

Globally Optimal Resource Allocation and Time Scheduling in Downlink Cognitive CRAN Favoring Big Data Requests

MOHAMMAD BIGDELI¹, SHAHROKH FARAHMAND¹, BAHMAN ABOLHASSANI¹,
AND HA H. NGUYEN², (Senior Member, IEEE)

¹School of Electrical Engineering, Iran University of Science and Technology (IUST), Tehran 13114-16846, Iran

²Department of Electrical and Computer Engineering, University of Saskatchewan, Saskatoon, SK S7N 5A9, Canada

Corresponding author: Ha H. Nguyen (ha.nguyen@usask.ca)

ABSTRACT This paper is concerned with a cognitive cloud radio access network (CRAN) with a special attention to efficient and reliable downlink transmission of big data for secondary users (SUs). Existing approaches either try to maximize the number of accepted SUs or the sum data rate of accepted SUs. The first approach unfairly favors users with small data requests, whereas the second approach allocates most resources to users with better channel conditions. In contrast, this paper develops a novel approach that favors big data requests while simultaneously maintaining a certain degree of fairness among SUs. To this end, we first introduce a novel objective function that allows us to jointly optimize deadline-aware time scheduling, spectrum allocation, SU selection, and remote radio head (RRH) allocation for SUs. Second, we demonstrate that finding the global optimum solution entails the enumeration of all colorful independent sets on a generalized interval graph, which is known to be NP-hard. Third, we propose a dynamic programming (DP) approach, which yields the global optimum solution at a reduced computational cost. Fourth, we analyze the complexity of the proposed DP approach and assess its performance against existing baseline algorithms. Simulation results reveal that our solution favors big data users while incurring only a small degradation in the fairness index. Our proposed solution is practical for small-to-medium size networks. Furthermore, it offers an optimum benchmark for any new sub-optimal low-complexity algorithm.

INDEX TERMS Time scheduling, resource allocation, user selection, cloud radio access network (CRAN), big data transmission.

I. INTRODUCTION

Tremendous increases in both the number of mobile devices and their data demand in the last few years [1] have led to many new challenges in the design of modern wireless communication networks. Efficient utilization of time, frequency, and spatial resources is imperative in providing efficient wireless communication of big data. The fifth-generation (5G) [2], [3] and sixth-generation (6G) cellular networks [4] are expected to support both licensed primary users (PUs) and unlicensed secondary users (SUs), as well as much higher data rates [5], [6]. The first commercially developed version of 5G technology is being designed based on specifications published by 3GPP [7], which put a special focus on unlicensed users and spectrum bands [8]. Cisco's annual

internet report emphasizes that more than 10% of the wireless connections will be on 5G by 2023 [1]. Unsurprisingly, current mobile network operators, such as AT&T, Vodafone, and NTT DoCoMo, are evolving from 2G/3G/4G to 5G and 6G [9], [10].

The term "big data" is characterized by five big V features: *Volume*, *Velocity*, *Variety*, *Veracity* and *Value* [11], [12]. Here, *Volume* explains the size of data (in gigabytes) generated per time unit. *Velocity* explains how fast data is generated, and *Variety* refers to various data types which are generated. *Veracity* explains the accuracy and quality of the generated data. Lastly, *Value* refers to the priority or usefulness of data.

To serve the increasing number of mobile devices and handle big data, a massive number of micro/macro base stations, and remote radio heads (RRHs) could be employed, which causes densification of wireless networks. On the other hand, cloud radio access network (CRAN) [13] has

The associate editor coordinating the review of this manuscript and approving it for publication was Fang Yang¹.

been introduced as a revolutionary redesign of the centralized cellular architecture [14], which integrates existing radio access networks to achieve feasible, flexible and scalable solutions [15]. A CRAN consists of three basic components: RRHs, wireless/wired fronthaul links with low-latency and high bandwidth [16], [17], and central baseband processing unit (BBU) pool cloud. Using technologies such as software-defined networks [18], CRAN detaches the BBUs from traditional base stations and integrates a large number of them in the BBU pool. As a result, a CRAN enables cross-layer design, efficient cloud-based resource allocation, optimal central routing and controlling, and load equalization in the network. A use case of CRAN can be seen in resource sharing among several service providers. This helps to reduce the cost of establishing network infrastructure for each operator. In general, service providers rent resources from a network to support data transmission [19]. With such an arrangement, unlike conventional wireless networks that usually have large idle times, BBU pool cloud-based resource allocation can be designed to improve the overall efficiency for big data transmission. In the next subsection, we provide a brief overview of existing literature on this topic and point to their limitations. Then, we will enumerate the contributions of this work.

A. RELATED WORKS

Our proposed scheduling and resource allocation algorithm will answer the following three questions: (1) Which SUs and when are they scheduled for optimal utilization of available resources? (2) How to handle big data requests and prioritize them over small data demands while ensuring a certain degree of fairness? (3) How to develop efficient algorithms that can address these issues with both satisfactory performance and complexity? Focusing on these questions, we categorize related works into two parts: (1) Works that investigate RRH and spectrum allocation, user selection, and scheduling, (2) Works related to big data transmission. A summary of prior art and the topics they address are provided in Table 1.

1) RRH AND SPECTRUM ALLOCATION, USER SELECTION, AND SCHEDULING

The authors in [20] consider constant transmit power during each radio resource block and assign users to these resource blocks and RRHs. The work done in [20] is then extended in [21] to perform joint user scheduling and power allocation by maximizing the overall CRAN data rate. Then, in [22], a low computational complexity graph-based approach is developed to construct a power allocation graph that is responsible for transmitted frame synchronization, power level control and user scheduling. Subsequently, in [23], a multiple CRAN framework is considered for a hybrid scheduling with a constraint that users are assigned to a single or multiple RRHs of a single cloud. In [24], the authors solve the joint user scheduling and beamforming problem to maximize the overall CRAN data rate. In a similar framework, the work in [25] presents a greedy algorithm for a coordinated

interference aware scheduling for the downlink of a CRAN in order to maximize the overall CRAN throughput.

The main goal in [26]–[28] is to minimize the delay in the CRAN. In [26], a queuing model is established to minimize the mean response delay and power by using joint BBU allocation and load scheduling. A two-phase interference aware scheduling for a CRAN is investigated in [27]. First, all users are grouped into clusters based on estimated delay cost function and interference levels. Then, channels are matched to user groups to minimize the sum delay of the system. Under the total power constraint, the authors in [28] minimize the traffic delay in a hierarchical CRAN by using a joint virtual machine scheduling and RRH allocation.

In [19] and [29], the overall throughput of a CRAN is maximized via resource allocation. Resource sharing in a CRAN with fronthaul constraints is studied in [19], where service providers rent radio resources to provide services. A threshold-based version is used to both manage interference between RRHs, and provide minimum resource requirements. A multi-scale global and local channel allocation and user association mechanism is performed at different time scales. Classification and machine learning are examined in [29], which led to a conclusion that a fairer scheduling should be done in a distributed manner.

RRH selection and resource allocation are studied in [30]–[33]. In [30], a traffic-aware RRH clustering algorithm is introduced for efficiently improving the QoS of a CRAN. Then, an optimal spectrum assignment algorithm is designed for clusters. In [31], RRH selection for coordinated multi-point transmission and resource allocation are jointly optimized to improve user's capacity performance in a CRAN. Furthermore, the optimal resource allocation is solved under fixed RRH selection. In [32], MISO downlink multi-cast is studied for a CRAN where an energy-aware joint scheduling and RRH selection for coordinated beamforming are investigated. The goal is to increase CRAN performance by considering interference management and energy efficiency.

