Hindawi
Wireless Communications and Mobile Computing
Volume 2020, Article ID 6290375, 11 pages
https://doi.org/10.1155/2020/6290375

Research Article

Slicing Resource Allocation for eMBB and URLLC in 5G RAN

Tengteng Ma,[1,2] **Yong Zhang**,[1,2] **Fanggang Wang**,[3] **Dong Wang**,[3] **and Da Guo**[1]

[1]*School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China*
[2]*Beijing Key Laboratory of Work Safety Intelligent Monitoring, Beijing University of Posts and Telecommunications, Beijing 100876, China*
[3]*State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, Beijing 100044, China*

Correspondence should be addressed to Yong Zhang; yongzhang@bupt.edu.cn

Received 9 September 2019; Accepted 30 December 2019; Published 31 January 2020

Academic Editor: Carles Gomez

Copyright © 2020 Tengteng Ma et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper investigates the network slicing in the virtualized wireless network. We consider a downlink orthogonal frequency division multiple access system in which physical resources of base stations are virtualized and divided into enhanced mobile broadband (eMBB) and ultrareliable low latency communication (URLLC) slices. We take the network slicing technology to solve the problems of network spectral efficiency and URLLC reliability. A mixed-integer programming problem is formulated by maximizing the spectral efficiency of the system in the constraint of users' requirements for two slices, i.e., the requirement of the eMBB slice and the requirement of the URLLC slice with a high probability for each user. By transforming and relaxing integer variables, the original problem is approximated to a convex optimization problem. Then, we combine the objective function and the constraint conditions through dual variables to form an augmented Lagrangian function, and the optimal solution of this function is the upper bound of the original problem. In addition, we propose a resource allocation algorithm that allocates the network slicing by applying the Powell–Hestenes–Rockafellar method and the branch and bound method, obtaining the optimal solution. The simulation results show that the proposed resource allocation algorithm can significantly improve the spectral efficiency of the system and URLLC reliability, compared with the adaptive particle swarm optimization (APSO), the equal power allocation (EPA), and the equal subcarrier allocation (ESA) algorithm. Furthermore, we analyze the spectral efficiency of the proposed algorithm with the users' requirements change of two slices and get better spectral efficiency performance.

1. Introduction

With the increase of the demand for mobile phones and networks, the new Internet of things (IoT) for consumers and vertical industries has developed rapidly. The mobile Internet and the IoT tend to be the main forces to drive the development of mobile communication in the future network. In particular, the equipment of massive IoT that connected to sensors such as humidity sensors and remote industrial robots needs to provide divergent services [1]. For the multiservice requirement of vertical industries in the 5th generation (5G), the one-size-fits-all design concept is not viable anymore, and it is a promising approach to meet the requirement that slicing one physical network into several logical networks according to different demands [2]. Currently, network slicing is one of the most cost-effective methods to meet the different requirements of services with multiple logical networks. It has been a key enabler for 5G to accommodate a variety of services in a flexible manner [3]. Therefore, it is critical to study multiservice problems in the virtual radio access network (RAN) by network slicing, especially different service requirements in various scenarios, such as enhanced mobile broadband (eMBB), ultrareliable low latency communication (URLLC), and massive machine-type communication (mMTC) [4, 5].

The network slicing architecture consists of a core network (CN) and RAN slicing. Since the research of network slicing in the CN has been relatively mature and our research focuses on RAN slicing, the work of CN slicing is briefly introduced. For instance, in [6], the authors investigated how to combine fog node and network slicing of CN to safely access remote service data while ensuring low

latency. Network orchestration architecture for dynamic network slicing is considered in [7], demonstrating how to provide dynamic network slicing for enterprise-networking services and mobile metro-core networks. Rayani et al. in [8] proposed virtualized EPC-based 5G architecture and considered the resource allocation problem, which aims at minimizing the cost of network slicing while ensuring QoS requirement.

In terms of the RAN, some researchers have studied the problem of network slicing resource allocation. One of the main challenges of RAN slicing is to provide different levels of resource isolation through resource abstraction, virtualization, and separation of services [9]. Many papers have already studied the main challenges. The network slicing could be allocated resources for each service to provide performance guarantees and isolate it from other services [10]. Three typical network slices focused on service-oriented deployment by providing different deployment strategies, which can effectively improve the utilization rate of resources [11]. The development of technology focuses on the CN slicing, but it is essential to study the network slicing of the RAN, especially the resource allocation problem with RAN slicing. In addition, the pivotal problem of the RAN slicing algorithm is that the mathematical formula is the NP-hard problem [12]. Most scholars focus on the architecture of RAN slicing, and the research on resource allocation optimization of RAN slicing is insufficient. One of the principal research projects is resource allocation among different slices in the RAN [13]. These issues motivate our paper.

Different from these papers, we pay more attention to the multiservice virtual resource allocation and management of the RAN slicing. Our interest in this article focuses on the methods and mathematical models of resource allocation algorithms for 5G network slicing eMBB and URLLC. The purpose of our research is to improve the resource utilization of the overall system and the reliability of the URLLC service. Our contribution in this paper can be summarized as follows:

(i) We propose a network architecture with slice eMBB and slice URLLC in Figure 1 and aim to maximize the spectral efficiency of the overall network system.

(ii) To efficiently allocate the slice resources, we formulate the spectral efficiency maximization problem with the high capacity of eMBB slices and the low delay of URLLC slices.

(iii) The original mixed-integer nonlinear programming problem is approximately transformed into a nonlinear problem by relaxing integer variables, and we prove that the nonlinear problem is a convex optimization problem.

(iv) Finally, we propose a slice resource allocation algorithm by applying the Powell–Hestenes–Rockafellar (PHR) and the branch and bound method. The simulation results show that the proposed algorithm can ensure the users' requirement of the eMBB slice and the users' requirement of the URLLC slice with a high probability and improve the spectral efficiency of the overall system.

The remainder of this paper is organized as follows: Section 3 introduces the system model and the problem formulation. In Section 4, we determine the optimal power and the subcarrier allocation spectral efficiency design with network slicing. The simulation results are shown in Section 5. Section 6 concludes this paper.

2. Related Work

The main idea of RAN slicing resource allocation research is optimizing the performance of the entire network system while ensuring the various users' requirements. In this section, we analyze the resource allocation of RAN slicing from the perspective of single service and multiservice. Regarding the single service of RAN slicing, the researchers focused on the price of slice resources [3, 8, 14–16] and the throughput [17, 18]. In [3], wang et al. studied resource pricing as a Stackelberg pricing game, in which network slice customers (NSCs) maximize their profit by adjusting their slice's resource requirement. Lee et al. [18] considered the multitenant cellular network slicing, aiming to achieve a higher network throughput, fairness, and QoS performance.

