Dynamic Virtual Resource Allocation for 5G and Beyond Network Slicing

FEI SONG [1] (Student Member, IEEE), JUN LI [1,2] (Senior Member, IEEE), CHUAN MA [1],
YIJIN ZHANG [1,3] (Senior Member, IEEE), LONG SHI [4] (Member, IEEE), AND DUSHANTHA NALIN K.
JAYAKODY [2,5] (Senior Member, IEEE)

[1]School of Electronic and Optical Engineering, Nanjing University of Science Technology, Nanjing 210094, China
[2]School of Computer Science and Robotics, National Research Tomsk Polytechnic University, Tomsk 634050, Russia
[3]State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, China
[4]Science and Math Cluster, Singapore University of Technology and Design, Singapore 487372, Singapore
[5]Centre for Telecommunications Research, School of Engineering, Sri Lanka Technical Campus, Padukka 10500, Sri Lanka

CORRESPONDING AUTHOR: JUN LI (e-mail: jun.li@njust.edu.cn)

ABSTRACT The fifth generation and beyond wireless communication will support vastly heterogeneous services and user demands such as massive connection, low latency and high transmission rate. Network slicing has been envisaged as an efficient technology to meet these diverse demands. In this paper, we propose a dynamic virtual resources allocation scheme based on the radio access network (RAN) slicing for uplink communications to ensure the quality-of-service (QoS). To maximum the weighted-sum transmission rate performance under delay constraint, formulate a joint optimization problem of subchannel allocation and power control as an infinite-horizon average-reward constrained Markov decision process (CMDP) problem. Based on the equivalent Bellman equation, the optimal control policy is first derived by the value iteration algorithm. However, the optimal policy suffers from the widely known curse-of-dimensionality problem. To address this problem, the linear value function approximation (approximate dynamic programming) is adopted. Then, the subchannel allocation *Q-factor* is decomposed into the per-slice *Q-factor*. Furthermore, the *Q-factor* and Lagrangian multipliers are updated by the use of an online stochastic learning algorithm. Finally, simulation results reveal that the proposed algorithm can meet the delay requirements and improve the user transmission rate compared with baseline schemes.

INDEX TERMS Network slicing, RAN slicing, constrained Markov decision process (CMDP), resource allocation.

I. INTRODUCTION

Driven by an astounding increase of mobile users and their bandwidth-hungry and varied applications, wireless data traffic will growth exponentially in the next couple years [1], [2]. The fifth generation (5G) and next-generation wireless communication networks will go beyond bringing only high transmission rates for mobile users, as it will support a wider communication ecosystem for the machine-type communications, Internet of Vehicles and Internet of Things [3]. To promote this evolution, the 5G and beyond wireless communications are envisaged to be the cornerstone for abundant emerging applications [4]. Due to the diversity of these applications and the complexity of the real environment [5], the 5G and beyond wireless communications try to guarantee the heterogeneous user quality of service (QoS) requirements [6]. For instance, it is required to provide extremely high reliability and low latency communications for some vertical industries, like industrial automation control system and Internet of Vehicles, in order to satisfy the stringent QoS demands. Reexamining the networking technologies and

network architecture for 5G and beyond networks is necessary according to the varied and customized QoS requirements. To establish the service classification, the International Telecommunication Union (ITU) has characterized three types of users [7]: 1) Enhanced mobile broadband (eMBB) communications, which demand high transmission rates; 2) Ultra-reliable low-latency communications (URLLC), devised to satisfy stringent latency and reliability requirements; 3) Massive machine type communications (mMTC), aimed to support massive devices, each one uploading very short data packets.

Network slicing is a vital network architecture innovation in 5G, which is also envisioned to play an important role in the next generation [8], [9]. Network slicing permits multiple independent and isolated virtual networks to coexist in the same physical network (PN) infrastructure, the virtual network is defined as slice. Software-defined networking (SDN) and network functions virtualization (NFV) are key technologies to implement network slicing for accommodating new services with wide different requirements over the same PN [10]. On the one hand, the slices are established through an abstract set of control logic and resources provided by the SDN controller [11]. In addition, with SDN the network slicing enables the share of the same PN resources among different tenants. On the other hand, NFV is developed to solve the shortage of existing special communication equipment. The core of NFV is virtual network functions, running on virtual machines of general servers without requiring dedicated hardware [12]. Generally, the concept of network slicing prevails due to the following enticing advantages: First, network slicing supports multi-tenancy through the virtual networks multiplexing, which leads to the same physical infrastructure can be shared by several virtual network operators [13]. Second, network slicing can realize differentiated service, ensure the service level agreement for every service type [4]. Third, network slicing enhances the adaptability and flexibility of network management as slices can be created as needed and changed on-demand [14].

Fig. 1 describes the architecture of a 5G and beyond network scenario with network slicing. In this model, the underlying physical infrastructure are abstracted into network slices, which are managed and coordinated uniformly by orchestrator. The physical resources, such as core networks (CNs), radio access networks (RANs), are divided into several logical parts, thereby forming different network slices based on the user requirements. The slice layer contains multiple slices dedicated to heterogeneous services and runs on the top of underlying layer. By providing flexible and scalable network architecture to guarantee various QoS demands of heterogeneous services, RAN slicing is considered to be one of the most promising technologies in 5G and beyond networks. RAN slicing can provide customized services for the isolated logical networks by dividing the same PN into multiple isolated logical networks. On the basis of sharing PN, RAN slicing is an economic and high-efficient network management scheme. A research reported the global
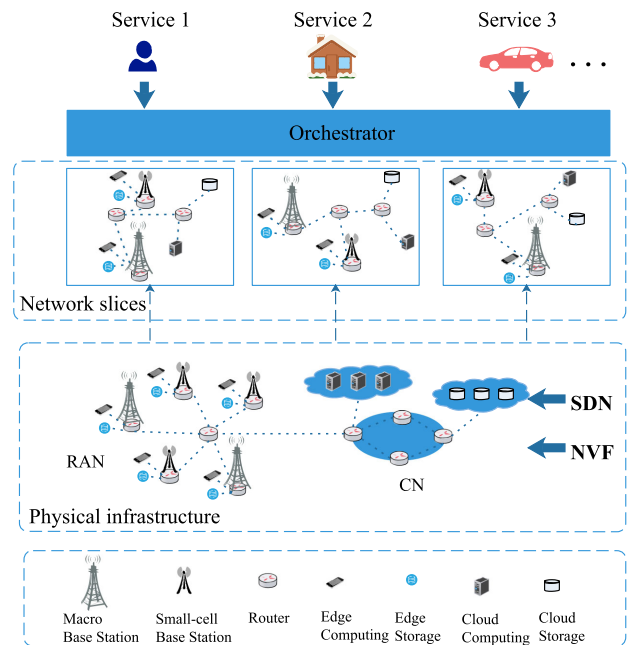

FIGURE 1. An illustration of network slicing based 5G and beyond system architecture.

capital expenditure and operational expenditure can save nearly 60 billion through RAN sharing by 2021 [15]. In practice, the 3GPP has implemented expansive research for 5G network slicing [16], [17].

Due to the properties of RAN slicing, the key problem of RAN slicing is how to flexibly and efficiently allocate the network resources to guarantee the varied QoS demands. In RAN slicing, operators allocate resources efficiently and flexibly at the network level according to different performance requirements of users. These resources include not only wireless resources, but also computing and cache resources. Resource allocation technology can realize the efficient utilization and improve the utilization rate of resources, and integrate different slices. Depending on the environment the resource allocation of network slicing is classified as static allocation and dynamic allocation [18]. Static resource allocation means that, after the resource allocation and mapping strategy of the network slicing is determined, it will remain the same regardless of how the environment changes. Dynamic resource management [19] refers to the ability to sense the changes in the environment, and then dynamically adjust resource allocation strategies to optimize the quality of communication services.

Network slicing has also attracted widespread research interest in academia. The authors of [20] studied the resource management of wireless access network slices from the perspective of user access control and wireless bandwidth allocation. The optimization problem of this paper is to minimize bandwidth consumption under the QoS and resource constraints. Finally, the Lagrangian decomposition theory is used to solve the optimization problem. In [21], a cell planning scheme for network slicing was proposed to maximize

the resource efficiency under different QoS requirements of different mobile services for wireless communications. This scheme jointly optimized the inter-slice resource allocation, the distribution of cells operated on different slices, and the users allocation for cells. Caballero *et. al.* [22] proposed a network slicing framework, including admission control, resource management and user dropping. However, the wireless channel set in [20] and [21] are fixed. In addition, the resource allocation scheme in [22] was fixed and will not change with the topology. In fact, the wireless channel is always interfered by the environment and the channel quality changes dynamically with time. And the changes in the environment may cause the server to stop working when the policy is fixed. Although the above papers studied and analyzed the resource allocation and optimization problem of network slicing, with the exponential growth of mobile terminal data traffic, the increase of device connections and the rise of multimedia and other emerging services, the static resource allocation scheme can not change the resource allocation strategy in real time. Not only the resource utilization is low, but also the performance requirements of all slices cannot be met, so the static resource allocation method cannot meet the current network environment.

For this reason, the static resource allocation cannot meet the dynamic performance requirements. Driven by this issue, [23] investigated the optimal dynamic resource allocation and mode selection to minimize the average delay with a dropping probability constraint in D2D communications. This paper formulated the resource control problem in D2D communications into a constrained Markov decision process (CMDP). The authors of [24] focused on the optimization of dynamic power control policy. The optimization problem was formulated as a CMDP to minimize the weighted sum of average delay and average transmit power. However, aforementioned works in [23] and [24] only investigated one type of service requirement for resource allocation.