Various works on CRANs' uplink and downlink resource sharing have been carried out in [33]–[36]. Joint resource allocation and interference cancellation in full-duplex CRAN are studied in [33] with the goal of optimizing the downlink capacity under transmit power constraints while guaranteeing QoS of the uplink. The orthogonal frequency-division multiple access (OFDMA) and non-orthogonal multiple access (NOMA) [37] techniques are used for spectrum allocation in [34], [35], [38] and [39]. The work in [34] addresses joint downlink and uplink resource sharing in OFDMA-based CRAN. In order to maximize the overall throughput, cooperative RRH selection, scheduling, and power assignment with maximum power constraints are performed. The work in [35] designs a joint RRH association and user scheduling for OFDMA-based CRAN in both uplink and downlink. Various multiple access approaches for cognitive radio networks (CRN) have been examined in [38]. Furthermore, the work in [39] investigates a CRN instead of a cognitive

CRAN with improving energy-efficiency in mind and allowing devices to select between NOMA and orthogonal multiple access options.

The problems of user selection in CRANs are studied in [36], [40], [41]. In particular, the authors in [36] minimize the power consumption of full duplex CRANs by jointly optimizing the transmit powers of users, RRH set, beamforming vector and compression ratio of the fronthaul. To make it more efficient in computational complexity, a two-level algorithm is employed. First, a user selection algorithm, which is based on minimizing square error between the minimum SINR that satisfies QoS requirements and the achievable SINR, is performed to find the largest cluster of users that can be served. Second, power optimization is carried out. The work in [40] applies a generalized Stackelberg game to the problem of overall data rate maximization by joint distributed power allocation and centralized user association in heterogeneous CRANs with guaranteed QoS for users. A two-level approach is developed in [41] to guarantee the minimum data rate for users in the downlink of an ultra-dense CRAN. The coverage probability of a CRAN is maximized by joint frequency allocation and user clustering with a satisfactory computational complexity. In the first level, there is a new binary user clustering. This clustering reduces the complexity of spectrum allocation and guarantees a minimum transmission data rate for users. Then, a new graph-based algorithm is proposed for serving clusters by considering a relationship among them without extra calculations.

Scheduling of computing resources is considered in [42] and [43]. In [42], empirical information of computation energy is used to propose a model for energy consumption in CRANs. Based on this model, power and bandwidth are allocated to all users to meet their QoS. Then, the number of active processing units in the BBU pool is optimized to minimize energy consumption. In [43], the authors present a unified framework to improve performance by using joint CRAN resource scheduling and allocation of computational resources. To achieve this, the authors formulate and solve a stochastic problem for resource scheduling and handling variable length requests.

Deadline requirement is studied for data transmission over CRANs in [44] and [45]. In [44], a complexity-efficient resource scheduler is explored to increase the CRAN throughput and to meet the deadline requirements for sub-frames of RRHs. In [45], power-efficient RRH allocation and processor scheduling are studied for CRANs in which the minimum SINR, different transmission times and deadlines required for each of users are guaranteed by the scheduling algorithm. Beside these works on CRANs, the earliest deadline first scheduling [46] is the basis for many other scheduling algorithms developed in recent papers [47], [48]. In [49], the authors implement a small-cell scheduler for a CRAN to improve its capacity, energy and spectral efficiency in a 3D indoor environments.

While joint scheduling and resource allocation have been considered in the aforementioned references, none of them

TABLE 1. Topics of related works.

| Topic | Related works |
|----------------------------|---|
| User scheduling/clustering | [20]–[25], [27], [29], [35], [36], [40], [41] |
| Power allocation | [21], [22], [26], [34], [36], [40], [42] |
| BBU allocation | [26], [28], [42], [43] |
| Data transmission time | [26]–[28] |
| Spectrum allocation | [19], [27], [30], [41], [42] |
| RRH scheduling/clustering | [28], [30]–[36] |
| Throughput | [19], [29], [33], [34], [49] |
| Up and down link | [33]–[36] |
| Deadline based scheduling | [44]–[48] |
| Big data transmission | [6], [50]–[56] |

pays attention to efficient communication of big data. Next, we will provide a brief review of references that have focused on big data requests.

2) BIG DATA TRANSMISSION

The general interplay between big data and communication networks, as well as some open problems in the field are elaborated in [6], [50]. The work in [51] utilizes channel state information statistics to obtain the distribution of energy consumed in big data transmissions. In [52], big data transmission requirements in the context of internet of things (IoT) are investigated in terms of expected delay, data length, link's capacity and load. The authors introduce a scheduling and routing algorithm to provide a lower transmission time for big data and to improve the end user's experience. In [53], a scheduling policy is developed for real time video delivery as a specific case of big data transmission. However, their strategy is simply to selectively transmit those videos with fewer packets. Many to many wireless big data delivery is addressed in [54] by construction of a group communication structure. The authors of [55] consider security-aware resource allocation, e.g., processing resources needed to perform encryption for mobile social big data.

Most works mentioned above do not pose a specific optimization problem and just provide heuristic algorithms that can handle big data. On the other hand, the optimization problems formulated in [53] and [56] strive to maximize the number of accepted requests, which inherently favor smaller-sized data requests. Furthermore, these two references solve the formulated optimization problems using heuristic sub-optimal algorithms. In contrast, we introduce in this paper a new objective function that can ensure a certain degree of fairness while favoring big data transmissions. We also propose an algorithm that can find the globally-optimal solution for the formulated joint scheduling and resource allocation problem.

B. CONTRIBUTIONS

In this paper, we consider a cognitive CRAN serving both PUs and SUs. Our first objective is to develop an efficient joint resource allocation algorithm in time, frequency (spectrum) and spatial dimensions for SUs. Second, we propose a mechanism to favor "big data" users as much as possible

(which is determined by the network planner), while maintaining a certain degree of fairness in the network. To the best of our knowledge, there are no published works that consider these two aspects simultaneously. In this context, the contributions of our work are as follows:

- We introduce a new objective function that ensures certain degree of fairness among SUs while favoring big data requests. The objective parameters can be selected to trade-off between these two requirements. The allocated resources are time, spectrum and RRHs to connect to.
- Since the computational complexity of both exhaustive search and the branch-and-bound method is prohibitive, we propose a novel dynamic programming (DP) approach to reduce the complexity as much as possible while reaching the global optimum.
- We establish a connection between the formulated resource allocation problem and the problem of finding colorful independent sets on a graph [57]. As a result, we show how to apply advanced methods (that exist for finding colorful independent sets problem) to solve our resource allocation problem with even lower complexity than the DP approach.
- Performance of the proposed DP resource allocation is numerically compared to that of two well-known heuristic resource scheduling algorithms and the superiority of the proposed method is thoroughly demonstrated under various criteria.

C. PAPER ORGANIZATION AND NOTATIONS

The remainder of this paper is organized as follows. Section II introduces the system model under consideration and describes various parameters. Section III formulates the optimization problem of interest. Section 4 establishes the connection between our resource allocation/scheduling problem and the problem of finding colorful independent sets on a graph. Furthermore, our proposed DP approach is also developed in this section. The computational complexity of exhaustive search and that of the proposed algorithm are analyzed in Section V. Simulation results are presented in Section VI, and the conclusions are drawn in Section VII.

We use calligraphic letters to represent various sets. Bold capital (non-capital) letters are used to denote matrices (vectors). Scalars are represented by non-bold and non-calligraphic letters.