As to multiservice network resource allocation, throughput [10, 19], delay tolerance [20, 21], operator revenue [22, 23], and the number of network slicing [24] were analyzed with the RAN slicing. Especially, in [19], a joint eMBB and URLLC scheduler was proposed, which can meet the requirement of URLLC while maximizing the utility of eMBB traffic. The radio resource allocation of different network slices was utilized in the downlink fog RAN in [20], where the network was logically divided into high transmission rate slices and low delay slices, and the main objective was to minimize delay tolerance. In [21], Alsenwi et al. proposed proportional fair resource allocation formula that allocates resources to incoming URLLC traffic, while ensuring the reliability of eMBB and URLLC. Wang et al. in [23] studied network slice dimensioning with resource pricing policy, by exploring the relationship between resource efficiency and profit maximization, aiming to maximize the operator's revenue from the perspective of slice price. To solve the multiservice resource allocation problem in the interslice, Smpokos et al. [24] designed a scheduling algorithm that the scheduler schedules a single type of service with their features in each slice. Although the above authors mentioned the multiservice virtual network with network slicing, the design of the spectral efficiency with multiservice network slicing and URLLC network reliability were not considered.

3. System Model and Problem Formulation

In this section, we first introduce a slice model of a downlink orthogonal frequency division multiple access (OFDMA) network. Then, the types of slices are analyzed. Finally, we formulate a spectral-efficient maximization problem with

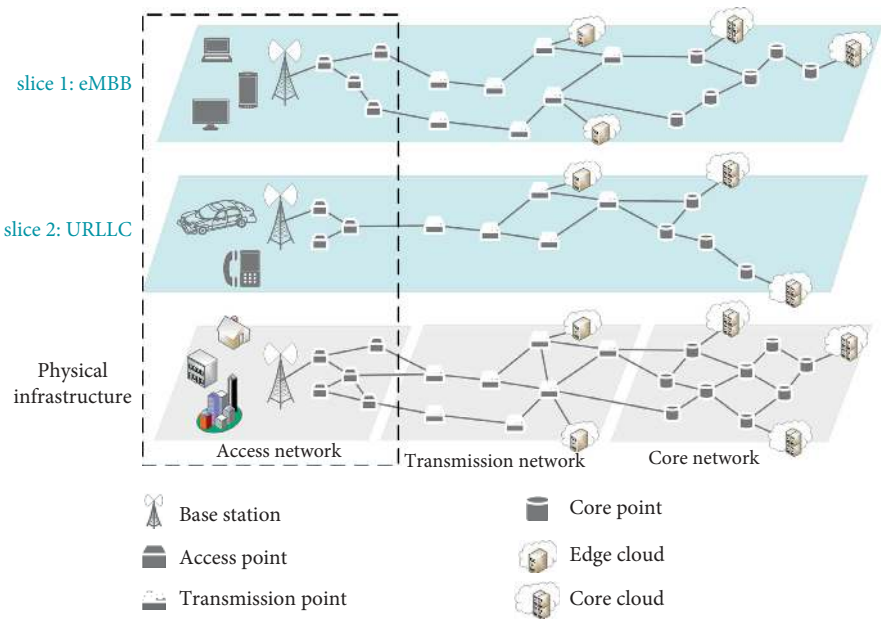

Figure 1: The downlink transmission architecture for multislice connectivity.

slices design in the constraints of the subcarrier and the power.

3.1. System Model. We consider a wireless virtual network with a single cell downlink OFDMA system. The users $\mathcal{K} = \{1, 2, \ldots, K\}$ are randomly located in the coverage area of the base station (BS). The allocation round is one transmission time interval (TTI) [25]. During each TTI, the total bandwidth of the system is divided into N subbandwidth B on average and its corresponding subcarrier sets are $\mathcal{S}_n = \{1, 2, \ldots, N\}$. These resource blocks are shared by two slices, i.e., eMBB slice with user set $\mathcal{K}_1, |\mathcal{K}_1| = K_1$ and URLLC slice with user set $\mathcal{K}_2, |\mathcal{K}_2| = K_2$, where $\mathcal{K} = \mathcal{K}_1 \cup \mathcal{K}_2$ and $\mathcal{K}_1 \cap \mathcal{K}_2 = \varnothing$. In this model, we aim to ensure that the two kinds of slices meet different users' requirement when the spectral efficiency of the network is maximized.

3.2. Problem Formation. For OFDMA downlink wireless virtual network, we apply the slicing technology to solve problems with two types of traffic, i.e., eMBB and URLLC slices. Next, we analyze the constraints of the two kinds of traffic slices, respectively.

3.2.1. Slice 1: eMBB (Throughput Priority). When user k of slice 1 sends a request to a mobile virtual network operator, the corresponding throughput is given by

$$r_k = \sum_{n=1}^{N} a_{n,k} r_{n,k} \geq \sigma_0, \quad k \in \mathcal{K}_1, \tag{1}$$

where

$$a_{n,k} = \begin{cases} 1, & \text{assign subcarrier } n \text{ to user } k, \\ 0, & \text{others}, \end{cases} \tag{2}$$

$$r_{n,k} = B \log\left(1 + \frac{g_{n,k} p_n}{N_0 B}\right), \tag{3}$$

in which $r_{n,k}$ is the data rate of user k with respect to (w.r.t) subcarrier n; N_0 is the noise spectral density; $g_{n,k}$ and p_n denote the channel gain and the transmit power between the BS and user k w.r.t subcarrier n, respectively; and σ_0 is the throughput requirement of user k.

3.2.2. Slice 2: URLLC (Delay Priority). For the users with delay requirement, we assume that the maximum data arrival rate of users in each slice follows a Poisson distribution, and the packet length of each user follows an exponential distribution. From each user's data packet sequence, we infer that the overall system can be regarded as an $M/M/1$ system. Then, the corresponding delay outage probability of user k in slice 2 is given by

$$P_r\left\{D_k \geq D_{k,\max}\right\} = e^{-(R_k - d_{k,\max}) D_{k,\max}}, \quad k \in \mathcal{K}_2, \tag{4}$$

where D_k and $D_{k,\max}$ are the delay of user k and the maximum delay that user k tolerates, respectively, and $d_{k,\max}$ is the maximum data arrival rate of user k. In addition, we assume the maximum transmit power of BS is P_0. Combining the above two slices, the system throughput $R = \sum_{k=1}^{K} \sum_{n=1}^{N} a_{n,k} B \log(1 + (g_{n,k} p_n / N_0 B))$ and total bandwidth NB, the spectral-efficient problem can be formulated as