In general, the resources need to be allocate dynamically according to the different QoS, and network slicing can meet this demand. A systematic approach of dynamic resource allocation is via MDP and dynamic programming techniques. The authors of [25] focused on the fog computing system, wherein a orchestrator coordinated workload distribution among local fog nodes. The energy and computing resources were dynamically allocated to slices. In particular, the partially observable MDP (POMDP) was developed in this paper to maximum the total amount of offloaded workload. The authors of [26] proposed a virtual resource scheduling scheme in 5G slicing network with the non-orthogonal multiple access system. To solve the problem of virtual resource allocation of downlink RAN slicing, the problem was described as a CMDP aiming at maximizing the total user rate. Therefore, we focus on the dynamic virtual resource allocation for RAN slicing aided uplink communication systems. The subchannel allocation and power control policy is proposed to maximum transmission rate with the heterogeneous delay constrains. The rate maximization problem is formulated as an infinite-horizon average

reward CMDP. The conventional value iteration and policy iteration will leads to the curse of dimensionality. To reduce the complexity, we use linear value function approximation and distributed online stochastic learning to simplify the optimization problem with the subchannel allocation and power control. The contributions of this paper are summarized as follows.

- A dynamic virtual resource allocation algorithm is proposed for CMDP-based RAN slicing. The resource allocation scheme is dynamically adjusted to power control and subchannel allocation according to the continuous interaction process with external environment. Due to the specific structure of this problem, we derive a reduced-state equivalent Bellman equation.

- To further reduce the computational complexity, we use approximate dynamic programming to develop the proposed algorithm. Then, we propose an equivalent Bellman equation in terms of subchannel allocation *Q-factor* to solve the CMDP. Especially, the *Q-factor* is approximated as the sum of per-slice *Q-factor*. Furthermore, we put forth a distributed online learning algorithm based on channel state information (CSI), queue state information (QSI) and energy state information (ESI) to optimize the per-slice *Q-factor* and the Lagrange multipliers (LM).

The rest of the paper is organized as follows. The system model is briefly described in Section II. The problem formulation and optimal solution are proposed in Section III. In Section IV, we develop a low-complexity distributed online learning algorithm. Simulation results are shown in Section V. Finally, Section VI concludes this paper. The main notations used in this paper are listed in Table 1.

II. SYSTEM MODEL
In this section, we first present physical layer model, queue dynamic model and energy dynamic model of RAN slicing, and then elaborate the subchannel allocation and power control policy and data dropping and delay constraints.

A. PHYSICAL LAYER MODEL
Here, we develop a uplink wireless communication scenario for an RAN slicing consisting of a single BS and heterogeneous user equipments (UEs). The network first determines the service types of UEs, and then UEs are assigned to the related slices based on the QoS demands. The time dimension is divided into multiple time slots and each slot lasts Δ seconds. Given the communication scenario for RAN-slicing uplink wireless communication, in which three types of UEs (eMBB, mMTC and URLLC) access the same BS, as shown in Fig. 2.

The wireless resources are divided into M slices to support M different types of services, each slice accommodates k_m active UEs, $m \in \{1, 2, \ldots, M\}$. The BS locate in the center and the UEs are uniformly distributed in a disk with radius D. The whole uplink bandwidth of the BS is divided into N subchannels, the bandwidth of each subchannel is B.

TABLE 1. Main Notations and Their Definitions

Category	Notations	Definitions
Constant	B	Subchannel bandwidth
	$D_m^{\max}$	The maximum delay of slice m
	$B_m^Q, \beta_m, \lambda_m^Q$	The buffer capacity of UE in slice m, the probability that the UE in slice m generate a new queue task, the average data arrival rate of the UE in slice m
	B_m^E, λ_m^E	The battery capacity of UE in slice m, the average energy arrival rate of the UE in slice m
Variable	$M,$	The number of network slices
	N	The number of subchannels
	k_m	The active UEs of slice m
	$h_{i,m}^n$	The channel from the UE i in slice m connect with subchannel n to the BS
	$\phi, P_{i,m}$	The FPC factor, the transmit power of the UE i in slice m
	$R_{i,m}, \bar{R}_m$	The transmission rate of the UE i in slice m, the total rate of the slice m
	$D_{i,m}, \bar{D}_m$	The delay of the UE i in slice m, the average delay of the slice m
	$Q_{i,m}, A_{i,m}^Q, L_{i,m}^Q$	The queue length, the amount of data arrived and the amount of transmitted packets of UE i in slice m
	$E_{i,m}, A_{i,m}^E, L_{i,m}^E$	The energy length, the amount of energy arrived and the amount of transmission energy consumption of UE i in slice m
State	$\mathbf{S}_t$	The global system state at time slot t, $\mathbf{S}_t = (\mathbf{H}_t, \mathbf{Q}_t, \mathbf{E}_t)$
	$\mathbf{H}_t$	The channel at time slot t
	$\mathbf{Q}_t$	The queue at time slot t
	$\mathbf{E}_t$	The energy at time slot t
Action	$\Omega = (\Omega_c, \Omega_p)$	A stationary subchannel allocation and power control policy
	$\Omega_c(s) = \{c_{i,m}^n\}_{i \in k_m, m \in M, n \in N},$	The adjustment action of subchannel allocation
	$\Omega_p(s) = \{\phi_{i,m}\}_{i \in k_m, m \in M}$	The adjustment action of power control

FIGURE 2. System scenario for RAN-slicing uplink wireless communication, with three example network slices.

We focus on the instantaneous channel gain [27], which consists of the fast-fading and path loss on each subchannel n. The channel state between the UE in slice m and the subchannel n of BS is denoted by $h_{i,m}^n \in \mathcal{H}$, where i is the UE index in slice m and $\mathcal{H}$ denotes the finite discrete complex channel state space. Let $h_{i,m}^n = g_{i,m}^n \sqrt{A d_{i,m}^{-\alpha}}$, where $g_{i,m}^n$ is the fading that the UE i in slice m connect with the subchannel n, $d_{i,m}$ denotes the distance from the UE i in the slice m to the BS, α is the path loss factor, and A represents the free space power received when the reference distance is 1m. We have the following assumptions on CSI.

Assumption 1: The fading follows a Rayleigh distribution with variance σ_g^2 and mean $\bar{g}$, and stays the same in a time slot.

The fading distribution of each UE is independent identically distributed (i.i.d.) for each time slot and independent with respect to each subchannel [28]. During the communication session, the path loss stays the same. Therefore, the CSI stays the same in a time slot. ∎

We define the transmit power for UE i in slice m as $P_{i,m}$. Given the baseline power P_m of slice m, $P_{i,m}$ is given by

$$P_{i,m} = P_m (A d_{i,m}^{-\alpha})^{-\phi}, \tag{1}$$

where ϕ is the fractional power control (FPC) factor.

Let $c_{i,m}^n \in \{0, 1\}$ denote the subchannel allocation for the UE i in slice m on the nth subchannel. Concretely, $c_{i,m}^n = 1$ means that the nth subchannel is used by the UE i in slice m, and $c_{i,m}^n = 0$ otherwise. Specifically, considering that the UE i in slice m is scheduled on subchannel n, the received signal at the BS from the UE i in slice m on subchannel n is given by

$$y_{i,m}^n = h_{i,m}^n \sqrt{P_{i,m}} x_{i,m} + \omega_n. \tag{2}$$

Here, $x_{i,m}$ is the transmit signal of UE i in slice m. ω_n denotes a zero-mean complex additive white Gaussian noise (AWGN) random variable on the subchannel n with variance σ_n^2.

Hence, the transmission rate of the UE i in slice m on subchannel n is given by

$$R_{i,m}^n = B \log_2 \left(1 + \frac{P_{i,m} |h_{i,m}^n|^2}{\sigma_n^2} \right). \tag{3}$$

B. QUEUE DYNAMIC MODEL

Each UE continuously buffer the arriving packets by maintains a traffic queue. To reduce the queueing time waiting in the buffer before transmission, it is crucial to properly schedule the queues to maximize the system capacity. The generation of new queue task for each UE in slice m follows independent Bernoulli distribution with probability β_m, which

represents the probability of UEs in slice m that generate new queue tasks at the current time slot. In addition, the packet arrival process of UE in slice m is assumed to i.i.d. for each time slot and follows Poisson distribution with the average arrival rate λ_m^Q. To be more formal, we assume that the packet arrivals take place around the slot boundaries.

In time slot t, the amount of data arrived at UE is denoted as $A_{i,m}^Q(t)$ and the current queue length of UE i in slice m is denoted as $Q_{i,m}(t)$. The subsequent arriving packet will be dropped if the queue length reached the buffer capacity B_m^Q. Hence, the queue dynamics can be expressed as

$$Q_{i,m}(t+1) = \min\left\{\left[Q_{i,m}(t) + A_{i,m}^Q(t) - L_{i,m}^Q(t)\right]^+, B_m^Q\right\}, \tag{4}$$

where $x^+ \triangleq \max\{x, 0\}$, and $L_{i,m}^Q(t) = R_{i,m}(t)\Delta$ denotes the amount of transmitted packets of UE i in slice m during time slot t. Note that $A_{i,m}^Q(t) = 0$ when the new queue task is not generated in time slot t.

Define the average queue length of UE i in slice m is $\overline{Q_{i,m}}$, which is given by

$$\overline{Q_{i,m}} = \lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T} E^\Omega[Q_{i,m}(t)] = E^\Omega[Q_{i,m}], \tag{5}$$

where $E[\cdot]$ means the expectation operation.