II. SYSTEM MODEL

Fig. 1 illustrates the proposed CRAN system model, which is composed of a macro cell overlaid with several small cells to serve PUs and a set $U := \{1, 2, \dots, U\}$ of U randomly distributed SUs, respectively. We focus on data transmission to SUs in the downlink mode. Time is divided into equal length time slots of duration Δt , whose value generally depends on subcarrier spacing [59]. For instance, in the IEEE 802.11 family standard $\Delta t = 9 \mu s$ [58], whereas the values

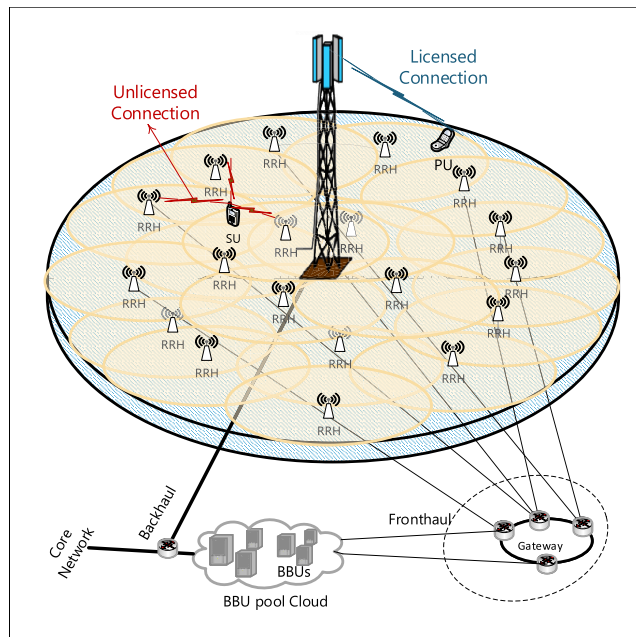


FIGURE 1. Proposed system model.

for Δt in 5G are recommended in [59]. Time slot t is the slot whose time interval belongs to $[(t - 1)\Delta t, t\Delta t)$. Both PUs and SUs utilize the same time slots structure and are thus mutually synchronized.

Small cells include a set of R RRHs (e.g., micro cell and pico cell RRHs), which are distributed in the service area at fixed locations. Each RRH can simultaneously serve at most u^{\max} SUs. RRHs are connected to the BBU pool via high speed and low latency fronthaul links [17]. All RRHs and SUs have single antennas. Extension to the cases where each RRH and/or SUs have multiple antennas shall be explored in future research. We assume that the BBU pool has perfect knowledge of large-scale and shadow fading parameters for all the RRH-SU links. However, it only knows the statistics of small-scale fading. In order to decode transmitted messages accurately, RRHs should have perfect knowledge of the channels connecting those SUs that are assigned to them. While accurate small-scale CSI values can be obtained via training at the beginning of each coherence time, it is out of the scope of this work as we only exploit small-scale fading statistics in our design.

The vacant spectrum bands for SUs are arranged in the spectrum pool [60] and this spectrum pool is divided into units of equal bandwidth Δf . Although techniques such as NOMA and/or beamforming may be used to simultaneously assign a vacant spectrum unit to more than one SU, in this work we focus on orthogonal association problem, where only one SU or PU can access a particular spectrum unit during each time slot. Each SU can use at most s^{\max} spectrum units.

The (possibly multiple) service provider(s) collects all the data requests that are sent by the SUs through RRHs, and

then forwards these requests to the BBU pool for resource allocation and scheduling. Our design aims to determine which spectrum units and RRHs, and at what times should be assigned to each SU. The objective is to satisfy QoS requirements for as many SUs as possible, while providing relative priority for big data transmission requests. Our design performs batch scheduling, which means that all requests over a certain time epoch are first collected and then are jointly scheduled.

A. RESOURCE ALLOCATION AND SCHEDULING PARAMETERS

Resource allocation time epoch or T is the number of time slots that transmission requests from SUs are jointly served. During each resource allocation time epoch, the SUs who want to receive data, send their requests to the service provider via a separate control channel. Upon joint resource allocation by the BBU pool, these SUs will be served in the next resource allocation epoch. For simplicity we only consider one such epoch, i.e., a T time slots epoch as the proposed procedure can be repeated for future resource allocation time epochs. Let $\mathcal{S} := \{1, 2, \dots, S\}$ denote the set of all spectrum units and let \mathcal{S}^t denote its subset of vacant spectrum units at time slot t . Since PUs' spectrum activities are time dependent, \mathcal{S}^t is a time-dependent set.

The SUs request various types of data. Thus, we set $L_u \times L$ to be the size of data requested by SU u , where L_u is a positive integer that represents the number of data frames, and L is the standard frame size (e.g., about 1500 bytes for Ethernet II and IEEE 802.3, or 2304 bytes for WLAN, and could be larger for extended versions [58]). Some previous studies consider that service providers rent some resources from the CRAN entity [19]. Then, service providers decide by themselves the resources required by each of the SUs. Thus, a service provider should collect information about the size and QoS of the data demanded by each of the SUs, as well as CSI of the rented resources in order to perform resource allocation and scheduling for SUs. However, in practice, a service provider has no access to full information of both the CRAN and characteristics of the data (size, type, etc.) requested by the SUs subscribing to other service providers. This may lead to an unfair and sub-optimal resource allocation in the network, especially when users are distributed non-uniformly in a service area [29]. In this work, by exploiting advantages of cloud processing of the BBU pool, we assign the duty of resource allocation and scheduling to the BBU pool. To this end, service providers communicate to the BBU pool the data requested by SU u , which is the quadruplet $\langle L_u \times L, T_u^s, T_u^w, \gamma_u \rangle$. In this quadruplet, T_u^s (in terms of the number of time slots) denotes the earliest time that user u can begin receiving service. T_u^s is usually set to zero as users can begin receiving service right upon request. However, it may be non-zero when user u has a slow processor and a full buffer from past data it has received. Similarly, T_u^w denotes the maximum satisfactory waiting time (in terms of the number of time slots) before beginning to receive data by the u th SU,

and γ_u is the minimum SNR required by SU u . The service provider determines the pair of $\langle T_u^w, \gamma_u \rangle$ based on the type of the data requested by SU u .

The BBU pool receives the quadruplet $\langle L_u \times L, T_u^s, T_u^w, \gamma_u \rangle$ for all SUs and then performs joint scheduling using this information. Before proceeding to the next section, some definitions are given next.

- $\gamma_{r,u}$: This is SNR seen by the receiver of SU u when associated with RRH r . The data rate [bits/sec] of the wireless communication link between SU u and the associated RRH r depends on $\gamma_{r,u}$. We denote $h_{r,u} \in \mathbb{C}$ as the downlink channel coefficient of this link, which includes the effect of RRH transmitter's antenna gain, SU receiver's antenna gain, small scale fading, and large scale fading, and shadowing. Therefore, $\gamma_{r,u}$ is given by

$$\gamma_{r,u} = \frac{|h_{r,u}|^2 P_{r,u}}{\Gamma \sigma^2 \Delta f}, \tag{1}$$

where $P_{r,u}$ is the transmit power of RRH r to SU u , σ^2 denotes the power spectral density of background noise, and Γ is the SNR gap, which represents the mismatch between theoretical capacity at a specified SNR and its actual throughput that may be achieved in practice. It should be pointed out that (1) is independent of spectrum band utilized or $s \in \mathcal{S}$. This is because we assume a narrowband system, i.e., a system whose total bandwidth is smaller than the coherence bandwidth of the channel. Therefore, the system experiences frequency-flat fading and all spectrum bands observe the same channel gain.¹ Since the BBU pool has access only to the statistics of small-scale fading, the ergodic rate is considered [28], [61].

- $\tau_u(\mathcal{R}_u, \mathcal{S}_u)$: This is the data transmission time duration for SU u . Let us define $\mathcal{R}_u := \{\mathcal{R}_u^1, \dots, \mathcal{R}_u^t, \dots, \mathcal{R}_u^T\} \subseteq \mathcal{R}^T$, and $\mathcal{S}_u := \{\mathcal{S}_u^1, \dots, \mathcal{S}_u^t, \dots, \mathcal{S}_u^T\} \subseteq \mathcal{S}^T$, where \mathcal{R}_u^t and \mathcal{S}_u^t , respectively, represent the set of RRHs and spectrum units assigned to user u at time slot t , whereas $\mathcal{R}^T, \mathcal{S}^T$ represent a T times Cartesian products of the sets \mathcal{R}, \mathcal{S} , respectively. Thus, \mathcal{R}_u and \mathcal{S}_u denote the sets of allocated RRHs and spectrum units to SU u over the whole resource allocation epoch T . Assuming the maximum ratio transmission (MRT) of distributed downlink beamforming to every user, the data transmission time for data request by SU u is given by

$$\tau_u(\mathcal{R}_u, \mathcal{S}_u) = \left\lceil \frac{L_u L}{\Delta t \Delta f \mathbb{E}_h \left[\sum_{t \in T} |\mathcal{S}_u^t| \log_2 \left(1 + \sum_{r \in \mathcal{R}_u^t} \gamma_{r,u} \right) \right]} \right\rceil. \tag{2}$$

Note that, to obtain the transmission time as an integer multiple of Δt , we have used the ceiling function, $\lceil x \rceil$,

¹Extension to wideband systems and frequency-selective fading is an interesting research direction.

which maps x to the smallest integer greater than or equal to x .