$$\max_{\{p_n,a_{n,k}\}} \frac{1}{N}\sum_{k=1}^{K}\sum_{n=1}^{N} a_{n,k}\log\left(1+\frac{g_{n,k}p_n}{N_0B}\right), \tag{5a}$$

$$\text{s.t.} \sum_{n=1}^{N} a_{n,k}r_{n,k} \geq \sigma_0, \quad k \in \mathscr{K}_1, \tag{5b}$$

$$P_r\left\{D_k \geq D_{k,\max}\right\} \leq \varepsilon, \quad k \in \mathscr{K}_2, \tag{5c}$$

$$\sum_{k=1}^{K}\sum_{n=1}^{N} a_{n,k}p_n = P_0, \quad p_n \in \mathbb{R}^+, \tag{5d}$$

$$\sum_{k=1}^{K}\sum_{n=1}^{N} a_{n,k} = N, \quad a_{n,k} \in \{0,1\}, \tag{5e}$$

$$\sum_{k=1}^{K} a_{n,k} \leq 1, \quad n \in \mathscr{I}_N, \tag{5f}$$

where (5a) aims to maximize the spectral efficiency of the network system; (5b) ensures that the transmission rate received by user k is no less than the rate requirement; and (5c) guarantees the low outage probability e of user k in slice 2. For (5d)–(5f), we assume that all users make full use of the subcarriers with the maximal transmit power, and each subcarrier associates one user at most.

4. Joint Power and Subcarrier Allocation

Since the original problem (5a)–(5f) is difficult to solve, an approximation method is adopted. From [26], the original problem (5a)–(5f) can be approximated as

$$\max_{\{p_k,N_k\}} \frac{1}{N}\sum_{k=1}^{K} N_k\log\left(1+\frac{\gamma_k p_k}{N_k}\right), \tag{6a}$$

$$\text{s.t.} \quad \sigma_0 - R_k \leq 0, \quad k \in \mathscr{K}_1, \tag{6b}$$

$$P_r\left\{D_k \geq D_{k,\max}\right\} - \varepsilon \leq 0, \quad k \in \mathscr{K}_2, \tag{6c}$$

$$\sum_{k=1}^{K} p_k - P_0 = 0, \quad p_k \in \mathbb{R}^+, \tag{6d}$$

$$\sum_{k=1}^{K} N_k - N = 0, \quad N_k \in \mathbb{Z}^+, \tag{6e}$$

where N_k is the number of subcarrier assigned to user k, $R_k = N_kB\log(1+(\gamma_k p_k/N_k))$, γ_k is the channel gain-to-noise ratio (CNR) for user k, and p_k is transmit power of user k [26]. Equations (6d) and (6e) are to make full use of the radio resource. Unfortunately, the problem in (6a)–(6e) is a mixed-integer programming problem, which is nonconvex. We can rewrite (6a) as

$$\min_{\{p_k,N_k\}} -\frac{1}{N}\sum_{k=1}^{K} N_k\log\left(1+\frac{\gamma_k p_k}{N_k}\right), \tag{7}$$

where (6a) and (7) are equivalent. The optimization problem in (6a)–(6e) is formulated by combining (7) and (6b)–(6e). Since the variable N_k is discrete, we can see that the nonconvex optimization problem in (6a)–(6e) is a mixed-integer programming problem. The branch and bound method can be applied to solve the problem, combing the relaxed problem. By relaxing the variable N_k and enlarging its range: $N_k \in \mathscr{A}$, where $\mathscr{A} = \{N_k \in \mathbb{R}^+ \mid \sum_{k\in\mathscr{K}} N_k = N\}$, we can obtain the relaxed problem:

$$\min_{\{p_k,N_k\}} -\frac{1}{N}\sum_{k=1}^{K} N_k\log\left(1+\frac{\gamma_k p_k}{N_k}\right), \tag{8a}$$

$$\text{s.t.} \quad \sigma_0 - BN_k\log\left(1+\frac{\gamma_k p_k}{N_k}\right) \leq 0, \quad k \in \mathscr{K}_1, \tag{8b}$$

$$e^{-\left(R_k-d_k\right)D_{k,\max}} - \varepsilon \leq 0, \quad k \in \mathscr{K}_2, \tag{8c}$$

$$\sum_{k=1}^{K} p_k - P_0 = 0, \quad p_k \in \mathbb{R}^+, \tag{8d}$$

$$N_k \in \mathscr{A}. \tag{8e}$$

It can be proved that the problem in (8a)–(8e) is a convex optimization problem, which is exhibited in the following theorem.

Theorem 1. *The problem in (8a)–(8e) is convex.*

Proof. See Appendix.

From Theorem 1, the convex optimization problem in (8a)–(8e) can be solved by the PHR augmented Lagrangian algorithm [27, 28]. We first relax the inequality constraint by introducing auxiliary variables x_k and y_k to enable the equality in (8b) and (8c). Then, we can get the augmented Lagrangian function L_ρ in (9), where $p = [p_1, p_2, \ldots, p_K]$; $n = [N_1, N_2, \ldots, N_K]$; $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_{K_1}]$; $\beta = [\beta_1, \beta_2, \ldots, \beta_{K_2}]$; α_k, β_k, μ, and λ is the Lagrange multiplier; and ρ is the penalty factor. Next, we further eliminate the variables to reduce the complexity of the algorithm. We calculate the partial derivatives with respect to the auxiliary variables, which can be written as

$$L_\rho(p,n,\alpha,\beta,\mu,\lambda) = -\frac{R}{NB} + \sum_{k\in\mathscr{K}_1}\left\{\alpha_k\left(\sigma_0 - N_k B\log\left(1+\frac{\gamma_k p_k}{N_k}\right) + x_k^2\right) + \frac{\rho}{2}\left(\sigma_0 - N_k B\log\left(1+\frac{\gamma_k p_k}{N_k}\right) + x_k^2\right)^2\right\} + \sum_{k\in\mathscr{K}_2}\left\{\beta_k\left(e^{-(R_k-d_k)D_{k,\max}} - \varepsilon + y_k^2\right) + \frac{\rho}{2}\left(e^{-(R_k-d_k)D_{k,\max}} - \varepsilon + y_k^2\right)^2\right\} + \mu\left(P_0 - \sum_{k=1}^{K} p_k\right) + \frac{\rho}{2}\left(\sum_{k=1}^{K} p_k - P_0\right)^2 + \lambda\left(N - \sum_{k=1}^{K} N_k\right) + \frac{\rho}{2}\left(\sum_{k=1}^{K} N_k - N\right)^2. \tag{9}$$

$$\frac{\partial L_\rho}{\partial x_k} = 2x_k^3 + 2\alpha_k x_k + 2\rho(\sigma_0 - R_k)\alpha_k x_k, \quad k \in \mathscr{K}_1. \tag{10}$$