C. ENERGY DYNAMIC MODEL

Define the energy arriving at UE in time slot t as $A_{i,m}^E(t)$. B_m^E denotes the battery capacity of slice m. The energy arrival $A_{i,m}^E(t)$ is assumed to i.i.d. for each scheduling slot and independent follows a general distribution with average arrival rate λ_m^E. At time slot t, $L_{i,m}^E(t)$ unit of energy is consumed for packet transfer. Therefore, the energy dynamics can be expressed as

$$E_{i,m}(t+1)$$
$$= \min\left\{\left[E_{i,m}(t) + A_{i,m}^E(t) - L_{i,m}^E(t)\right]^+, B_m^E\right\}, \tag{6}$$

where $L_{i,m}^E(t) = P_{i,m}(t)\Delta$ denotes transmission energy consumption of UE i in slice m at time slot t.

D. SUBCHANNEL ALLOCATION AND POWER CONTROL POLICY

Due to the system states change dynamically with time, the resource manager dynamically adjust subchannel allocation and power control strategy in real-time based on global CSI (GCSI), global QSI (GQSI) and global ESI (GESI). At the beginning of the tth time slot, the resource manager determines the subchannel allocation, and perform power control in terms of a stationary subchannel allocation and power control policy described as follow.

Definition 1: The mapping from the system state $\mathbf{S} \in \mathcal{S}$ to the subchannel allocation and power control actions is regarded as a stationary subchannel allocation and power control policy $\Omega = (\Omega_c, \Omega_p)$, where $\Omega_c(\mathbf{S}) = \{c_{i,m}^n\}_{i \in k_m, m \in M, n \in N} \in \mathcal{C}$, and $\Omega_p(\mathbf{S}) = \{\phi_{i,m}\}_{i \in k_m, m \in M} \in \mathcal{P}$.

Based on (1), the FPC factor ϕ can adjust the transmit power of UEs to meet the slice QoS requirements. Therefore, the action of power control is given by $\phi_{i,m} \in (0, 1]$. ∎

Note that the instantaneous transmission rate is influenced by the subchannel allocation and power control at time slot t, the total rate of UE i in slice m is given by $R_{i,m}(t) = \sum_{n=1}^N c_{i,m}^n(t)R_{i,m}^n(t)$. Hence, the total transmission rate for slice m is given by

$$R_m(t) = \sum_{n=1}^N \sum_{i=1}^{k_m} c_{i,m}^n(t)R_{i,m}^n(t)$$
$$= \sum_{n=1}^N \sum_{i=1}^{k_m} c_{i,m}^n(t)B\log_2\left(1 + \frac{P_{i,m}(t)|h_{i,m}^n(t)|^2}{\sigma_n^2(t)}\right). \tag{7}$$

In order to ensure that transmission is not interrupted by energy exhaustion, the power control policy Ω_p satisfies the power consumption constraint

$$E_{i,m}(t) > \min\left\{P_{i,m}(t)\frac{Q_{i,m}(t)}{R_{i,m}(t)}, P_{i,m}(t) \cdot \Delta\right\}. \tag{8}$$

Therefore, the UE upload queue tasks only when energy constraints are met. In this study, we redefine the immediate transmission rate as

$$R_{i,m}(t) = \begin{cases} R_{i,m}(t), & \text{if (8) is satisfied,} \\ 0, & \text{otherwise.} \end{cases} \tag{9}$$

E. DATA DROPPING AND DELAY

We consider the dropping probability in the system model when the traffic queue buffer size is limited [23]. At this time, the packet discard is inevitable when the buffer overflows. For the traffic queue with limited buffer size, the dropping probability for UE i in slice m can be written as

$$\overline{dr_{i,m}} = 1 - \frac{\text{Average \# of data transmitted in a time slot}}{\text{Average \# of data arrived in a time slot}}$$
$$= 1 - \frac{\overline{R_{i,m}} \cdot \Delta}{\lambda_m^Q}. \tag{10}$$

Considering the impact of packet dropping, the Little's Law is extended to a situation with packet dropping [29]. The average delay of UE i in slice m is given by

$$\overline{D_{i,m}} = \frac{\overline{Q_{i,m}}}{(1 - \overline{dr_{i,m}})\lambda_m^Q} = \frac{\overline{Q_{i,m}}}{\overline{R_{i,m}} \cdot \Delta}. \tag{11}$$

The average delay of the slice m can be written as

$$\overline{D_m} = \frac{\sum_{i=1}^{k_m} \overline{D_{i,m}}}{k_m}. \tag{12}$$

III. PROBLEM FORMULATION AND OPTIMAL SOLUTION
A. DYNAMIC TRANSITION OF SYSTEM STATE

The current system state depends on the state and action of the previous time slot at each time slot. Define the network state as $\mathbf{S}(t)$, and it changes over time slots $t = \{1, 2, \ldots\}$. $\mathcal{S}$ denote the state space. Let $\mathbf{S}(t) = (\mathbf{H}(t), \mathbf{Q}(t), \mathbf{E}(t))$ at time

slot t, where the $\mathbf{H}(t)$ describes the channel gain state, $\mathbf{Q}(t)$ describes the queue state, and $\mathbf{E}(t)$ describes the energy state. All the elements in each set are finite discrete values.

Furthermore, the $\mathbf{H}(t)$ is defined as $\mathbf{H}(t) = \{\mathbf{H}_m^n(t)\}_{m\in M, n\in N}$, where $\mathbf{H}_m^n(t) = \{h_{i,m}^n(t)\}_{i\in k_m}$, $\mathbf{Q}(t) = \{\mathbf{Q}_m(t)\}_{m\in M}$ with $\mathbf{Q}_m(t) = \{Q_{i,m}(t)\}_{i\in k_m}$, and $\mathbf{E}(t) = \{\mathbf{E}_m(t)\}_{m\in M}$ with $\mathbf{E}_m(t) = \{E_{i,m}(t)\}_{i\in k_m}$.

We represent the channel model, queue model and energy model as discrete-time Markov chain (DTMC). Therefore, $\{\mathbf{S}(t)\}$ for given Ω is Markovian, and the state transition probability is given by

$$\Pr\{\mathbf{S}(t+1)|\mathbf{S}(t), \Omega(\mathbf{S}(t))\}$$
$$= \Pr\{\mathbf{H}(t+1)|\mathbf{S}(t), \Omega(\mathbf{S}(t))\} \Pr\{\mathbf{Q}(t+1)|\mathbf{S}(t), \Omega(\mathbf{S}(t))\}$$
$$\times \Pr\{\mathbf{E}(t+1)|\mathbf{S}(t), \Omega(\mathbf{S}(t))\}$$
$$= \Pr\{\mathbf{H}(t+1)\} \Pr\{\mathbf{Q}(t+1)|\mathbf{S}(t), \Omega(\mathbf{S}(t))\}$$
$$\times \Pr\{\mathbf{E}(t+1)|\mathbf{S}(t), \Omega(\mathbf{S}(t))\}. \tag{13}$$

Firstly, the queue state transition probability $\Pr\{\mathbf{Q}(t+1)|\mathbf{S}(t), \Omega(\mathbf{S}(t))\}$ is derived. Given the state $\mathbf{S}(t)$ and policy $\Omega(\mathbf{S}(t))$, the conditional probability of $Q_{i,m}(t+1)$ based on (4) is given by

$$\Pr\{Q_{i,m}(t+1)|\mathbf{S}(t), \Omega(\mathbf{S}(t))\}$$
$$= \beta_m \Pr(A_{i,m}^Q(t)=q) = \beta_m \frac{(\lambda_m^Q)^q}{q!} \exp\left(-\lambda_m^Q\right). \tag{14}$$

The $\Pr\{\mathbf{Q}(t+1)|\mathbf{S}(t), \Omega(\mathbf{S}(t))\}$ can be written as the product of $\Pr\{Q_{i,m}(t+1)|\mathbf{S}(t), \Omega(\mathbf{S}(t))\}$ over all UEs as

$$\Pr\{\mathbf{Q}(t+1)|\mathbf{S}(t), \Omega(\mathbf{S}(t))\}$$
$$= \prod_{m=1}^{M} \prod_{i=1}^{k_m} \Pr\{Q_{i,m}(t+1)|\mathbf{S}(t), \Omega(\mathbf{S}(t))\}. \tag{15}$$

Similarly, the energy state transition probability $\Pr\{\mathbf{E}(t+1)|\mathbf{S}(t), \Omega(\mathbf{S}(t))\}$ is derived. Given the state $\mathbf{S}(t)$ and policy $\Omega(\mathbf{S}(t))$, the conditional probability of $E_{i,m}(t)$ based on (6) is given by

$$\Pr\{E_{i,m}(t+1)|\mathbf{S}(t), \Omega(\mathbf{S}(t))\}$$
$$= \Pr\left(A_{i,m}^E(t)=e\right) = \frac{(\lambda_m^E)^e}{e!} \exp\left(-\lambda_m^E\right). \tag{16}$$

The energy state transition probability $\Pr\{\mathbf{E}(t+1)|\mathbf{S}(t), \Omega(\mathbf{S}(t))\}$ can be written as the product of $\Pr\{E_{i,m}(t+1)|\mathbf{S}(t), \Omega(\mathbf{S}(t))\}$ over all UEs as

$$\Pr\{\mathbf{E}(t+1)|\mathbf{S}(t), \Omega(\mathbf{S}(t))\}$$
$$= \prod_{m=1}^{M} \prod_{i=1}^{k_m} \Pr\{E_{i,m}(t+1)|\mathbf{S}(t), \Omega(\mathbf{S}(t))\}. \tag{17}$$

B. PROBLEM FORMATION
The objective is maximizing the average weighted-sum transmission rate by optimizing the subchannel allocation and

power control policy under the heterogeneous delay constraints. It is common knowledge that desirable data transmission rate is a key factor to improve user satisfaction [30]. The induced Markov chain $\{\mathbf{S}(t)\}$ is ergodic and has a unique steady state distribution π_s given a unichain policy Ω. Consequently, the average reward for mth slice can be derived as

$$\overline{R_m}(\Omega) = \limsup_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T} \mathbb{E}^{\Omega}[R_m(t)]$$
$$= \mathbb{E}^{\pi(\Omega)}[R_m]. \tag{18}$$

Problem 1: The objective of the dynamic virtual resource allocation is to maximize the reward function subject to the constraints, which can be expressed as

$$\max \sum_{m=1}^{M} \omega_m \overline{R_m}(\Omega) = \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T} \mathbb{E}^{\Omega}[d(\mathbf{S}(t), \Omega(\mathbf{S}(t)))], \tag{19}$$

$$\text{s.t. } \overline{D_m} \le D_m^{\max} \quad \forall m = 1, 2\ldots, M, \tag{20}$$

where ω_m is the weight for slice m, $D_m^{\max}$ is the maximum delay that slice m can tolerate, and $d(\mathbf{S}(t), \Omega(\mathbf{S}(t))) = \sum_{m=1}^{M} \omega_m R_m(t)$.