III. PROBLEM FORMULATION

Before describing the objective function, we explain the temporal and resource constraints that should be satisfied by any feasible solution.

A. SCHEDULING AND RESOURCE ALLOCATION CONSTRAINTS

A service provider registers data requests of those SUs that arrive in the previous scheduling epoch T in the BBU pool to be served in the next resource allocation epoch of length T . By scheduling, the BBU pool first decides about acceptance/rejection of this request. If it is accepted, SU u may experience t_u^w time slots as waiting time to start receiving data. For successful data transmission, t_u^w should be in the range of $[T_u^s, T_u^w]$. When requests of some SUs are rejected, these SUs repeat their requests in the next scheduling epoch. Below we provide definitions that are relevant to the problem formulation.

Definition 1: Each RRH R_i maintains a service capacity of at most u^{\max} users in every time slot. To model this constraint, we define u^{\max} fictitious RRH units per every real RRH and represent them by $\tilde{R}_{i,j}$. Here, i denotes the real RRH to which $\tilde{R}_{i,j}$ refers and $j = 1, 2, \dots, u^{\max}$ denotes the fictitious RRH unit index. Every RRH unit may be assigned to only one SU in each time slot. The set of all RRH units in one epoch is denoted by $\tilde{\mathcal{R}} := \{\tilde{R}_{i,j}^t : i = 1, 2, \dots, R, j = 1, 2, \dots, u^{\max}, t = 1, 2, \dots, T\}$. Furthermore, the set of RRH units assigned to user u in time slot t is represented by $\tilde{\mathcal{R}}_u^t$. Finally, the set of all RRH units assigned to user u is given by $\tilde{\mathcal{R}}_u := \{\tilde{\mathcal{R}}_u^1, \tilde{\mathcal{R}}_u^2, \dots, \tilde{\mathcal{R}}_u^T\}$.

Definition 2: Let $I_u(t_u^w, \tilde{\mathcal{R}}_u, S_u)$ denote the transmission time interval for data requested by SU u with waiting time t_u^w and $(\tilde{\mathcal{R}}_u, S_u)$ as assigned resources. It is given as

$$I_u(t_u^w, \tilde{\mathcal{R}}_u, S_u) = t_u^w + [0, \tau_u(\tilde{\mathcal{R}}_u, S_u)], \quad \forall u \in \mathcal{U}. \quad (3)$$

Here, the scalar t_u^w is added to every member of the set $[0, \tau_u(\tilde{\mathcal{R}}_u, S_u)]$.

Definition 3: The set of all possible data transmission time intervals and resources for SU u is expressed by

$$\mathcal{I}_u = \left\{ (I_u, t_u^w, \tilde{\mathcal{R}}_u, S_u) \mid t_u^w \in [T_u^s, T_u^w], \tilde{\mathcal{R}}_u \subseteq \tilde{\mathcal{R}}, S_u \subseteq S^T \right\}, \quad \forall u \in \mathcal{U}. \quad (4)$$

Definition 4: The set of all possible transmission time intervals and resources for the data requests of \mathcal{U} is $\bigcup_{u \in \mathcal{U}} \mathcal{I}_u$, which is the union of all \mathcal{I}_u s.

Definition 5: Any set \mathcal{I} which is a subset of all possible transmission time intervals and resources, i.e., $\mathcal{I} \subseteq \bigcup_{u \in \mathcal{U}} \mathcal{I}_u$, is a disjunctive set, if it meets all the following six resource and temporal constraints.

- 1) Each spectrum unit and RRH unit are assigned to only a single SU in each time slot:

$$S_u^t \cap S_{u'}^t = \emptyset, \quad \forall t \in \{1, 2, \dots, T\}, \forall u \neq u', \quad (5a)$$

$$\tilde{\mathcal{R}}_u^t \cap \tilde{\mathcal{R}}_{u'}^t = \emptyset, \quad \forall t \in \{1, 2, \dots, T\}, \forall u \neq u'. \quad (5b)$$

- 2) The total assigned spectrum units to SUs are no larger than the spectrum pool capacity in each time slot:

$$\bigcup_{u \in \mathcal{U}} S_u^t \subseteq S^t, \quad \forall t \in \{1, 2, \dots, T\}, \quad (6)$$

- 3) The total number of assigned spectrum units for each SU is limited by s^{\max} :

$$|S_u^t| \leq s^{\max}, \quad \forall (t, u) \in (\{1, 2, \dots, T\}, \mathcal{U}). \quad (7)$$

Note that, in general, s^{\max} could be different for various requested data types.

- 4) In RRH assignment, minimum required SNR is enforced:

$$\mathbb{E}_h[\gamma_{r,u}] \geq \gamma_u, \quad \forall (r, u) : r \in \tilde{\mathcal{R}}_u^t. \quad (8)$$

- 5) Every user is allocated either none or exactly one fictitious RRH unit of a real RRH:

$$\sum_{j=1}^{u^{\max}} \text{Indicator}_{(\tilde{R}_{i,j}^t \in \tilde{\mathcal{R}}_u^t)} \leq 1, \quad \forall t, \forall i = 1, \dots, R. \quad (9)$$

where we have used an indicator function which assumes one if its logical argument is true and zero otherwise.

- 6) For all $(I_u, t_u^w, \tilde{\mathcal{R}}_u, S_u) \in \mathcal{I}$, we have:

$$T_u^s \leq t_u^w \leq T_u^w, \quad \max I_u \leq T. \quad (10)$$

If \mathcal{I} is a disjunctive set, we write it as $\mathcal{I} \subseteq_D \bigcup_{u \in \mathcal{U}} \mathcal{I}_u$.

Constants s^{\max} and u^{\max} could be determined in an adaptive fashion based on the number of SUs, CSI, distribution of the SUs, and requested data types. Other constraints can also be added. For example, in Section VI (performance evaluation), to avoid starvation of certain users in the CRAN that could happen when some SUs consume all resources, we force each SU to occupy a minimum number of spectrum units.

B. OPTIMUM SCHEDULING WITH BIG DATA PRIORITY

We define $\tau(\mathcal{I})$ as the maximum data transmission time among SUs that are accepted by \mathcal{I} . Mathematically, it is written as

$$\tau(\mathcal{I}) := \max_{u: (I_u, t_u^w, \tilde{\mathcal{R}}_u, S_u) \in \mathcal{I}} \tau_u(\tilde{\mathcal{R}}_u, S_u). \quad (11)$$

Note that the maximum operation in (11) is performed over accepted users only. At this point, the optimization problem in our proposed scheduling and resource allocation framework can be formally stated as follows:

$$\max_{\mathcal{I} \subseteq_D \bigcup_{u \in \mathcal{U}} \mathcal{I}_u} J(\mathcal{I}), \quad (12)$$

where maximization is carried out over all possible disjunctive sets. The objective function is defined as

$$J(\mathcal{I}) := \frac{1}{\tau(\mathcal{I})} \sum_{u:(I_u, t_u^w, \tilde{\mathcal{R}}_u, \mathcal{S}_u) \in \mathcal{I}} \alpha_u L_u L, \quad (13)$$

where α_u represents the weight or priority assigned to user u . As will be demonstrated later on, the proposed objective has the capability to simultaneously address both fairness and big data issues.