By setting (10) to be zero, we can determine the auxiliary variable that minimizes the augmented Lagrange function. Then, we obtain

$$x_k^2 = \max\left\{0, R_k - \sigma_0 - \frac{\alpha_k}{\rho}\right\}, \tag{11}$$

$$x_k^2 = \frac{1}{2}\left|R_k - \sigma_0 - \frac{\alpha_k}{\rho}\right| + \frac{1}{2}\left(\sigma_k - \sigma_0 - \frac{\alpha_k}{\rho}\right). \tag{12}$$

Plugging (12) into (9), we get

$$\sum_{k\in\mathscr{K}_1}\alpha_k\left(\sigma_0 - R_k + x_k^2\right) + \frac{\rho}{2}\left(\sigma_0 - R_k + x_k^2\right)^2 = \frac{1}{2\rho}\sum_{k\in\mathscr{K}_1}\left(\min\{0,\rho(R_k-\sigma_0)-\alpha_k\}\right)^2 - \alpha_k^2. \tag{13}$$

Similarly, when $k \in \mathscr{K}_2$, we can easily obtain

$$y_k^2 = \max\left\{0, \varepsilon - P_r\{D_k \geq D_{k,\max}\} - \frac{\beta_k}{\rho}\right\} = \frac{1}{2}\left|\varepsilon - P_r\{D_k \geq D_{k,\max}\} - \frac{\alpha_k}{\rho}\right| + \frac{1}{2}\left(\varepsilon - P_r\{D_k \geq D_{k,\max}\} - \frac{\beta_k}{\rho}\right). \tag{14}$$

Thus,

$$\sum_{k\in\mathscr{K}_2}\beta_k\left(P_r\{D_k \geq D_{k,\max}\} - \varepsilon + y_k^2\right) + \frac{\rho}{2}\left(P_r\{D_k \geq D_{k,\max}\} - \varepsilon + y_k^2\right)^2 = \frac{1}{2\rho}\sum_{k\in\mathscr{K}_1}\left(\min\{0,\rho(\varepsilon - P_r\{D_k \geq D_{k,\max}\}) - \beta_k\}\right)^2 - \beta_k^2. \tag{15}$$

Based on (13) and (4), the augmented Lagrangian function can be formulated as (16). In order to facilitate the description of the algorithm, we define

$$L_\rho(p,n,\alpha,\beta,\mu,\lambda) = -\frac{R}{NB} + \frac{1}{2\rho}\sum_{k\in\mathscr{K}_1}\left(\min\{0,\rho(R_k-\sigma_0)-\alpha_k\}\right)^2 + \frac{1}{2\rho}\sum_{k\in\mathscr{K}_2}\left(\min\{0,\rho(\varepsilon - P_r\{D_k \geq D_{k,\max}\}) - \alpha_k\}\right)^2 + \mu\left(P_0 - \sum_{k=1}^{K} p_k\right) + \frac{\rho}{2}\left(\sum_{k=1}^{K} p_k - P_0\right)^2 + \lambda\left(N - \sum_{k=1}^{K} N_k\right) + \frac{\rho}{2}\left(\sum_{k=1}^{K} N_k - N\right)^2. \tag{16}$$

$$f(p_k, N_k) = BN_k\log\left(1+\frac{\gamma_k p_k}{N_k}\right) - \sigma_0, \tag{17}$$

$$g(p_k, N_k) = \varepsilon - e^{-(R_k-d_k)D_{k,\max}}, \tag{18}$$

and the equality function

$$h(p) = \sum_{k=1}^{K} p_k - P_0, \qquad \tilde{h}(n) = \sum_{k=1}^{K} N_k - N. \tag{19}$$

In addition, the gradient of equality function and inequality function are used in the PHR augmented Lagrangian algorithm, which are, respectively, shown as

Initialize $\alpha, \beta, \mu, \lambda, \rho > 0$, $i = 0$, $\theta \in (0,1)$, $i_{\max} = 500$, $V_{i,\text{old}} = 10$, $\eta > 1$, $\delta \in [0,1]$.
while $V_i > \delta$ and $i < i_{\max}$ **do**
  Solve augmented Lagrangian function with fixed $\rho, \alpha, \beta, \mu, \lambda$, through BFGS algorithm with (10) to obtain $\mathcal{X}^i$, where $\mathcal{X}^i = \{p_k^i, N_k^i \mid k = 1, 2, \ldots, K\}$.
  $V_i = \omega$, where ω is shown in (22).
  if $V_i > \delta$ then
    If $i \geq 2$ and $V_i > \theta V_{i,\text{old}}$ **then**
      $\rho = \eta\rho$
    End if
    $\alpha_k = \max\{0, \alpha_k - \rho f(p_k, N_k)\}$, $k \in \mathcal{K}_1$
    $\beta_k = \max\{0, \beta_k - \rho g(p_k, N_k)\}$, $k \in \mathcal{K}_2$
    $\mu = \mu - \rho(\Sigma_{k=1}^{K} - P_0)$
    $\lambda = \lambda - \rho(\sum_{k=1}^{K} N_k - N)$.
  End if
  $i = i + 1$, $x_0 = x$.
End while

Algorithm 1: The Relaxed optimization problem.

$$\begin{aligned}
\nabla g(p_k, N_k) &= \left[\frac{\partial g}{\partial p_k}, \frac{\partial g}{\partial N_k}\right], \quad k \in \mathcal{K}_1, \\
\nabla f(p_k, N_k) &= \left[\frac{\partial f}{\partial p_k}, \frac{\partial f}{\partial N_k}\right], \quad k \in \mathcal{K}_2, \\
\nabla h(p) &= \left[\frac{\partial h}{\partial p_1}, \frac{\partial h}{\partial p_1}, \ldots, \frac{\partial h}{\partial p_K}\right], \\
\nabla \tilde{h}(n) &= \left[\frac{\partial \tilde{h}}{\partial N_1}, \frac{\partial \tilde{h}}{\partial N_1}, \ldots, \frac{\partial \tilde{h}}{\partial N_K}\right],
\end{aligned} \tag{20}$$

where

$$\begin{aligned}
\frac{\partial f}{\partial p_k} &= \frac{BN_k\gamma_k}{(N_k + \gamma_k p_k)\ln 2}, \\
\frac{\partial f}{\partial N_k} &= B\log\left(1 + \frac{\gamma_k p_k}{N_k}\right) - \frac{B\gamma_k p_k}{(N_k + \gamma_k p_k)\ln 2}, \\
\frac{\partial g}{\partial p_k} &= \frac{BN_k\gamma_k D_{k,\max}}{(N_k + \gamma_k p_k)\ln 2} e^{-(R_k - d_k)D_{k,\max}}, \\
\frac{\partial g}{\partial N_k} &= BD_{k,\max}\log\left(1 + \frac{\gamma_k p_k}{N_k}\right) e^{-(R_k - d_k)D_{k,\max}} \\
&\quad - BD_{k,\max}\frac{\gamma_k p_k}{(N_k + \gamma_k p_k)\ln 2} e^{-(R_k - d_k)D_{k,\max}}, \\
\frac{\partial h}{\partial p_k} &= 1, \quad k \in \mathcal{K}, \\
\frac{\partial \tilde{h}}{\partial N_k} &= 1, \quad k \in \mathcal{K}.
\end{aligned} \tag{21}$$

