IFIP TrustCom 2010, the Leadership Opportunity Fund Award from Canada Foundation of Innovation in 2009 and the Best Paper Awards at IEEE ICNC 2018, VTC 2017 Spring, ICC 2014, Globecom 2012, IEEE/IFIP TrustCom 2009 and Intl Conference on Networking 2005. His research interests include wireless cyber-physical systems, connected/autonomous vehicles, security, distributed ledger technology, and deep learning.

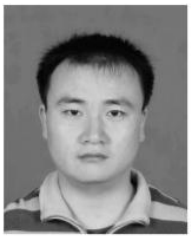
He serves on the editorial boards of several journals, including Co-Editor-in-Chief for Ad Hoc & Sensor Wireless Networks, Lead Series Editor for IEEE Transactions on Vehicular Technology, IEEE Transactions on Green Communications and Networking, and IEEE Communications Surveys & Tutorials. He has served as the Technical Program Committee (TPC) Co-Chair of numerous conferences. Dr. Yu is a registered Professional Engineer in the province of Ontario, Canada, a Fellow of the Institution of Engineering and Technology (IET), and a Fellow of the IEEE. He is a Distinguished Lecturer, the Vice President (Membership), and an elected member of the Board of Governors (BoG) of the IEEE Vehicular Technology Society.



Xiaoli Chu (M06CSM15) received the B.Eng. degree in electronic and information engineering from Xian Jiao Tong University, Xian, China, in 2001, and the Ph.D. degree in electrical and electronic engineering from the Hong Kong University of Science and Technology, Hong Kong, in 2005. She is a Senior Lecturer with the Department of Electronic and Electrical Engineering, University of Sheffield, Sheffield, U.K. From September 2005 to April 2012, she was with the Centre for Telecommunications Research, Kings College London. She has published more than 100 peer-reviewed journal and conference papers. She is the Lead Editor/author of the book Heterogeneous Cellular Networks: Theory, Simulation and Deployment (Cambridge University Press, 2013) and the book 4G Femtocells: Resource Allocation and Interference Management (Springer 2013).



Jianbo Du received the B.S. degree and M.S. degree from Xi'an University of Posts and Telecommunications in 2007 and 2013, respectively, and the Ph.D. in communication and information systems at Xidian University, Xian, Shaanxi, China, in 2018. She is now a teacher with the department of Communication and Information Engineering, Xian University of Posts and Telecommunications. Her research interests include mobile edge computing, resource management, NOMA, deep reinforcement learning, convex optimization, stochastic network optimization and heuristic algorithms and their applications in wireless communications.



Li Zhu received the Ph.D. degree in traffic control and information engineering from Beijing Jiaotong University, Beijing, China, in 2012. He is currently a Faculty Member at Beijing Jiaotong University and a Visiting Scholar at Carleton University, Ottawa, ON, Canada, and The University of British Columbia, Vancouver, BC, Canada. His research interests include intelligent transportation systems, train-ground communication technology in communication base train ground communication systems, and cross layer design in train-ground communication systems.