By placing $\tau(\mathcal{I})$ in the denominator of the objective function, we favor small $\tau(\mathcal{I})$ values. This enforces the maximum of SU's data transmission times to be as small as possible. Furthermore, a big volume of data leads to a greater data transmission time, hence $\tau(\mathcal{I})$ will be dominated by big data requests. By minimizing (11), the scheduling problem allocates more resource to big data requests. To further emphasize on big data requests, the overall volume of transmitted data is placed in the numerator of the objective function, and also, α_u can be adjusted properly to reflect priorities for certain users. For example, $\alpha_u = \frac{L_u}{\sum_{u \in \mathcal{U}} L_u}$ assigns priorities directly proportional to L_u , which gives a higher priority to the request with bigger data volume. The quantities α_u s affect fairness in the CRAN as well. We will study performance of the proposed algorithm for both $\alpha_u = 1$ and $\alpha_u = \frac{L_u}{\sum_{u \in \mathcal{U}} L_u}$.

The BBU pool should jointly perform the selection of SUs, resource allocation and time scheduling to optimize (12), which is accomplished by finding the optimal \mathcal{I} , denoted by \mathcal{I}^* . Unfortunately, the size of the search space to exhaustively find the global optimum \mathcal{I}^* is exponential in terms of the number of optimization variables. In the next section, we apply dynamic programming to reduce the computational burden as much as possible while guaranteeing to reach the globally optimum solution.

IV. DYNAMIC PROGRAMMING ALGORITHM

First, we describe a well-known problem which bears close resemblance to our optimization problem. Then, we point to the relevant references that have tackled this well-known problem. Building on this existing techniques, we develop a dynamic programming algorithm to solve our optimization problem. To clearly demonstrate how the proposed method works, a toy example is also presented.

A. SCHEDULING TASKS ON A SINGLE PROCESSOR

Suppose there are n tasks which can be processed by a single processor. Each task can be completed only at certain time intervals. These intervals are given as part of the problem formulation. If two intervals for different tasks overlap, only one can be completed. The problem is to schedule these tasks so that a maximum number of tasks can be completed [62]. If a single completion interval is given per task, this problem can be solved optimally via greedy scheduling. However, it is well-known that if more than two possible processing intervals are given per task, this problem is NP-hard [57]. It has been shown that finding the global optimum solution for this problem amounts to determining the maximum colorful

independent set on an interval graph. Below, we provide the definitions for these concepts.

- 1) *Interval graph*: A graph is an interval graph if its vertices represent time intervals. Furthermore, there exists an edge between two vertices if corresponding intervals overlap.
- 2) *Colorful set with respect to a function*: For a finite set A and a coloring function $\phi : A \rightarrow \{1, \dots, N\} \subseteq \mathbb{N}$, where \mathbb{N} is the set of natural numbers, $B \subseteq A$ is colorful with respect to ϕ if $\forall a, b \in B$, and $a \neq b$ we have $\phi(a) \neq \phi(b)$. Indeed, we can assume that the set $\{1, \dots, N\}$ represents N different colors. Then, we call B is colorful if no two of its members have the same color.
- 3) *Independent set of a graph*: It is a subset of graph vertices with the property that neither two of its members are connected by a direct edge.

Returning to our task scheduling problem, vertices are connected by a direct edge if they overlap in time. Hence, we can simultaneously schedule those vertices that are not connected by an edge. If we assign the same color to all those vertices that correspond to possible intervals for processing a single task, then we want our set of scheduled intervals to be colorful. This is because each task needs to be processed only once. Hence, the problem of scheduling a maximum number of tasks amounts to finding the maximum colorful independent set of the corresponding interval graph. Although this problem is NP-hard, various advanced methods have been proposed to minimize the complexity of solving this problem (see [57], [63] for example). Here, we apply a simple dynamic programming approach to solve our problem. Yet, the methods in these references can be leveraged to obtain even more complexity-efficient optimal algorithms.

B. CRAN SCHEDULING AS A COLORFUL INDEPENDENT SET PROBLEM

To solve (12), we first enumerate all possible disjunctive sets. Then, we evaluate each to obtain the one that maximizes our objective in (13). First, we define a generalization of the interval graph.

Our equivalent interval graph, G is defined by V as the set of vertices, and E as the set of edges. Every $(I_u, t_u^w, \tilde{\mathcal{R}}_u, \mathcal{S}_u)$ represents a vertex in the graph. For each $(I_u, t_u^w, \tilde{\mathcal{R}}_u, \mathcal{S}_u), (I_{u'}, t_{u'}^w, \tilde{\mathcal{R}}_{u'}, \mathcal{S}_{u'}) \in V$, they are connected by an edge if they cannot be scheduled simultaneously. This amounts to an overlap in the time domain while simultaneously having an overlap either in frequency or RRH unit resources. Mathematically, E is defined by

$$E = \left\{ (I_u, t_u^w, \tilde{\mathcal{R}}_u, \mathcal{S}_u)(I_{u'}, t_{u'}^w, \tilde{\mathcal{R}}_{u'}, \mathcal{S}_{u'}) \mid (I_u \cap I_{u'} \neq \emptyset) \text{ and } \left[(\mathcal{S}_u^f \cap \mathcal{S}_{u'}^f \neq \emptyset) \text{ or } (\tilde{\mathcal{R}}_u \cap \tilde{\mathcal{R}}_{u'} \neq \emptyset) \right] \right\}. \quad (14)$$

Also, we define a coloring function $\phi : V \rightarrow \mathcal{U}$ such that $\phi((I_u, t_u^w, \tilde{\mathcal{R}}_u, \mathcal{S}_u)) = u$ where every user is represented by a different color. So, any disjunctive set \mathcal{I} corresponds

to an independent colorful subset of vertices in graph G which can be simultaneously scheduled. It is noticeable that constraint (5) is taken into account when an edge exists between two vertices, and constraints (6)-(8) are considered in definition of vertices. Contrary to the processor scheduling problem that targets the maximum colorful independent set, our objective function in (13) is not necessarily maximized by the maximum colorful independent set. Hence, we have to enumerate over all colorful independent sets of various sizes and evaluate them one by one to find the maximum. This task is carried out via dynamic programming.

C. PROPOSED DP APPROACH

Let us define \mathcal{A}_v^t as the collection of all colorful independent sets of size v whose time intervals end at or before t . Note that v represents the number of users that are scheduled, and hence it is a non-negative integer that assumes values between 0 and U . The variable t is also a non-negative integer taking on a value between 0 and T . Based on the principle of optimality, the independent colorful sets of size v may be calculated from the independent colorful sets with smaller sizes or with a shorter end time. Mathematically, this relationship is captured in the following theorem.

Theorem 1: Initially, set $\mathcal{A}_0^t = \emptyset$ and $\mathcal{A}_v^0 = \emptyset$. Then, \mathcal{A}_v^t can be determined by the following iteration:

$$\mathcal{A}_v^t = \mathcal{A}_v^{t-1} \cup \left\{ \bigcup_{(I_u, t_u^w, \tilde{\mathcal{R}}_u, S_u): \max I_u = t} \{ (I_u, t_u^w, \tilde{\mathcal{R}}_u, S_u), \mathcal{A}_{v-1}^t \} \right\}_D \quad (15)$$

where the subscript D ensures that only disjunctive augmentations are accepted i.e., colorfulness and independence are maintained.