C. CONSTRAINED MDP
The proposed *Problem 1* is considered as an infinite-horizon average reward CMDP. All elements in each set are finite discrete values. Based on the global system state $\mathbf{S} = (\mathbf{H}, \mathbf{Q}, \mathbf{E}) \in \mathcal{S}$, where $\mathcal{S} = \mathcal{H} \times \mathcal{Q} \times \mathcal{E}$, the channel allocation and power control policy is optimized. Hence, the *Problem 1* is a CMDP with the following definition:
- State Space: $\{(\mathbf{H}, \mathbf{Q}, \mathbf{E}) : \forall \mathbf{H} \in \mathcal{H}, \mathbf{Q} \in \mathcal{Q}, \mathbf{E} \in \mathcal{E}\}$
- Action Space: $\{\Omega(\mathbf{H}, \mathbf{Q}, \mathbf{E}) : \forall \mathbf{H} \in \mathcal{H}, \mathbf{Q} \in \mathcal{Q}, \mathbf{E} \in \mathcal{E}\}$, where $\Omega = (\Omega_c, \Omega_p)$ is described in *Definition 1*.
- State Transition Probability: $\Pr[\mathbf{S}'|\mathbf{S}, \Omega(\mathbf{S})]$ can be formulated as (13).
- Per-stage Reward Function: $d(\mathbf{S}(t), \Omega(\mathbf{S}(t))) = \sum_{m=1}^{M} \omega_m R_m(t)$.

The MDP problem is convex programming problem, which is shown in [23]. Therefore, the constraints of *Problem 1* can be converted into an expanded objective function with a weighted-sum of the constraint functions by the Lagrangian duality method. The dual problem of the CMDP in *problem 1* is considered. For any given nonnegative Lagrangian multipliers (LMs) $\boldsymbol{\eta} = \{\eta_m|m \in M\}$, the Lagrangian function of *Problem 1* is given by

$$L(\Omega, \boldsymbol{\eta}) = \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T} \mathbb{E}^{\Omega}[g(\boldsymbol{\eta}, \mathbf{S}(t), \Omega(\mathbf{S}(t)))], \tag{21}$$

where

$$g(\boldsymbol{\eta}, \mathbf{S}, \Omega(\mathbf{S})) = \sum_{m=1}^{M} \left(\omega_m R_m + \eta_m(\overline{D_m} - D_m^{\max})\right). \tag{22}$$

As a result, we divide the *problem 1* into two subproblems, which is given by

$$\text{Subproblem 1-1}: G(\boldsymbol{\eta}) = \max_{\Omega} L(\Omega, \boldsymbol{\eta}),$$

$$\text{Subproblem 1-2}: G(\boldsymbol{\eta}^*) = \min_{\boldsymbol{\eta}} G(\boldsymbol{\eta}),$$

where Subproblem 1-1 and Subproblem 1-2 are the Lagrange dual function and the dual problem, respectively.

Consequently, we can get the optimizing policy by recursively solving the Bellman equation [31], which is given by

$$\begin{aligned}
&\theta + V(\mathbf{S}) \\
&= \max_{\Omega(\mathbf{S})} \{g(\boldsymbol{\eta}, \mathbf{S}, \Omega(\mathbf{S})) + \sum_{\mathbf{S}'} \Pr[\mathbf{S}'|\mathbf{S}, \Omega(\mathbf{S})]V(\mathbf{S}')\}, \quad (23)
\end{aligned}$$

where θ is the per-period optimal average reward for the steady-state system; $V(\mathbf{S})$ represents the value function, which represents the average reward attained according to the control policy Ω of each state $\mathbf{S}$; $\Omega(\mathbf{S}) = (\Omega_c(\mathbf{S}), \Omega_p(\mathbf{S}))$ are the subchannel allocation and power control policy performed in state $\mathbf{S}$; $g(\boldsymbol{\eta}, \mathbf{S}, \Omega(\mathbf{S}))$ in (22) is the per-stage reward function under the $\mathbf{S}$ and $\Omega(\mathbf{S})$. If $(\theta, \{V(\mathbf{S})\})$ can satisfy the fixed-point equations in (23), the $\theta = \max_{\Omega} L(\Omega, \boldsymbol{\eta})$ will be the optimal solution in *Problem 1*. In addition, the stationary policy is optimal when $\Omega^*(\boldsymbol{\eta}) = (\Omega_c^*, \Omega_p^*)$ obtains the maximum of the R.H.S. of (23) for all states.

D. EQUIVALENT BELLMAN EQUATION

Using the i.i.d. property of the CSI, the state space can be reduced. According to the partial system state $(\mathbf{Q}, \mathbf{E})$, a reduced-state Bellman equation from (23) is derived, instead of studying the complete system state $\mathbf{S}$. Specifically, the conditional actions of a policy Ω is given by following.

Definition 2: Given a control policy $\Omega = (\Omega_c, \Omega_p)$, $\Omega(\mathbf{Q}, \mathbf{E}) = \{(\mathbf{c}, \mathbf{p}) = (\Omega(\mathbf{S})): \mathbf{S} = (\mathbf{H}, \mathbf{Q}, \mathbf{E}) \forall \mathcal{H}\}$ is defined as the collection of actions $(\mathbf{c}, \mathbf{p})$ for all possible CSI $\mathbf{H}$ conditioned on a given QSI and ESI pair $(\mathbf{Q}, \mathbf{E})$. Thus, the policy Ω is the union of all conditional actions, i.e., $\Omega = \bigcup_{(\mathbf{Q}, \mathbf{E})} \Omega(\mathbf{Q}, \mathbf{E})$. ∎

By solving a reduced-state equivalent Bellman equation, we can attain the optimal control policy in *Problem 1* according to *Definition 2*, which is summarized in the below.

Lemma 1: Given a LM $\boldsymbol{\eta}$, the optimization problem in *Problem 1* is transformed into the unconstrained optimization problem. Then, the equivalent Bellman equation ($\forall (\mathbf{Q}, \mathbf{E}) \in \mathcal{Q} \times \mathcal{E}$) is given by

$$\begin{aligned}
\theta + V(\mathbf{Q}, \mathbf{E}) = \max_{\Omega(\mathbf{Q}, \mathbf{E})} & \{g(\boldsymbol{\eta}, (\mathbf{Q}, \mathbf{E}), \Omega(\mathbf{Q}, \mathbf{E})) \\
& + \sum_{(\mathbf{Q},'\mathbf{E}')} \Pr[(\mathbf{Q},'\mathbf{E}')|(\mathbf{Q}, \mathbf{E}), \Omega(\mathbf{Q}, \mathbf{E})] \\
& \times V(\mathbf{Q},'\mathbf{E}')\}, \quad (24)
\end{aligned}$$

where $V(\mathbf{Q}, \mathbf{E}) = \mathbb{E}[V(\mathbf{H}, \mathbf{Q}, \mathbf{E})|(\mathbf{Q}, \mathbf{E})] = \sum_{\mathbf{H} \in \mathcal{H}} \Pr[\mathbf{H}]V(\mathbf{H}, \mathbf{Q}, \mathbf{E})$ is the conditional expectation of value function $V(\mathbf{S})$. $\Pr[(\mathbf{Q}', \mathbf{E}')|(\mathbf{Q}, \mathbf{E}), \Omega(\mathbf{Q}, \mathbf{E})] = \mathbb{E}[\Pr[(\mathbf{Q}',$

$\mathbf{E}')|\mathbf{S}, \Omega(\mathbf{S})]|(\mathbf{Q}, \mathbf{E})]$ is the conditional expectation of state transition probability, and $g(\boldsymbol{\eta}, (\mathbf{Q}, \mathbf{E}), \Omega(\mathbf{Q}, \mathbf{E})) = \mathbb{E}[g(\boldsymbol{\eta}, \mathbf{S}, \Omega(\mathbf{S}))|(\mathbf{Q}, \mathbf{E})]$ is the conditional per-stage reward. ∎

Remark 1: The equivalent Bellman equation in (24) focus on the $(\mathbf{Q}, \mathbf{E})$ only. However, by solving (24), a stationary subchannel allocation policy Ω_c^* can be attained, which is a function of $(\mathbf{H}, \mathbf{Q}, \mathbf{E})$. Then, the power control policy is determined locally when the action for subchannel allocation is given. ∎

IV. LOW-COMPLEXITY SOLUTION

Owing to the high dimensionality of the state space, it is still very complex to obtain the optimal solution in (24). Brute-force solution brings high computational complexity. Hence, it is preferable to get a low-complexity solution. In this section, the subchannel allocation *Q-factor* is approximately converted into the sum of per-slice *Q-factor*. Then, a distributed online learning algorithm is proposed to optimize the per-slice *Q-factor* and the LM of per-slice.