The relaxed problem is solved by the PHR augmented Lagrangian algorithm, which is shown in Algorithm 1. For the optimal subcarrier and power allocation, we apply Algorithm 1 and the branch and bound method to solve it. By relaxing the integer constraints in the mixed-integer programming problem, each relaxed problem is solved by Algorithm 1. By adopting the joint power and subcarrier allocation (JPSA) algorithm in Algorithm 2, the iteration is carried out with the branch and bound method until the obtained integer solution satisfies the iterative condition. Therefore, we can obtain the optimal resource allocation and the spectral efficiency s:

$$\begin{aligned}
\omega = \Bigg(&\left(\sum_{k=1}^{K} p_k - P_0\right)^2 + \left(\sum_{k=1}^{K} N_k - N\right)^2 \\
&+ \sum_{k \in \mathcal{K}_1}\left[\min\left(f(p_k, N_k), \frac{\alpha_k}{\rho}\right)\right]^2 \\
&+ \sum_{k \in \mathcal{K}_2}\left[\min\left(g(p_k, N_k), \frac{\beta_k}{\rho}\right)\right]^2\Bigg)^{1/2}
\end{aligned} . \tag{22}$$

5. Simulation

In this section, we evaluate the effectiveness of the proposed JPSA algorithm and compare it with the corresponding relaxed problem. The adaptive particle swarm optimization (APSO) [29], the equal power allocation (EPA), and the equal subcarrier allocation (ESA) algorithm are compared with the JPSA algorithm, where in the EPA the transmit power for each user from BS is P_0/K and in the ESA the assigned subcarrier number for each user is N/K. In order to describe the result of simulation conveniently, we assume that users in the same slice require the same requirement of throughput or delay but different CNRs. The total number of users is 16, the total bandwidth of the network system is 5 MHz, and the number of subcarriers is 20. To show the utility of network slicing, we evaluate the reliability of the network system with a single slice (eMBB slice or URLLC slice) with that of the network system supporting both eMBB slices and URLLC slices. In Figure 2, we evaluate the spectral efficiency design of the proposed algorithm and other algorithms versus the total power. We can see that the proposed joint JPSA algorithm improves the spectral efficiency more effectively. When the total power is small, the APSO

Solve the nonlinear programming problem of the relaxed N in (8a)–(8e) by Algorithm 1, we get l the lower bound of (7) and the initial resource allocation N_k^*, p_k^*.
While $\|u - l\| \le \varepsilon$ **do**
If $N_k^* \in \mathbb{Z}$ **then**
Break
Else
Find the upper bound of the problem. Initialize the subcarrier number $N_i \in \mathbb{Z}$ and the power $P_i, i = 1, 2, \ldots, K$. We determine the feasible solution, and we obtain u as an upper bound of (7).
End if
Branch the problem in (7) into problems $\mathscr{P}_1$ with $N_k \le \lceil N_k^* \rceil$ and $\mathscr{P}_2$ with $N_k \ge \lceil N_k^* \rceil + 1$ by the branch and bound method. We solve the two problems $\mathscr{P}_1$ and $\mathscr{P}_2$ that are still the convex optimization problem. Combining Algorithm 1, we get $u_1, N_k^{(1)}, p_k^{(1)}$ and $u_2, N_k^{(2)}, p_k^{(2)}$.
If $\exists u_i, i = 1, 2, \forall N_k \in \mathbb{Z}$ **then**
$u = \min\{u_i, u\}$ or $u = \min\{u_1, u_2, u\}$
Else
If $u_i < u$ **then**
$l = \max\{u_1, u_2, l\}, N_k^* = N_k^{(i)}$
End if
End if
$s = u$
End while
Output: p_k^*, N_k^*, s

Algorithm 2: Joint power and subcarrier allocation (JPSA).

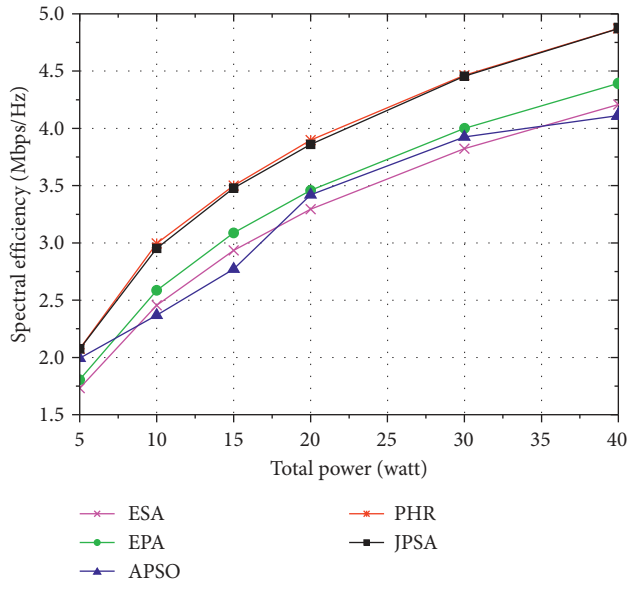

Figure 2: The spectral efficiency performance versus the total power with $\gamma_k = 5$ dB of eMBB and $\gamma_k = 15$ dB of URLLC. It shows the network using the JPSA algorithm has better frequency efficiency performance.

[29] and our proposed method are close to the optimal relaxed solution. When the total power is large, the spectral efficiency of APSO [29], ESA, and EPA are not as high as that of the JPSA algorithm. It can be seen from Figure 2 that the solution of the APSO is not stable enough and that the greater the total power of the system is, the more effective the JPSA algorithm is to improve the spectral efficiency.