Proof: To prove this theorem, we consider the following cases separately:

- 1) The first case is when our generalized interval graph contains no vertex whose time interval ends in time slot t . It is obvious that $\mathcal{A}_v^t = \mathcal{A}_v^{t-1}$.
- 2) For the case that there exists a vertex whose time interval ends at t , it can be added to any vertex in \mathcal{A}_{v-1}^t as long as the augmented set is still colorful and independent. These newly formed feasible schedules should be added to those in \mathcal{A}_v^{t-1} . □

Our proposed DP method in Theorem 1 is succinctly summarized as Algorithm 1. After all colorful independent sets are collected via Algorithm 1, Algorithm 2 evaluates the objective over all these sets and selects the one yielding the maximum value as the global optimum \mathcal{I}^* . To further reduce complexity, the update in (15) needs not to be computed at every time slot but only at those time slots when a user's service is completed. With this approach, the computational complexity is reduced, especially in case of having big data requests that span over many time slots. The computational

Algorithm 1: Finding Colorful Independent Sets

Input: $\{ < L_u \times L, T_u^s, T_u^w, \gamma_u > \forall u \in \mathcal{U} \}$, S^T, \mathcal{R}^T , and statistical CSI for RRHs-SUs.
Output: $\mathcal{A}_v^T, \forall v \leq U$: Independent colorful sets.

- 1 $\forall (r, u) \in \mathcal{R} \times \mathcal{U}$ calculate $\mathbb{E}[\gamma_{r,u}]$
- 2 Exclude (r, u) if $\mathbb{E}[\gamma_{r,u}] < \gamma_u$
- 3 $\forall u \in \mathcal{U}$ form all possible $\tau_u(\tilde{\mathcal{R}}_u, S_u)$ such that $|S_u^t| \leq s^{\max}$
- 4 $\forall u \in \mathcal{U}$ form \mathcal{I}_u
- 5 $\forall t \in [0, T]$: $\mathcal{A}_0^t \leftarrow \emptyset$
- 6 $\forall v \leq | \mathcal{U} |$: $\mathcal{A}_v^0 \leftarrow \emptyset$
- 7 **for** $v \leftarrow 1 : U$ **do**
- 8 **for** $t \leftarrow 1 : T$ **do**
- 9 $\mathcal{A}_v^t \leftarrow \mathcal{A}_v^{t-1}$
- 10 **foreach** $\mathcal{I}_{cand} = (I_u, t_u^w, \tilde{\mathcal{R}}_u, S_u) \in \bigcup_{u \in \mathcal{U}} \mathcal{I}_u$ **that** $\max I_u = t$ **do**
- 11 **foreach** $\mathcal{I} \in \mathcal{A}_{v-1}^t$ **do**
- 12 **if** $\mathcal{I}_{cand} \neq \mathcal{I}$ **then**
- 13 **if** $\{ \mathcal{I}_{cand}, \mathcal{I} \}$ **is disjunctive then**
- 14 $\mathcal{A}_v^t \leftarrow \mathcal{A}_v^t \cup \{ \mathcal{I}_{cand}, \mathcal{I} \}$
- 15 **return** $\forall v : \mathcal{A}_v^T$

Algorithm 2: Finding the Optimal Solution

Input: $\forall v \leq U : \mathcal{A}_v^T, \{ \alpha_u, L_u \times L \mid u \in \mathcal{U} \}$
Output: \mathcal{I}^*

- 1 $Temp \leftarrow -\infty$
- 2 **foreach** $\mathcal{I} \in \mathcal{A}_v^T, \forall v = 1, \dots, U$ **do**
- 3 $\tau(\mathcal{I}) \leftarrow \max_{u: (I_u, t_u^w, \tilde{\mathcal{R}}_u, S_u) \in \mathcal{I}} \tau_u(\mathcal{R}_u, S_u)$
- 4 Calculate $J(\mathcal{I})$
- 5 **if** $J(\mathcal{I}) > Temp$ **then**
- 6 $\mathcal{I}^* \leftarrow \mathcal{I}, Temp \leftarrow J(\mathcal{I})$
- 7 **return** \mathcal{I}^*

complexities of Algorithm 1 and Algorithm 2 are analyzed in the next section. Before closing this section, we present a toy example to illustrate how Algorithm 1 works.

D. A SIMPLE EXAMPLE

Consider a simple CRAN where the values for U, S, R, s^{\max} , and u^{\max} are all set to two. The proposed scheduling interval begins at time slot one and ends at $T = 22$. Also, $T_1^s = 0, T_1^w = 1$ and $T_2^s = T_2^w = 2$. The spectrum availability for SUs over time S^t s is given on top of Table 2. We assume that the SNR constraint (8) is always met for any user-RRH pair. Since $u^{\max} = 2$ and there are exactly two users, the RRH service limit is always satisfied. Thus, it is optimal for both users to connect to both RRHs over the service time. Table 2 also shows all possible data transmission time

TABLE 2. Available spectrum units and all possible schedules for the two users in the toy example.

| | | | | | | | | | | | | | | | | | | | | | | | | |
|-----------------|-------|---|---|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|----|
| \mathcal{S}^t | t | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 |
| | s = 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 |
| | s = 2 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 0 | 0 |

$$\textcircled{1} (I_1, t_1^w, \mathcal{R}_1, \mathcal{S}_1) \rightarrow I_1 = [1, 9] \quad t_1^w = 1$$

| | | | | | | | | | | | | | | | | | | | | | | |
|-------------------|-------|---|---|---|---|---|---|---|---|---|-------------------|-------|---|---|---|---|---|---|---|---|---|---|
| \mathcal{S}_1^t | t | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | \mathcal{R}_1^t | t | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | |
| | s = 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | r = 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | s = 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | r = 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

$$\textcircled{2} (I_1, t_1^w, \mathcal{R}_1, \mathcal{S}_1) \rightarrow I_1 = [2, 10] \quad t_1^w = 2$$

| | | | | | | | | | | | | | | | | | | | | | |
|-------------------|-------|---|---|---|---|---|---|---|---|----|-------------------|-------|---|---|---|---|---|---|---|---|----|
| \mathcal{S}_1^t | t | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | \mathcal{R}_1^t | t | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| | s = 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | r = 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | s = 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | r = 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

$$\textcircled{3} (I_1, t_1^w, \mathcal{R}_1, \mathcal{S}_1) \rightarrow I_1 = [1, 9] \quad t_1^w = 1$$

| | | | | | | | | | | | | | | | | | | | | | |
|-------------------|-------|---|---|---|---|---|---|---|---|---|-------------------|-------|---|---|---|---|---|---|---|---|---|
| \mathcal{S}_1^t | t | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | \mathcal{R}_1^t | t | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| | s = 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | r = 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | s = 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | r = 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

$$\textcircled{4} (I_1, t_1^w, \mathcal{R}_1, \mathcal{S}_1) \rightarrow I_1 = [2, 10] \quad t_1^w = 2$$

| | | | | | | | | | | | | | | | | | | | | | |
|-------------------|-------|---|---|---|---|---|---|---|---|----|-------------------|-------|---|---|---|---|---|---|---|---|----|
| \mathcal{S}_1^t | t | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | \mathcal{R}_1^t | t | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| | s = 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | | r = 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | s = 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | r = 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

$$\textcircled{5} (I_1, t_1^w, \mathcal{R}_1, \mathcal{S}_1) \rightarrow I_1 = [1, 5] \quad t_1^w = 1$$

| | | | | | | | | | | | | | | | | | | | | | |
|-------------------|-------|---|---|---|---|---|---|---|---|---|-------------------|-------|---|---|---|---|---|---|---|---|---|
| \mathcal{S}_1^t | t | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | \mathcal{R}_1^t | t | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| | s = 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | | r = 1 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |
| | s = 2 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | | r = 2 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 |

$$\textcircled{6} (I_1, t_1^w, \mathcal{R}_1, \mathcal{S}_1) \rightarrow I_1 = [2, 6] \quad t_1^w = 2$$

| | | | | | | | | | | | | | | | | | | | | | |
|-------------------|-------|---|---|---|---|---|---|---|---|---|-------------------|-------|---|---|---|---|---|---|---|---|---|
| \mathcal{S}_1^t | t | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | \mathcal{R}_1^t | t | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| | s = 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | | r = 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |
| | s = 2 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | | r = 2 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 |

$$\textcircled{7} (I_2, t_2^w, \mathcal{R}_2, \mathcal{S}_2) \rightarrow I_2 = [2, 21] \quad t_2^w = 2$$