A. LINEAR APPROXIMATION OF THE SUBCHANNEL ALLOCATION Q-FACTOR

To facilitate the subchannel allocation, we consider the subchannel allocation *Q-factor* $\mathbb{Q}(\mathbf{Q}, \mathbf{E}, \mathbf{c})$ w.r.t. the subchannel allocation action $\mathbf{c}$. According to *Lemma 1*, the optimal subchannel allocation is described in the following corollary.

Corollary 1: The optimal subchannel allocation is derived as ($\forall (\mathbf{Q}, \mathbf{E}) \in \mathcal{Q} \times \mathcal{E}$)

$$\Omega_c^*(\mathbf{Q}, \mathbf{E}) = \arg\max_{\mathbf{c} \in \mathcal{C}} \mathbb{Q}(\mathbf{Q}, \mathbf{E}, \mathbf{c}), \quad (25)$$

where $\mathbb{Q}(\mathbf{Q}, \mathbf{E}, \mathbf{c})$ is the subchannel allocation *Q-factor*. The Bellman equation w.r.t. $(\theta, \mathbb{Q}(\mathbf{Q}, \mathbf{E}, \mathbf{c}))$ is given by

$$\begin{aligned}
\theta + \mathbb{Q}(\mathbf{Q}, \mathbf{E}, \mathbf{c}) = \max_{\Omega_p(\mathbf{Q}, \mathbf{E})} & \{g(\boldsymbol{\eta}, (\mathbf{Q}, \mathbf{E}), \mathbf{c}, \Omega_p(\mathbf{Q}, \mathbf{E})) \\
& + \sum_{(\mathbf{Q},'\mathbf{E}')} \Pr[(\mathbf{Q},'\mathbf{E}')|(\mathbf{Q}, \mathbf{E}), \mathbf{c}, \Omega_p(\mathbf{Q}, \mathbf{E})] \\
& \times \max_{\mathbf{c}' \in \mathcal{C}} \mathbb{Q}(\mathbf{Q},'\mathbf{E},'\mathbf{c}')\}, \quad (26)
\end{aligned}$$

where $\Pr[(\mathbf{Q},'\mathbf{E}')|(\mathbf{Q}, \mathbf{E}), \mathbf{c}, \Omega_p(\mathbf{Q}, \mathbf{E})] = \mathbb{E}[\Pr[(\mathbf{Q}', \mathbf{E}')|\mathbf{S}, \mathbf{c}, \Omega_p(\mathbf{S})]|(\mathbf{Q}, \mathbf{E})]$ is the conditional expectation of state transition probability, and $g(\boldsymbol{\eta}, (\mathbf{Q}, \mathbf{E}), \mathbf{c}, \Omega_p(\mathbf{Q}, \mathbf{E})) = \mathbb{E}[g(\boldsymbol{\eta}, \mathbf{S}, \mathbf{c}, \Omega_p(\mathbf{S}))|(\mathbf{Q}, \mathbf{E})]$ is the conditional per-stage reward.

To further reduce the computational complexity and the cardinality of state space, the *Q-factor* in (26) is approximated by a linear approximation

$$\mathbb{Q}(\mathbf{Q}, \mathbf{E}, \mathbf{c}) \approx \sum_{m=1}^{M} \hat{\mathbb{Q}}_m(\mathbf{Q}_m, \mathbf{E}_m, \mathbf{c}_m), \quad (27)$$

where $\mathbf{c}_m$ denotes the per-slice subchannel allocation actions of the slice m, and $\hat{\mathbb{Q}}_m(\mathbf{Q}_m, \mathbf{E}_m, \mathbf{c}_m)$ is the per-slice *Q-factor* for the slice m of the slice QSI (SQSI) and slice ESI (SESI)

$(\mathbf{Q}_m, \mathbf{E}_m)$ and action $\mathbf{c}_m$. In addition, the per-slice *Q-factor* is the solution of per-slice fixed-point equation, which is given by

$$
\theta_m + \hat{\mathbb{Q}}_m(\mathbf{Q}_m, \mathbf{E}_m, \mathbf{c}_m)
$$
$$
= \max_{\Omega_{\mathbf{p}_m}} \{ \hat{g}_m(\eta_m, \mathbf{Q}_m, \mathbf{E}_m, \mathbf{c}_m)
$$
$$
+ \sum_{(\mathbf{Q}'_m, \mathbf{E}'_m)} \Pr\left[(\mathbf{Q}'_m, \mathbf{E}'_m)|(\mathbf{Q}_m, \mathbf{E}_m), \mathbf{c}_m\right]
$$
$$
\times \mathbb{E}[\max_{\Omega_{\mathbf{c}_m}}[\hat{\mathbb{Q}}_m(\mathbf{Q}'_m, \mathbf{E}'_m, \mathbf{c}'_m)]]\}, \tag{28}
$$

where

$$
\hat{g}_m(\eta_m, \mathbf{Q}_m, \mathbf{E}_m, \mathbf{c}_m)
$$
$$
= \mathbb{E}\left[\left(\sum_{i=1}^{k_m} \sum_{n=1}^{N} \omega_m c_{i,m}^n B \log_2\left(1 + \frac{P_{i,m}|h_{i,m}^n|^2}{\sigma_n^2}\right) \right) \right.
$$
$$
\left. + \eta_m(\overline{D_m} - D_m^{\max})) \, | \mathbf{Q}_m, \mathbf{E}_m, \mathbf{c}_m \right], \tag{29}
$$
$$
\Pr[(\mathbf{Q}'_m, \mathbf{E}'_m)|(\mathbf{Q}_m, \mathbf{E}_m), \mathbf{c}_m]
$$
$$
= \mathbb{E}[\Pr[\mathbf{Q}'_m, \mathbf{E}'_m|\mathbf{H}_m, \mathbf{Q}_m, \mathbf{E}_m, \mathbf{c}_m]|\mathbf{Q}_m, \mathbf{E}_m, \mathbf{c}_m]. \tag{30}
$$

Based on the linear approximation of the subchannel allocation *Q-factor* by the sum of per-slice subchannel allocation *Q-factor* in (27), the resource manager decides the optimal subchannel allocation according to

$$
\Omega_c^*(\mathbf{Q}, \mathbf{E}) = \arg\max_c \sum_{m=1}^{M} \hat{\mathbb{Q}}_m(\mathbf{Q}_m, \mathbf{E}_m, \mathbf{c}_m). \tag{31}
$$

Due to the power control can be obtained locally based on the subchannel allocation action. According to the slice CSI (SCSI), SQSI and SESI ($\mathbf{H}_m, \mathbf{Q}_m, \mathbf{E}_m$), the resource manager determines $\Omega_p^*(\mathbf{Q}_m, \mathbf{E}_m) = \{\Omega_p^*(\mathbf{H}_m, \mathbf{Q}_m, \mathbf{E}_m) : (\mathbf{H}_m, \mathbf{Q}_m, \mathbf{E}_m)\forall \mathbf{H}_m\}$, which gets the maximum of the R.H.S. of (28). Therefore, the whole power control policy is attained by $\Omega_p^*(\mathbf{S}) = \{\Omega_{\mathbf{p}_m}^*((\mathbf{H}_m, \mathbf{Q}_m, \mathbf{E}_m) : \mathbf{c}_n \in \Omega_c^*(\mathbf{Q}, \mathbf{E})\}$.

B. ONLINE LEARNING ALGORITHM VIA STOCHASTIC APPROXIMATION

To solve the (28) and get the LM for each slice, the per-slice delay constraint need to be satisfied, since the control policy $\Omega^* = (\Omega_c^*, \Omega_p^*)$ is function of per-slice *Q-factor* $\{\hat{\mathbb{Q}}(\mathbf{Q}_m, \mathbf{E}_m, \mathbf{c}_m)\}$. In this section, the per-slice *Q-factors* are distributedly estimated by stochastic learning [32] based on the SCSI, SQSI and SESI, and the LMs also be estimated for each slice. The Distributed Online Learning Algorithm using Stochastic Approximation needs to obtain according to the SCSI, SQSI and SESI, which is summarized in the following.

Algorithm 1 (Distributed Online Learning Algorithm using Stochastic Approximation):

- **Step 1 [Initialization]**: Let $t = 0$. Each slice manager initialize per-slice *Q-factor* $\{\mathbb{Q}_m^0(\mathbf{Q}_m, \mathbf{E}_m, \mathbf{c}_m)\}$, and the LM η_m^0.

- **Step 2 [Subchannel Allocation]**: Based on SCSI, SQSI and SESI and LM, each slice manager obtain the channel allocation $\mathbf{c}(t) = \{c_{i,m}^n(t)\}_{i \in k_m, m \in M, n \in N} \in \mathcal{C}$ at the beginning of the tth slot.
- **Step 3 [Power Control]**: Based on SCSI, SQSI, SESI and LM, each slice manager implements the power control $\mathbf{p}_m(t) = \{\phi_{i,m}(t)\}_{i \in k_m, m \in M} \in \mathcal{P}$ on the basis of $\Omega_{p_m}^*(\mathbf{H}_m(t), \mathbf{Q}_m(t), \mathbf{E}_m(t))$.
- **Step 4 [Update Per-Slice Q-Factors and LMs]**: Each slice updates per-slice *Q-factor* on the basis of the real-time SCSI, SQSI and SESI after the control action is determined according to (32). Furthermore, the LMs $\eta_m(t)$ for each slice m is updated to $\eta_m(t+1)$ according to the iterative formula given by (34).
- **Step 5 [Termination]**: If $||\hat{\mathbb{Q}}(t+1) - \hat{\mathbb{Q}}(t)|| < \mu_Q$ and $||\boldsymbol{\eta}(t+1) - \boldsymbol{\eta}(t)|| < \mu_\eta$, stop; otherwise, set $t := t+1$ and go to step 2.