In Figure 3, we evaluate the spectral efficiency of the eMBB slice, the URLLC slice and the overall network system under different user throughput requirements of the eMBB slice. When σ_0 increases, the spectral efficiency of the eMBB slice is almost the same, and the spectral efficiency of the URLLC slice and the whole system decreases. Therefore, when we pay more attention to the eMBB slice, we can appropriately increase the throughput requirement of the users.

As can be seen from Figure 4, the network slice of URLLC is more reliable when the delay requirement is

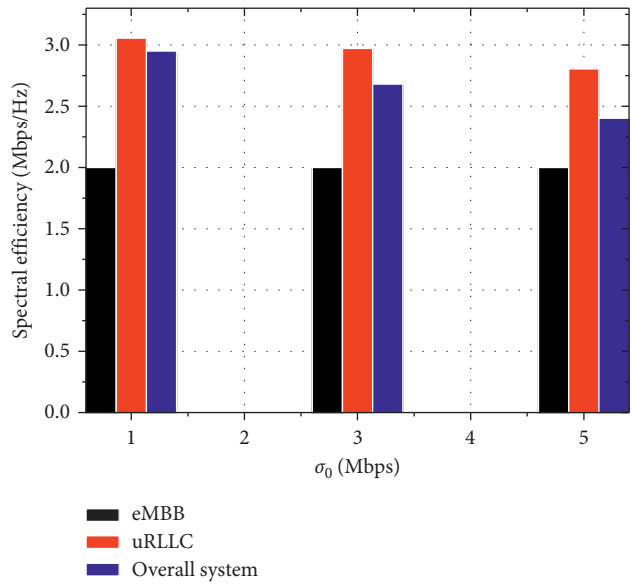

Figure 3: The spectral efficiency performance of eMBB slice, URLLC slice, and the total system versus the rate requirement of the eMBB slice σ_0 = 1 Mbps, 3 Mbps, 5 Mbps.

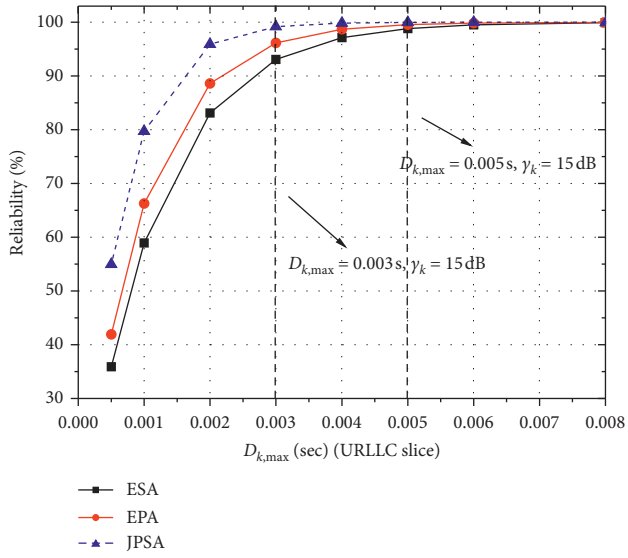

Figure 4: Reliability of URLLC versus $D_{k,\max}$. The reliability of the JPSA algorithm obviously outperforms other algorithms, when $D_{k,\max} < 0.005$ sec.

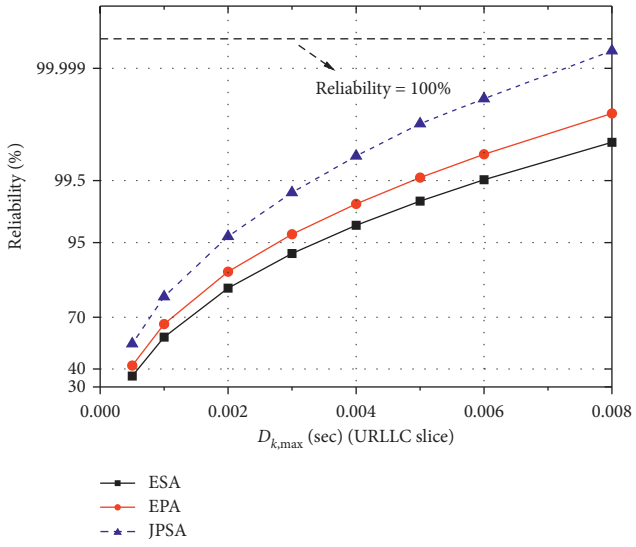

Figure 5: For reliability versus $D_{k,\max}$ of URLLC, the magnification of the corresponding figure part shows the proposed algorithm makes URLLC slice more reliable.

greater than 0.005 sec. When there is a higher delay requirement of $D_{k,\max} \leq 0.003$ sec, the URLLC slice is more reliable when using the JPSA algorithm, and its overall reliability is higher than the EPA and ESA algorithms. In Figure 5, in order to facilitate the analysis of the part of Figure 4 whose reliability is close to 1, we increase the magnification ratio of the corresponding figure part as the reliability increases. It is obvious that our proposed algorithm makes URLLC slice more reliable, especially when the reliability of the URLLC slice is larger than 90%.

The cumulative distribution function of the low delay requirement for URLLC slices with different CNRs is analyzed in Figure 6. It can be seen from the figure that when $\gamma_k \geq 10$ dB and $d_{\max} \geq 0.004$ sec, the probability of $D \leq D_{\max}$ is close to 1. From the overall curve, we can infer that when $\gamma_k \geq 10$ dB, the cumulative distribution probability is relatively high. That is, the slice URLLC has higher reliability.

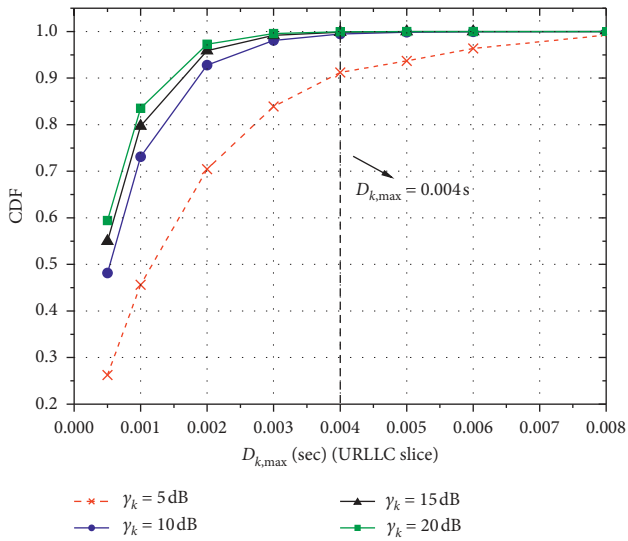

FIGURE 6: CDF versus $D_{k,\max}$ of URLLC with different CNRs with different resource allocation. When $\gamma_k > 10$ dB, the slice URLLC has a higher reliability.