| | | | | | | | | | | | | | | | | | | | | | |
|-------------------|-------|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|
| \mathcal{S}_2^t | t | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 |
| | s = 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | s = 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| \mathcal{R}_2^t | r = 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | r = 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

$$\textcircled{8} (I_2, t_2^w, \mathcal{R}_2, \mathcal{S}_2) \rightarrow I_2 = [2, 18] \quad t_2^w = 2$$

| | | | | | | | | | | | | | | | | | | | | | |
|-------------------|-------|---|---|---|---|---|---|---|---|----|----|----|----|----|----|----|----|----|----|----|----|
| \mathcal{S}_2^t | t | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 |
| | s = 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | s = 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 |
| \mathcal{R}_2^t | r = 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 |
| | r = 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 0 | 0 |

$$\textcircled{9} (I_2, t_2^w, \mathcal{R}_2, \mathcal{S}_2) \rightarrow I_2 = [2, 9] \quad t_2^w = 2$$

| | | | | | | | | | | | | | | | | | | | | | |
|-------------------|-------|---|---|---|---|---|---|---|---|---|-------------------|-------|---|---|---|---|---|---|---|---|---|
| \mathcal{S}_2^t | t | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | \mathcal{R}_2^t | t | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| | s = 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | r = 1 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| | s = 2 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | | r = 2 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

intervals and resources for these two SUs where $|\mathcal{I}_1| = 6$, and $|\mathcal{I}_2| = 3$. We use blue and red colors for SU 1 and SU 2, respectively. The corresponding colorful interval graph is presented in Fig. 2. In this graph, each vertex corresponds to one $(I_u, t_u^w, \mathcal{R}_u, \mathcal{S}_u)$ and is labeled by a number. As explained before, we only need to update the set \mathcal{A}_v^t at those time slots where a schedule ends. Therefore, we assume 6 update

slots, i.e., 5, 6, 9, 10, 18, and 21 instead of all the 22 time slots. Table 3 illustrates \mathcal{A}_v^t for $v = 0, 1, 2$ and update slot $= 0, 1, \dots, 6$. For example, \mathcal{A}_2^5 , which is the (2,5)th entry in the table, is formed according to following steps:

- 1- The contents of \mathcal{A}_2^4 in entry (2,4) are copied in \mathcal{A}_2^5 .
- 2- Is there $(I_u, t_u^w, \mathcal{R}_u, \mathcal{S}_u)$ which is completed in the 5th update slot corresponding to the 18th time slot? The

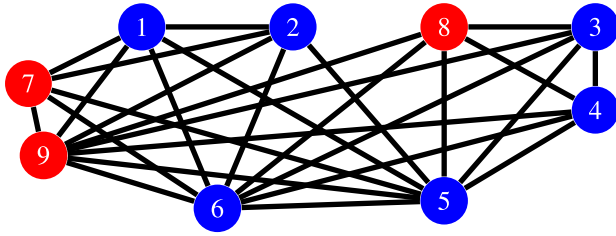


FIGURE 2. Colorful interval graph for our toy example. Each vertex corresponds to an schedule presented in Table 2.

answer is positive: 8 ends in the 5th update slot. Thus, 8 is combined with contents of \mathcal{A}_1^5 , i.e., 1, 2, 3, 4, 5, 6, 8, and 9. However, constraints on spectrum units and color gives combination permission only for 1 and 2. Finally, the content of \mathcal{A}_2^5 becomes {18} and {28}.

Similarly, \mathcal{A}_2^6 is formed according to the following steps:

- 1- The contents of \mathcal{A}_2^5 , which are {18} and {28}, are copied in \mathcal{A}_2^6 .
- 2- Is there $(I_u, t_u^w, \mathcal{R}_u, S_u)$ which is concluded in the 6th update slot corresponding to time slot 21? The answer is positive: 7 ends in the 6th update slot. So, 7 is combined with contents of \mathcal{A}_1^6 , i.e., 1, 2, 3, 4, 5, 6, 7, 8, and 9. However, constraints on spectrum units and color gives combination permission only for 3 and 4. So, the content of \mathcal{A}_2^6 is {18}, {28}, {37}, and {47}.

V. COMPLEXITY ANALYSIS

The number of possible intervals for SU u , $|\mathcal{I}_u|$, equals to $N(t_u^w) \times N(\mathcal{R}_u) \times N(S_u)$. $N(t_u^w)$ is equal to $T_u^w - T_u^s + 1$, and $N(\mathcal{R}_u)$ and $N(S_u)$ are equal to $\sum_{r=1}^R \binom{R}{r}$, and $\sum_{s=1}^{s^{\max}} \binom{|S|}{s} \leq (|S| + 1)^{s^{\max}}$, respectively. Therefore, the computational complexity of the exhaustive search is $\prod_{u \in \mathcal{U}} |\mathcal{I}_u|$.

In Algorithm 1, the computational complexity of lines 1 and 2 is $\mathcal{O}(R \times U)$. Also, the complexities for lines 3 and 4 are $\mathcal{O}\left(U \times (|S| + 1)^{s^{\max}} \times \sum_{r=1}^R \binom{Ru^{\max}}{r}\right)$ and $\mathcal{O}\left(U \times (|S| + 1)^{s^{\max}} \times \sum_{r=1}^R \binom{Ru^{\max}}{r} \times \max(T_u^w - T_u^s + 1)\right)$, respectively. The complexity of lines 7-14 is $\mathcal{O}\left(U^2 \times T \times (|S| + 1)^{s^{\max}} \times \sum_{r=1}^R \binom{Ru^{\max}}{r} \times \max(T_u^w - T_u^s + 1)\right)$. Overall, the complexity of Algorithm 1 is dominated by the latest term. Algorithm 2 finds optimal scheduling between independent colorful sets that are calculated in Algorithm 1, so its complexity is dominated by that of Algorithm 1. It can be observed that the proposed DP complexity is of order $\mathcal{O}(U^2 \times T \times \max(T_u^w - T_u^s + 1))$ times the complexity of forming all the possible candidates for a single u given by $|\mathcal{I}_u|$ if we assume u^{\max} is close to one. However, the complexity of exhaustive search equals to the product of complexities of forming schedules for a single user, or $|\mathcal{I}_u|$ which is larger by many orders. It is noteworthy that the problem of interest is NP-hard so finding the global optimum solution entails exponential complexity. However, the pro-

posed DP significantly reduces the exponential complexity order. As one of the main contributions of this work, reaching the global optimum is tractable over larger values of system parameters than what could have been achieved by exhaustive search. It is pointed out, however that, algorithms that are even more complexity-efficient could be found by leveraging prior works, such as [57] and [63].

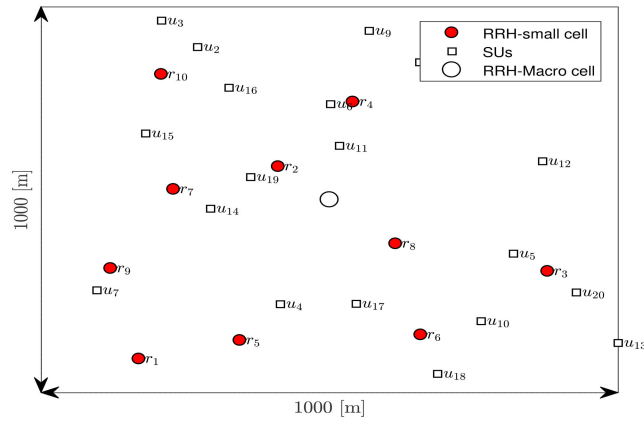
VI. SIMULATION RESULTS

The performance metrics for scheduling in CRAN include: (i) total data rate, (ii) maximum data transmission time, (iii) average waiting times, (iv) percentage of the scheduled SUs, and most importantly, (v) percentage of the scheduled SUs which request big data. The normalized distance between the minimum and maximum requested data sizes is divided into 5 ranges, where each range is recognized by ζ as upper bound of that range. We assume the 5th range that is recognized by $\zeta = 1$ and contains 20% of the largest requested data sizes, represents big data users. Raj Jain fairness index [65] is used to measure fair resource allocation for the scheduled SUs of each range.