The per-slice *Q-factor* update in Step 4 according to SCSI, SQSI and SESI at the current time slot is given by

$$
\hat{\mathbb{Q}}_m^{t+1}(\mathbf{Q}_m, \mathbf{E}_m, \mathbf{c}_m)
$$
$$
= \hat{\mathbb{Q}}_m^t(\mathbf{Q}_m, \mathbf{E}_m, \mathbf{c}_m) + \epsilon_t[U_m(\hat{\mathbb{Q}}_m^t, \mathbf{Q}_m, \mathbf{E}_m, \mathbf{c}_m)
$$
$$
- U_m(\hat{\mathbb{Q}}_m^t, \mathbf{Q}_m^I, \mathbf{E}_m^I, \mathbf{c}_m^I) - \hat{\mathbb{Q}}_m^t(\mathbf{Q}_m, \mathbf{E}_m, \mathbf{c}_m)], \tag{32}
$$

where

$$
U_m(\hat{\mathbb{Q}}_m^t, \mathbf{Q}_m, \mathbf{E}_m, \mathbf{c}_m)
$$
$$
= \hat{g}_m(\mathbf{Q}_m, \mathbf{E}_m, \mathbf{c}_m)
$$
$$
+ \sum_{(\mathbf{Q}'_m, \mathbf{E}'_m)} \Pr[(\mathbf{Q}'_m, \mathbf{E}'_m)|(\mathbf{Q}_m, \mathbf{E}_m), \mathbf{c}_m]
$$
$$
\times \mathbb{E}\left[\mathbb{Q}_m^t \left(\min\{Q_m^t + A_m^Q(t), B_m^Q\}, \right. \right.
$$
$$
\left. \left. \min\{E_m^t + A_m^E(t), B_m^E\}, \mathbf{c}'_m\right) \right]. \tag{33}
$$

The LMs $\eta_m(t)$ for each slice m can be updated as

$$
\eta_m(t+1) = \Gamma(\eta_m(t+1) + \epsilon_\eta(t)(\overline{D}_m(t) - D_m^{\max})). \tag{34}
$$

C. DECOMPOSITION OF THE PER-SLICE Q-FACTOR

Note that the cardinality is $(N_Q + 1)^{k_m}(N_E + 1)^{k_m}$ for the per-slice system state, which is exponential in the number of UEs at slice mth, i.e., k_m. We point out that the per-slice *Q-factor* can be further split into per-slice per-UE *Q-factor* in the following lemma, which results in a linear order for the state space cardinality, i.e., $k_m(N_Q + 1)k_m(N_E + 1)$.

Lemma 2: (Decomposition of Per-slice Q-factor): The per-slice *Q-factor* $\hat{\mathbb{Q}}_m(\mathbf{Q}_m, \mathbf{E}_m, \mathbf{c}_m)$ can be decomposed into the sum of the per-slice per-UE *Q-factors* $\hat{\mathbb{Q}}_{i,m}(Q_{i,m}, E_{i,m}, \mathbf{c}_{i,m})$, i.e., $\hat{\mathbb{Q}}_m(\mathbf{Q}_m, \mathbf{E}_m, \mathbf{c}_m) = \sum_{i=1}^{k_m} \hat{\mathbb{Q}}_{i,m}(Q_{i,m}, E_{i,m}, \mathbf{c}_{i,m})$, where $\hat{\mathbb{Q}}_{i,m}(Q_{i,m}, E_{i,m}, \mathbf{c}_{i,m})$ satisfies the per-slice per-UE *Q-factor*

equation in (35).

$$\theta_{i,m} + \hat{\mathbb{Q}}_{i,m}(Q_{i,m}, E_{i,m}, \mathbf{c}_{i,m})$$

$$= \max_{\Omega_{p_{i,m}}} \left\{ \hat{g}_{i,m}(\boldsymbol{\eta}_m, (Q_{i,m}, E_{i,m}), \mathbf{c}_{i,m}) \right.$$

$$+ \sum_{(Q'_{i,m}, E'_{i,m})} \Pr[(Q'_{i,m}, E'_{i,m})|(Q_{i,m}, E_{i,m}), \mathbf{c}_{i,m}]$$

$$\left. \mathbb{E}\left[\max_{\Omega_{\mathbf{c}_{i,m}}} [\hat{\mathbb{Q}}_{i,m}(Q'_{i,m}, E'_{i,m}, \mathbf{c}'_{i,m})] \right] \right\}, \qquad (35)$$

where, $\hat{g}_{i,m}(\boldsymbol{\eta}_m, Q_{i,m}, E_{i,m}, \mathbf{c}_{i,m})$ and $\Pr[(Q'_{i,m}, E'_{i,m}|(Q_{i,m}, E_{i,m}), \mathbf{c}_{i,m}]$ are given by

$$\hat{g}_{i,m}(\boldsymbol{\eta}_m, Q_{i,m}, E_{i,m}, \mathbf{c}_{i,m})$$

$$= \mathbb{E}\left[\left(\sum_{n=1}^{N} \omega_m c_{i,m}^n B \log_2 \left(1 + \frac{P_{i,m}|h_{i,m}^n|^2}{\sigma_n^2} \right) \right. \right.$$

$$\left. \left. + \frac{\eta_m}{k_m}\left(\overline{D}_{i,m} - D_m^{\max} \right) \right) \middle| Q_{i,m}, E_{i,m} \right], \qquad (36)$$

and

$$\Pr[(Q'_{i,m}, E'_{i,m}|(Q_{i,m}, E_{i,m}), \mathbf{c}_{i,m})]$$

$$= \mathbb{E}[\Pr[Q'_{i,m}, E'_{i,m}|\mathbf{H}_m, Q_{i,m}, E_{i,m},$$

$$\Omega_{c_{i,m}}(\mathbf{H}_m, Q_{i,m}, E_{i,m})]|Q_{i,m}, E_{i,m}]. \qquad (37)$$

∎

Recall that the subchannel $n \in \{1, 2, \dots, N\}$ can be allocated to at most one UE. In addition, observe that the summation index n in (36) is the index of subchannels. Consequently, for every n, at most one $c^n = 1$, while all the other $c^n = 0$. Due to the birth-death queue dynamics and symmetry of each subchannel [33], the per-slice per-UE *Q-factor* meeting (35) can be converted into the sum of the per-slice per-UE persubchannel *Q-factors*, which can be written as

$$\hat{\mathbb{Q}}_{i,m}(Q_{i,m}, E_{i,m}, \mathbf{c}_{i,m}) = \sum_{n=1}^{N} \hat{\mathbb{Q}}_{i,m}(Q_{i,m}, E_{i,m}, c_{i,m}^n), \qquad (38)$$

where per-slice per-UE per-subchannel *Q-factor* satisfies:

$$\bar{g}_{i,m}(\eta_m, Q_{i,m}, E_{i,m}, \Omega_{c_{i,m}^n}(Q_{i,m}, E_{i,m}))$$

$$= \mathbb{E}\left[\left(\omega_m c_{i,m}^n B \log_2 \left(1 + \frac{P_{i,m}|h_{i,m}^n|^2}{\sigma_n^2} \right) \right. \right.$$

$$\left. \left. + \frac{\eta_m}{Nk_m}(D_{i,m} - D_m^{\max}) \right) \middle| Q_{i,m}, E_{i,m} \right]. \qquad (39)$$

Then, the subchannel allocation action is determined as

$$c_{i,m}^n = \begin{cases} 1, & \text{if } F > \max_j F_j \\ 0, & \text{otherwise} \end{cases}, \qquad (40)$$

where $F = \hat{\mathbb{Q}}_{i,m}(Q_{i,m}, E_{i,m}, c_{i,m}^n)$ and $F_j = \hat{\mathbb{Q}}_{i,m}(Q_{i,m}, E_{i,m}, c_{i,m}^j)$.

Parameters	Values
Radius	100 m
Background noise power (σ^2)	-104 dBm
Subchannel Bandwidth	8 MHz
Number of subchannels	6
FPC factor	0.6, 0.8, 1

TABLE 3. The Parameters for Three Kinds of Slices

Parameters	eMBB	URLLC	mMTC
Number of UEs	2	2	2
Baseline power	-76 dBm	-73 dBm	-79 dBm
Queue state	6	6	4
Energy state	6	6	4
Packet size	100 kbits	10 kbits	2 kbits
Average packet arrival rate	3 packet/Δ	3 packet/Δ	3 packet/Δ
Average energy arrival rate	3 unit/Δ	3 unit/Δ	3 unit/Δ
Maximum delay	100 ms	10 ms	3 s

V. SIMULATION RESULTS

In this section, the proposed optimal and reduced complexity solutions is compared with three reference baselines to demonstrate the advantage of the proposed scheme.

- **Baseline 1 [Random Control Scheme]**: The virtual resources are assigned to each slice equally. On the basis of this policy, each slice can utilize the radio resources with the same opportunity. In this scheme, the UE always transmit with the maximum power.
- **Baseline 2 [QSI-Based Scheme]**: In order to optimize the weighted-sum transmission rate, the subchannel allocation and power control policy are determined according to the CSI and QSI.
- **Baseline 3 [Value Iteration Scheme]**: The value iteration scheme can find the optimal policy and value functions. Hence, it is regarded as the upper bound of performance.