6. Conclusion

In this paper, we studied the spectral efficiency with the eMBB slice and the URLLC slice for OFDMA virtual RAN. We proved that the approximately transformed continuous programming problem is a convex optimization problem, which can be solved by applying the PHR. We proposed a JPSA algorithm combined with the PHR and the branch and bound method. The simulation results show that the proposed algorithm improves the spectral efficiency performance of the network system while ensuring the users' requirement of the eMBB slice and the users' requirement of the URLLC slice with a high probability. In the paper, we have solved how to allocate virtual network resources in two types of slices in a single cell to ensure that the admitted users meet the requirement and maximize the spectrum efficiency of the system. In future research, we will also consider the problem of slice admission and resource allocation for slices in multiple cells.

Appendix

Proof of Theorem 1

We first prove that $\phi(x, y) = x\log(1 + (y/x))$ is a concave function while $x > 0$ and $y > 0$. By calculating the second partial derivative of the function $\phi(x, y)$, we can get the corresponding Hessian matrix:

$$H = \begin{bmatrix} \dfrac{\partial^2\phi(x,y)}{\partial x^2} & \dfrac{\partial^2\phi(x,y)}{\partial x\,\partial y} \\ \dfrac{\partial^2\phi(x,y)}{\partial y\,\partial x} & \dfrac{\partial^2\phi(x,y)}{\partial x^2} \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}, \tag{A.1}$$

where

$$\begin{aligned} b_{11} &= -\frac{y^2}{x^3(1+(y/x))^2\ln 2}, \\ b_{12} &= \frac{y}{x^2(1+(y/x))^2\ln 2}, \\ b_{21} &= \frac{y}{x^2(1+(y/x))^2\ln 2}, \\ b_{22} &= -\frac{1}{x(1+(y/x))^2\ln 2}. \end{aligned} \tag{A.2}$$

Combining $b_{11} = -(y^2/x^3(1+y/x)^2\ln 2) < 0$ and $b_{11}b_{22} - b_{12}b_{21} = 0$, we can obtain that the corresponding Hessian matrix of the function $\phi(x, y)$ is negative semidefinite. Without loss of generality, we consider the case of $K = 2$ in (8a), i.e., $\psi = -(1/N)\sum_{k=1}^{2}\phi(x_k, y_k)$, where $y_k = \gamma_k p_k > 0$ and $x_k = N_k > 0$. The Hessian matrix of the function ψ is

$$\tilde{H} = \begin{bmatrix} \dfrac{\partial^2\psi}{\partial x_1^2} & \dfrac{\partial^2\psi}{\partial x_1\partial y_1} & \dfrac{\partial^2\psi}{\partial x_1\partial x_2} & \dfrac{\partial^2\psi}{\partial x_1\partial y_2} \\ \dfrac{\partial^2\psi}{\partial y_1\partial x_1} & \dfrac{\partial^2\psi}{\partial y_1^2} & \dfrac{\partial^2\psi}{\partial y_1\partial x_2} & \dfrac{\partial^2\psi}{\partial y_1\partial y_2} \\ \dfrac{\partial^2\psi}{\partial x_2\partial x_1} & \dfrac{\partial^2\psi}{\partial x_2\partial y_1} & \dfrac{\partial^2\psi}{\partial x_2^2} & \dfrac{\partial^2\psi}{\partial x_2\partial y_2} \\ \dfrac{\partial^2\psi}{\partial y_2\partial x_1} & \dfrac{\partial^2\psi}{\partial y_2\partial y_1} & \dfrac{\partial^2\psi}{\partial y_2\partial x_2} & \dfrac{\partial^2\psi}{\partial y_2^2} \end{bmatrix}, \tag{A.3}$$

$$\tilde{H} = \begin{bmatrix} c_{11} & c_{12} & 0 & 0 \\ c_{21} & c_{22} & 0 & 0 \\ 0 & 0 & c_{33} & c_{34} \\ 0 & 0 & c_{43} & c_{44} \end{bmatrix}, \tag{A.4}$$

where

$$\begin{aligned} c_{11} &= \frac{y_1^2}{Nx_1^3(1+(y_1/x_1))^2\ln 2}, \\ c_{12} &= -\frac{y_1}{Nx_1^2(1+(y_1/x_1))^2\ln 2}, \\ c_{21} &= -\frac{y_1}{Nx_1^2(1+(y_1/x_1))^2\ln 2}, \\ c_{22} &= \frac{1}{Nx_1(1+(y_1/x_1))^2\ln 2}, \\ c_{33} &= \frac{y_2^2}{Nx_2^3(1+(y_2/x_2))^2\ln 2}, \\ c_{34} &= -\frac{y_2}{Nx_2^2(1+(y_2/x_2))^2\ln 2}, \\ c_{43} &= -\frac{y_2}{Nx_2^2(1+(y_2/x_2))^2\ln 2}, \\ c_{44} &= \frac{1}{Nx_2(1+(y_2/x_2))^2\ln 2}. \end{aligned} \tag{A.5}$$

Obviously, when $K = 2$, the Hessian matrix is positive semidefinite. We can determine whether the matrix (A.3) is a positive semidefinite matrix by analyzing its leading principle minor. There are four leading principle minors, one of order 1 $c_{11} = (y_1^2/Nx_1^3(1 + y_1/x_1)^2 \ln 2) > 0$, one of order 2 $c_{11}c_{22} - c_{12}c_{21} = 0$, and one of order 3 and one of order 4 are zero. Similarly, $K > 2$ also holds. Since one of order 1 $c_{11} = (y_1^2/Nx_1^3(1 + y_1/x_1)^2 \ln 2) > 0$, one of order 2 $c_{11}c_{22} - c_{12}c_{21} = 0$, one of order 3 and one of order 4 are zero, one of order 5, one of order 6, and so on, the leading principle minor whose order is greater than or equal to 2 is 0. Thus, when $K \geq 2$, the Hessian matrix of the function we proposed is positive semidefinite. (8a) is a convex function. Otherwise, the inequality constraint function (8b) and (8c) are convex by [30] and convexity of $\phi(x, y)$. The two equality constraints (8d) and (8e) satisfy the affine transformation. Regarding [30], we obtain this problem (8a)–(8e) is a convex optimization problem.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work is supported by the National Natural Science Foundation of China under Grant No. 61971057.

References

[1] L. Li, N. Deng, W. Ren, B. Kou, W. Zhou, and S. Yu, "Multi-service resource allocation in future network with wireless virtualization," *IEEE Access*, vol. 6, pp. 53854–53868, 2018.