Consider that a RAN coverage area with radius 100 m has three types of network slices to support URLLC, eMBB and mMTC, respectively. And the mean of fading is fixd to $\bar{g} = -10$ dB. The main parameter values of the RAN are listed in Table 2. Each slice supports a kind of UE. The main parameter values of the slice UEs are listed in Table 3. There are 6 subchannels with bandwidth 8 MHz. A data queue is maintained at each UE to cache the arrival packets. The UE keeps a queue to cache the arrival data packets. The UEs in different slices have different buffer size and battery capacity.

Fig. 3 to Fig. 5 compare the performance of the proposed optimization scheme with Random Control Scheme and QSI-Based Scheme. Simulation results indicate that the proposed optimization scheme achieves a performance improvement and outperforms the baseline 1 and baseline 2. The virtual resources are assigned to each slice equally and the maximum transmission power is adopted in Baseline 1, hence it realizes worse performance. Baseline 2 cares the CSI and QSI without considering the ESI, which is better than baseline 1 and worse
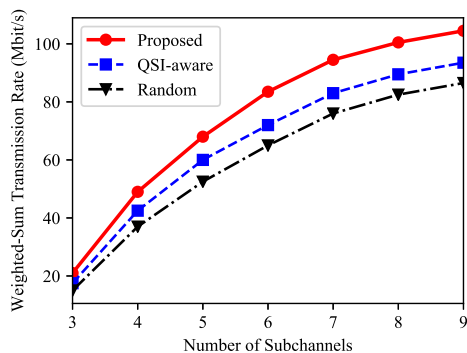
FIGURE 3. Weighted-sum transmission rate versus the number of subchannels.
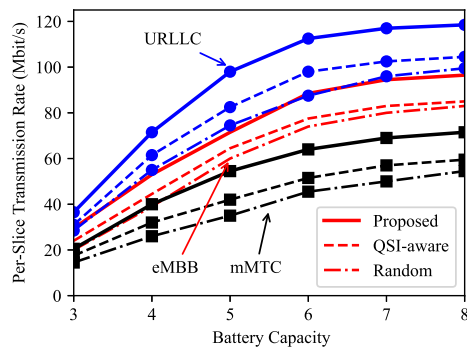

FIGURE 4. Per-slice transmission rate versus battery capacity.

FIGURE 5. Per-slice dropping probability versus the average packet arrival rate.

FIGURE 6. Convergence property of the proposed online stochastic learning algorithm.

than the proposed scheme. In addition to replacing the equality allocation with on-demand subchannel allocating among UEs, the performance improvement is also owing to the power control policy in the proposed optimization scheme. Fig. 6 indicates that the performance of the proposed optimization scheme is asymptotically optimal and highly close to the upper bound.

Fig. 3 illustrates weighted-sum transmission rate versus the number of subchannels. For all the schemes, we can observed that weighted-sum transmission rate increases with the number of subchannels. Due to the lack of channel supply at the beginning, only some UEs get the opportunities to transmit packets, which results in an increase in the number of packets being dropped and the number of packets remaining in the queue. Even if every UE can get the channel access opportunities, in order to reduce both the packet drops and delay, the slice manager may need to schedule more queued packets for transmission when the channel environment is poor, therefore the higher transmission power will be selected. As the number of channels increases, multiple subchannels may be allocated to one UE. At this time, the transmission rate of UEs grow slowly as a result of the limitation of buffer size and average packet arrival rate.

Fig. 4 illustrates per-slice transmission rate versus battery capacity. For all the schemes, simulation results show that the per-slice transmission rate increases with the battery capacity.
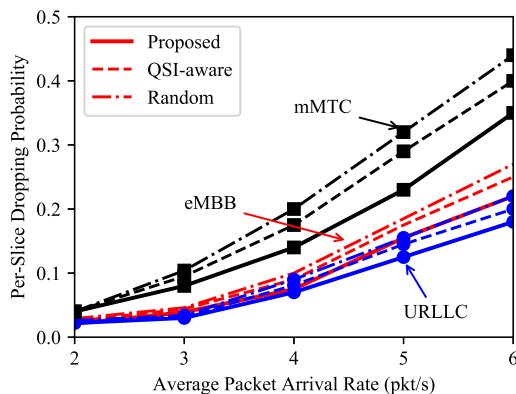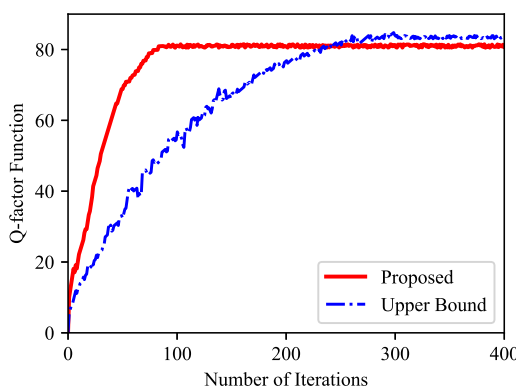
As the battery capacity gradually increases, the rate increases slowly. This is because even though the battery capacity increases, the average energy arrival does not increase. Moreover, the proposed optimization scheme has significant performance improvement compared with baseline 1 and baseline 2. The transmission rate of URLLC slice with strict delay constraints is higher than that of the other two slices. In addition, mMTC slice has the lowest delay requirement and the smallest packet size, resulting in the lowest transmission rate. Observe that the performance gaps between the baseline 1 and baseline 2 become smaller with the increase of battery capacity. This is because more energy is stored in the battery as the battery capacity increases, the slice manager is more inclined to choose the maximum transmit power.

Fig. 5 illustrates per-slice dropping probability versus the average packet arrival rate. As the average packet arrival rate increases, the dropping probability increases gradually. Observe that mMTC slice has the largest dropping probability due to the limitation of buffer capacity. The buffer capacity of mMTC slice is 3, and the buffer capacity of eMBB slice and URLLC slice are 6. After the average packet arrival rate is greater than 3, the dropping probability of mMTC slice increases rapidly, the dropping probability of eMBB slice and URLLC slice increase relatively slowly.

Fig. 6 illustrates the convergence property of the proposed optimization scheme for estimating per-slice *Q-factor* and the value iteration algorithm given in baseline 3. It turns out that the proposed optimization scheme converges quite fast. The performance at about 80-th number of iteration is already quite close to the converged value.

VI. CONCLUSION

In this paper, we propose a dynamic virtual resources allocation scheme based on RAN slicing for uplink communication to guarantee the QoS. The resources control problem is formulated as an infinite-horizon average-reward CMDP problem. To reduce the computational complexity, an equivalent Bellman equation on the basis of subchannel allocation *Q-factor* is proposed to solve the CMDP. In addition, we adopt a distributed online stochastic learning algorithm to optimize the value functions and LMs. Then, the subchannel allocation *Q-factor* is approximate to the sum of per-slice *Q-factors* to further reduce the computational complexity. Finally, simulation results reveal that the proposed optimization scheme can converge and improve the user performance compared with baseline 1 and baseline 2.

Going forward, several directions deserve further investigation. First, the multiple base station in the proposed RAN slicing networks is yet to be analyzed. In this situation, the introduction of users as interferers in the uplink will make the optimization even more challenging. Second, it is difficult to get all the state information or there are some inaccurate state information in practice. So, an optimal policy cannot be obtained based on the imperfect observation states. In order to solve the problem, it is of interest to consider that the resources control problem is further formulated as a Partially Observed Markov Decision Process.

REFERENCES

[1] J. Li, S. Chu, F. Shu, J. Wu, and D. N. K. Jayakody, "Contract-based small-cell caching for data disseminations in ultra-dense cellular networks," *IEEE Trans. Mobile Comput.*, vol. 18, no. 5, pp. 1042–1053, May 2019.

[2] J. Li, H. Chen, Y. Chen, Z. Lin, B. Vucetic, and L. Hanzo, "Pricing and resource allocation via game theory for a small-cell video caching system," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 8, pp. 2115–2129, Aug. 2016.

[3] R. A. Addad, T. Taleb, H. Flinck, M. Bagaa, and D. Dutra, "Network slice mobility in next-generation mobile systems: Challenges and potential solutions," *IEEE Netw.*, vol. 34, no. 1, pp. 84–93, Jan. 2020.

[4] X. Shen *et al.*, "AI-assisted network-slicing based next-generation wireless networks," *IEEE Open J. Veh. Technol.*, vol. 1, pp. 45–66, Jan. 2020.

[5] P. Wu, J. Li, L. Shi, M. Ding, K. Cai, and F. Yang, "Dynamic content update for wireless edge caching via deep reinforcement learning," *IEEE Commun. Lett.*, vol. 23, no. 10, pp. 1773–1777, Oct. 2019.

[6] 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.

[7] G. Tseliou, F. Adelantado, and C. Verikoukis, "A base station agnostic network slicing framework for 5G," *IEEE Netw.*, vol. 33, no. 4, pp. 82–88, Jul. 2019.

[8] Q. Ye, J. Li, K. Qu, W. Zhuang, X. Shen, and X. Li, "End-to-end quality of service in 5G networks: Examining the effectiveness of a network slicing framework," *IEEE Veh. Technol. Mag.*, vol. 13, no. 2, pp. 65–74, Jun. 2018.

[9] S. Yan, P. Yang, Q. Ye, W. Zhuang, X. S. X. Li, and J. Rao, "Transmission protocol customization for network slicing: A case study of video streaming," *IEEE Veh. Technol. Mag.*, vol. 14, no. 4, pp. 20–28, Dec. 2019.

[10] F. Z. Yousaf, M. Bredel, S. Schaller, and F. Schneider, "NFV and SDN - key technology enablers for 5G networks," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 11, pp. 2468–2478, Nov. 2017.

[11] R. Su *et al.*, "Resource allocation for network slicing in 5G telecommunication networks: A survey of principles and models," *IEEE Netw.*, vol. 33, no. 6, pp. 172–179, Nov. 2019.

[12] W. Zhuang, Q. Ye, F. Lyu, N. Cheng, and J. Ren, "SDN/NFV-empowered future IoV with enhanced communication, computing, and caching," *Proc. IEEE*, vol. 108, no. 2, pp. 274–291, Feb. 2020.

[13] K. Samdanis, X. Costa-Pérez, 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.

[14] X. Li *et al.*, "Network slicing for 5G: Challenges and opportunities," *IEEE Internet Comput.*, vol. 21, no. 5, pp. 20–27, Sep. 2017.

[15] "Active RAN sharing could save operators 60 billion," Sep. 2016. [Online]. Available: http://www.cellular-news.com/story/36831.php

[16] 3GPP, "Study on radio access network (RAN) sharing enhancements," Sophia Antipolis Cedex, France, TR 22.852, Sep. 2014.

[17] 3GPP, "Telecommunication management; network sharing; concepts and requirements," Sophia Antipolis Cedex, France, TS 32.130, Jun. 2018.

[18] L. Wang, Z. Lu, X. Wen, R. Knopp, and R. Gupta, "Joint optimization of service function chaining and resource allocation in network function virtualization," *IEEE Access*, vol. 4, pp. 8084–8094, Nov. 2016.

[19] M. Richart, J. Baliosian, J. Serrat, and J. Gorricho, "Resource slicing in virtual wireless networks: A survey," *IEEE Trans. Netw. Service Manage.*, vol. 13, no. 3, pp. 462–476, Sep. 2016.

[20] Y. Sun, G. Feng, L. Zhang, M. Yan, S. Qin, and M. A. Imran, "User access control and bandwidth allocation for slice-based 5G-and-beyond radio access networks," in *Proc. IEEE Int. Conf. Commun.*, May 2019, pp. 1–6.

[21] F. Bahlke, O. Ramos, S. Henneberger, and M. Pesavento, "Optimized cell planning for network slicing in heterogeneous wireless communication networks," *IEEE Commun. Lett.*, vol. 22, no. 8, pp. 1676–1679, Aug. 2018.

[22] P. Caballero, A. Banchs, G. V. X. Costa-Pérez, and A. Azcorra, "Network slicing for guaranteed rate services: Admission control and resource allocation games," *IEEE Trans. Wireless Commun.*, vol. 17, no. 10, pp. 6419–6432, Oct. 2018.

[23] L. Lei, Y. Kuang, N. Cheng, X. S. Shen, Z. Zhong, and C. Lin, "Delay-optimal dynamic mode selection and resource allocation in device-to-device communications—Part I: Optimal policy," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 3474–3490, May 2016.

[24] W. Wang, F. Zhang, and V. K. N. Lau, "Dynamic power control for delay-aware device-to-device communications," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 1, pp. 14–27, Jan. 2015.

[25] 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, pp. 2640–2654, Dec. 2018.

[26] L. Tang, Q. Tan, Y. Shi, C. Wang, and Q. Chen, "Adaptive virtual resource allocation in 5G network slicing using constrained Markov decision process," *IEEE Access*, vol. 6, pp. 61 184–61 195, Nov. 2018.

[27] F. Song *et al.*, "Probabilistic caching for small-cell networks with terrestrial and aerial users," *IEEE Trans. Veh. Technol.*, vol. 68, no. 9, pp. 9162–9177, Sep. 2019.

[28] J. Li, Y. Chen, Z. Lin, W. Chen, B. Vucetic, and L. Hanzo, "Distributed caching for data dissemination in the downlink of heterogeneous networks," *IEEE Trans. Commun.*, vol. 63, no. 10, pp. 3553–3568, Oct. 2015.

[29] Y. Cui, V. K. N. Lau, R. Wang, H. Huang, and S. Zhang, "A survey on delay-aware resource control for wireless systems-large deviation theory, stochastic Lyapunov drift, and distributed stochastic learning," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1677–1701, Mar. 2012.

[30] J. Li *et al.*, "On social-aware content caching for D2D-enabled cellular networks with matching theory," *IEEE Internet Things J.*, vol. 6, no. 1, pp. 297–310, Feb. 2019.

[31] D. P. Bertsekas, *Dynamic Programming and Optimal Control*, 3rd ed. Belmont, MA, USA: Athena Scientific, 2007.

[32] X. R. Cao, *Stochastic Learning and Optimization: A Sensitivity-Based Approach*, 1st ed. New York, NY, USA: Springer, 2007.

[33] Y. Cui, V. K. N. Lau, and Y. Wu, "Delay-aware BS discontinuous transmission control and user scheduling for energy harvesting downlink coordinated MIMO systems," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3786–3795, Jul. 2012.

FEI SONG (Student Member, IEEE) received the M.S. degree from Xi'an Polytechnic University, P. R. China in 2017. From September 2017 to now, she is a Ph.D. Student with the School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing, China. Her research interests include small-cell caching and network slicing.

JUN LI (Senior Member, IEEE) received the Ph.D degree in electronic engineering from Shanghai Jiao Tong University, Shanghai, P. R. China in 2009. From January 2009 to June 2009, he worked in the Department of Research and Innovation, Alcatel Lucent Shanghai Bell as a Research Scientist. From June 2009 to April 2012, he was a Postdoctoral Fellow at the School of Electrical Engineering and Telecommunications, the University of New South Wales, Australia. From April 2012 to June 2015, he is a Research Fellow at the School of Electrical Engineering, the University of Sydney, Australia. From June 2015 to now, he is a Professor with the School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing, China. He was a Visiting Professor at Princeton University from 2018 to 2019. His research interests include network information theory, game theory, distributed intelligence, multiple agent reinforcement learning, and their applications in ultra-dense wireless networks, mobile edge computing, network privacy and security, and industrial Internet of things. He has coauthored more than 200 papers in IEEE journals and conferences, and holds one US patents and more than ten Chinese patents in these areas. He was serving as an Editor of IEEE COMMUNICATION LETTERS and TPC member for several flagship IEEE conferences. He received Exemplary Reviewer of IEEE TRANSACTIONS ON COMMUNICATIONS in 2018, and Best Paper Award from IEEE International Conference on 5G for Future Wireless Networks in 2017.
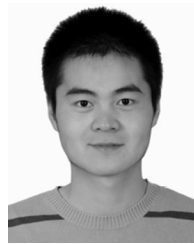
CHUAN MA received the B.S. degree from the Beijing University of Posts and Telecommunications, Beijing, China, in 2013 and Ph.D. degree from the University of Sydney, Australia, in 2018. He is now working as a Lecturer with the School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing, China. He has published more than ten journal and conference papers, including a Best Paper in WCNC 2018. His research interests include stochastic geometry, wireless caching networks and machine learning, and now focuses on the big data analysis and privacy preservation.

YIJIN ZHANG (Senior Member, IEEE) received the B.S. degree from the Nanjing University of Posts and Telecommunications, China, in 2004, the M.S. degree from Southeast University, China, in 2007, and the Ph.D. degree from The Chinese University of Hong Kong in 2010, all in information engineering. He joined the Nanjing University of Science and Technology, China, in 2011, where he is currently a Professor with the School of Electronic and Optical Engineering. His research interests include sequence design and resource allocation for communication networks.

LONG SHI (Member, IEEE) received the Ph.D. degree in electrical engineering from the University of New South Wales, Sydney, Australia, in 2012. He was a Visiting Student at the Chinese University of Hong Kong and University of Delaware in 2010 and 2011, respectively. From 2013 to 2016, he was a Postdoctoral Fellow at the Institute of Network Coding, Chinese University of Hong Kong, Hong Kong SAR, China. From 2014 to 2017, he was a Lecturer at the College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, China. Currently, he is a Research Fellow with the Singapore University of Technology and Design. His current research interests include mobile edge computing, wireless caching, and deep reinforcement learning.

DUSHANTHA NALIN K. JAYAKODY (Senior Member, IEEE) received the Ph. D. degree in electronics, electrical, and communications engineering, from the University College Dublin, Ireland in 2013. From 2014–2016, he was a Postdoc Research Fellow at the Institute of Computer science, University of Tartu, Estonia and Department of Informatics, University of Bergen, Norway. From 2016, he is a Professor at the School of Computer Science & Robotics, National Research Tomsk Polytechnic University (TPU), Russia. In addition, since 2019, he also serves as the Dean/ School of Postgraduate & Research, Sri Lanka Technological Campus (SLTC), Padukka Sri Lanka and Founding Director of Centre of Telecommunication Research, SLTC, Sri Lanka. He has received the Best Paper Award from the IEEE International Conference on Communication, Management and Information Technology (ICCMIT) in 2017 and International Conference on Emerging Technologies of Information and Communications, Bhutan, March 2019. In July 2019, He received the Education Leadership Award from the World Academic Congress in 2019. In 2017 and 2018, he received the Outstanding Faculty Award by National Research Tomsk Polytechnic University, Russia. He also received Distinguished Researcher in Wireless Communications in Chennai, India 2019. He has published over 140 international peer reviewed journal and conference papers and books. His research interests include PHY and NET layer prospective of 5G communications technologies such as NOMA for 5G etc, Cooperative wireless communications, device to device communications, LDPC codes, Unmanned Ariel Vehicle etc. He is a Senior Member of IEEE and he has organized or co-organized more than 20 workshops and special sessions of various IEEE conferences. He also served as Chair, Session Chair or Technical Program Committee Member for various international conferences, such as IEEE PIMRC 2013-2019, IEEE WCNC 2014-2018, IEEE VTC 2015-2018 etc. He currently serves as an Area Editor the *Elsevier Physical Communications Journal, MDPI Information journal and Wiley Internet of Technology Letters*. Also, he serves as a Reviewer for various IEEE Transactions and other journals.