[2] S. Zhang, "An overview of network slicing for 5G," *IEEE Wireless Communications*, vol. 26, no. 3, pp. 111–117, 2019.

[3] G. Wang, G. Feng, S. Qin, R. Wen, and S. Sun, "Optimizing network slice dimensioning via resource pricing," *IEEE Access*, vol. 7, pp. 30331–30343, 2019.

[4] N. Alliance, "5G white paper," in *Next Generation Mobile Networks*, pp. 1–125, NGMN, Frankfurt, Germany, 2015.

[5] S. Redana, *View on 5G Architecture*, pp. 1–61, July 2016, https://www.researchgate.net/publication/306107214.

[6] J. Ni, X. Lin, and X. S. Shen, "Efficient and secure service-oriented authentication supporting network slicing for 5G-enabled IoT," *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 3, pp. 644–657, Mar. 2018.

[7] R. Alvizu, S. Troia, V. M. Nguyen, G. Maier, and A. Pattavina, "Network orchestration for dynamic network slicing for fixed and mobile vertical services," in *Proceedings of the Optical Fiber Communications Conference and Exposition (OFC)*, pp. 1–3, San Diego, CA, USA, March 2018.

[8] M. Rayani, D. Naboulsi, R. Glitho, and H. Elbiaze, "Slicing virtualized EPC-based 5G core network for content delivery," in *Proceedings of the IEEE Symposium on Computers and Communications (ISCC)*, pp. 726–729, Natal, Brazil, June 2018.

[9] C. Chang, N. Nikaein, and T. Spyropoulos, "Radio access network resource slicing for flexible service execution," in *Proceedings of the IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 668–673, Honolulu, HI, USA, April 2018.

[10] P. Popovski, K. F. Trillingsgaard, O. Simeone, and G. Durisi, "5G wireless network slicing for eMBB, URLLC, and mMTC: a communication-theoretic view," *IEEE Access*, vol. 6, pp. 55765–55779, 2018.

[11] W. Guan, X. Wen, L. Wang, Z. Lu, and Y. Shen, "A service-oriented deployment policy of end-to-end network slicing based on complex network theory," *IEEE Access*, vol. 6, pp. 19691–19701, 2018.

[12] S. Vassilaras, L. Gkatzikis, N. Liakopoulos et al., "The algorithmic aspects of network slicing," *IEEE Communications Magazine*, vol. 55, no. 8, pp. 112–119, 2017.

[13] P. Caballero, A. Banchs, G. de Veciana, X. Costa-Perez, and A. Azcorra, "Network slicing for guaranteed rate services: admission control and resource allocation games," *IEEE Transactions on Wireless Communications*, vol. 17, no. 10, pp. 6419–6432, 2018.

[14] P. Rost, C. Mannweiler, D. S. Michalopoulos et al., "Network slicing to enable scalability and flexibility in 5G mobile networks," *IEEE Communications Magazine*, vol. 55, no. 5, pp. 72–79, 2017.

[15] D. Bega, M. Gramaglia, A. Banchs, V. Sciancalepore, K. Samdanis, and X. Costa-Perez, "Optimising 5G infrastructure markets: the business of network slicing," in *Proceedings of the IEEE Conference on Computer Communications*, pp. 1–9, Atlanta, GA, USA, May 2017.

[16] V. N. Ha and L. B. Le, "End-to-End network slicing in virtualized OFDMA-based cloud radio access networks," *IEEE Access*, vol. 5, pp. 18675–18691, 2017.

[17] D. Nojima, "Resource isolation in RAN Part While utilizing ordinary scheduling algorithm for network slicing," in *Proceedings of the IEEE Vehicular Technology Conference (VTC Spring)*, pp. 1–5, Porto, Portugal, June 2018.

[18] Y. L. Lee, J. Loo, T. C. Chuah, and L.-C. Wang, "Dynamic network slicing for multitenant heterogeneous cloud radio access networks," *IEEE Transactions on Wireless Communications*, vol. 17, no. 4, pp. 2146–2161, 2018.

[19] A. Anand, G. De Veciana, and S. Shakkottai, "Joint scheduling of URLLC and eMBB traffic in 5G wireless networks," in *Proceedings of the IEEE Conference on Computer Communications (IEEE INFOCOM)*, pp. 1970–1978, Honolulu, HI, USA, April 2018.

[20] T. Dang and M. Peng, "Delay-aware radio resource allocation optimization for network slicing in fog radio access networks," in *Proceedings of the IEEE International Conference on Communications Workshops (ICC Workshops)*, pp. 1–6, Kansas City, MO, USA, May 2018.

[21] M. Alsenwi, N. H. Tran, M. Bennis, A. Kumar Bairagi, and C. S. Hong, "eMBB-URLLC resource slicing: a risk-sensitive approach," *IEEE Communications Letters*, vol. 23, no. 4, pp. 740–743, 2019.

[22] 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 Journal on Selected Areas in Communications*, vol. 37, no. 4, pp. 881–895, 2019.

[23] G. Wang, G. Feng, W. Tan, S. Qin, R. Wen, and S. Sun, "Resource allocation for network slices in 5G with network resource pricing," in *Proceedings of the IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, Singapore, December 2017.

[24] G. Smpokos, A. Lioumpas, T. Mouroutis, Y. Stylianou, and V. Angelakis, "Performance aware resource allocation and traffic aggregation for user slices in wireless HetNets," in *Proceedings of the Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, pp. 1–5, June 2017.

[25] L. Gao, P. Li, Z. Pan, N. Liu, and X. You, "Virtualization framework and VCG based resource block allocation scheme for LTE virtualization," in *Proceedings of the IEEE 83rd Vehicular Technology Conference (VTC Spring)*, pp. 1–6, Nanjing, China, May 2016.

[26] C. Xiong, G. Y. Li, Y. Liu, Y. Chen, and S. Xu, "Energy-efficient design for downlink OFDMA with delay-sensitive traffic," *IEEE Transactions on Wireless Communications*, vol. 12, no. 6, pp. 3085–3095, 2013.

[27] R. Fletcher, *Practical Methods of Optimization*, John Wiley & Sons, Hoboken, NJ, USA, 2013.

[28] E. G. Birgin and J. M. Martínez, "Structured minimal-memory inexact quasi-Newton method and secant preconditioners for augmented lagrangian optimization," *Computational Optimization and Applications*, vol. 39, no. 1, pp. 1–16, 2008.

[29] A. R. Faisal, F. Hashim, N. K. Noordin, M. Ismail, and A. Jamalipour, "Efficient beamforming and spectral efficiency maximization in a joint transmission system using an adaptive particle swarm optimization algorithm," *Applied Soft Computing*, vol. 49, pp. 759–769, 2016.

[30] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.