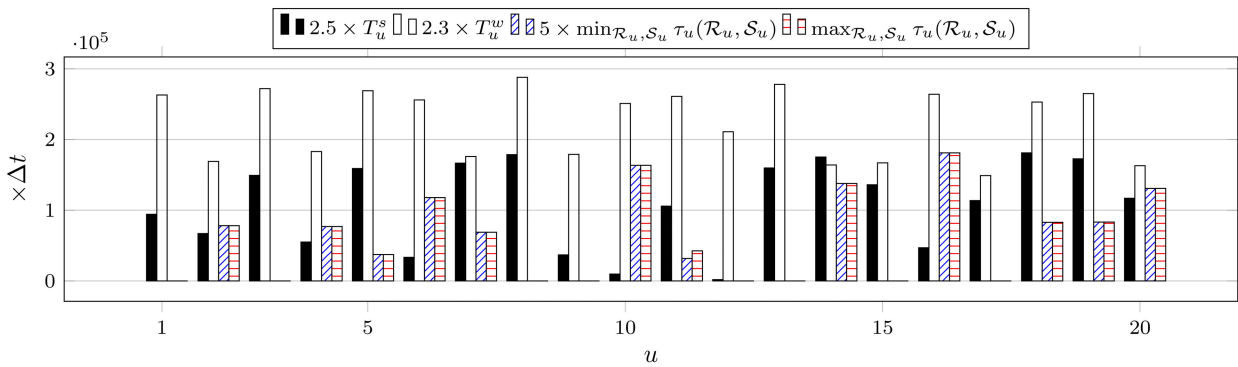
While the branch-and-bound method will reach the globally optimal solution, it does so at an exorbitant cost in computational complexity. Due to its impracticality, we did not compare the proposed algorithm with the branch-and-bound method. Instead, we considered two baseline algorithms for the cloud BBU based scheduling. The first one is referred to as *modified earliest deadline first scheduling* (MEDFS) and the second algorithm is referred to by *temporal resource demand-capacity ratio scheduling* (TRDCS). In MEDFS, we first allocate the RRH-spectrum resources to the SU with the earliest deadline [47], [48], [66]. In fact, nearest RRHs that meet the SNR constraint, and maximum possible spectrum units of s^{\max} are allocated to that SU. Afterwards, we remove all other intervals of this SU and reject those SUs whose QoS requirements cannot be satisfied by the remaining CRAN resources. Then, we repeat this process until all SUs are exhausted. When only one schedule per user is allowed, this algorithm will reach the global optimum. When more than two schedules per user exist, this algorithm will accept more than half of the number of scheduled users at the global optimum solution [67]. It is noteworthy that MEDFS strives to schedule as many users as possible, irrespective of their data types. The idea behind TRDCS is to schedule users at those intervals where resources are abundant and the demand is low. This will reduce the burden on heavily-loaded intervals. To achieve this goal, a priority metric is defined for every possible schedule. This metric equals the ratio of the CRAN resource capacity in that interval to the overall SUs' demands during the same interval. The overall demand is defined as resource requirement of the investigated interval, while excluding other intervals of that SU, plus demands of all intersecting parts of other SUs' intervals. For each SU, one interval with the maximum overlap is considered. We first accept the schedule with the maximum priority metric, then

TABLE 3. \mathcal{A}_i^t for $v = 0, 1, 2$ and update slot = $0, 1, \dots, 6$.

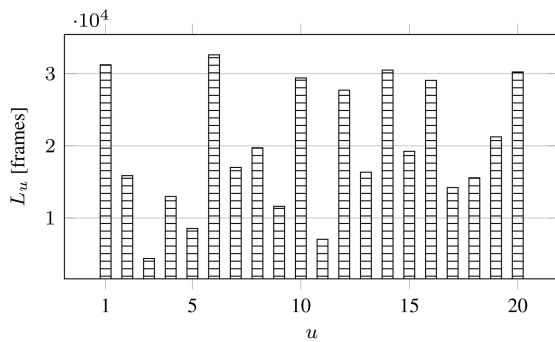
| | | | | | | | |
|---|-------------|-------------|-------------|-----------------|-----------------------|--------------------------|-----------------------------|
| | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| 0 | \emptyset | \emptyset | \emptyset | \emptyset | \emptyset | \emptyset | \emptyset |
| 1 | \emptyset | {5} | {5, 6} | {5, 6, 1, 3, 9} | {5, 6, 1, 3, 9, 2, 4} | {5, 6, 1, 3, 9, 2, 4, 8} | {5, 6, 1, 3, 9, 2, 4, 8, 7} |
| 2 | \emptyset | \emptyset | \emptyset | \emptyset | \emptyset | {1, 8}{2, 8} | {1, 8}{2, 8}{3, 7}{4, 7} |



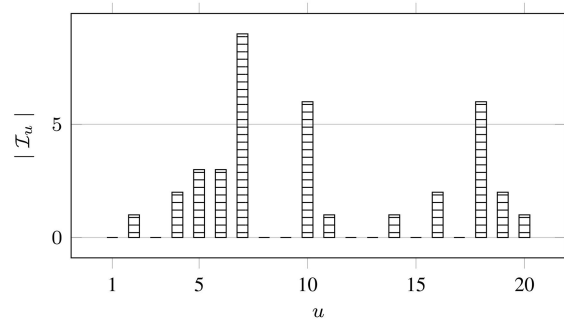
(a)



(b)



(c)



(d)

FIGURE 3. (a) An instance of the considered CRAN with $U = 20, R = 10$, and a single macro cell RRH that is fixed in center, (b) Minimum and maximum data transmission times, T_u^s and T_u^w , (c) Number of requested frames, L_u , (d) Histogram of the number of time intervals and resources, with respect to SU indices.

we remove all other intervals of the corresponding selected SU and all intervals of other users that do not meet the CRAN

resource constraints. Then, we iterate until all users are either scheduled or rejected.

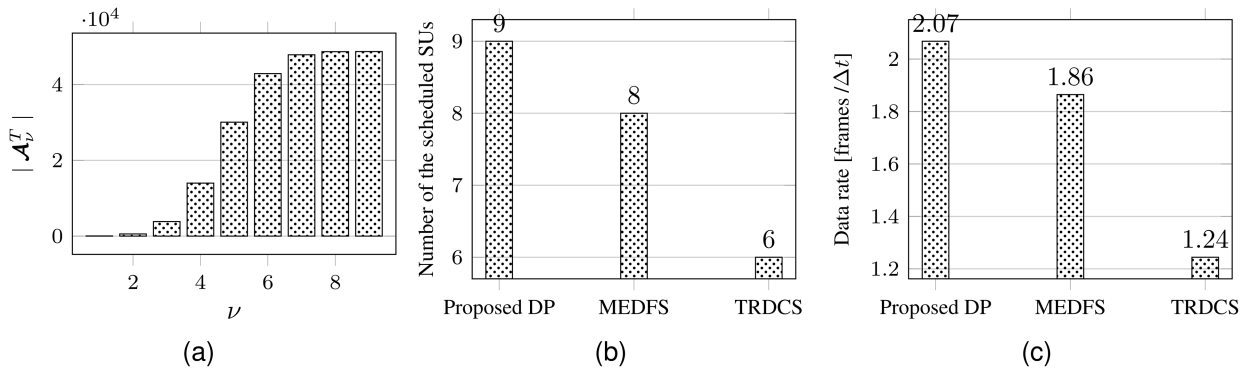


FIGURE 4. Results for an instance of the CRAN in Fig. 3: (a) $|\mathcal{A}_p^T|$, (b) The number of scheduled SUs, and (c) Total data rate.

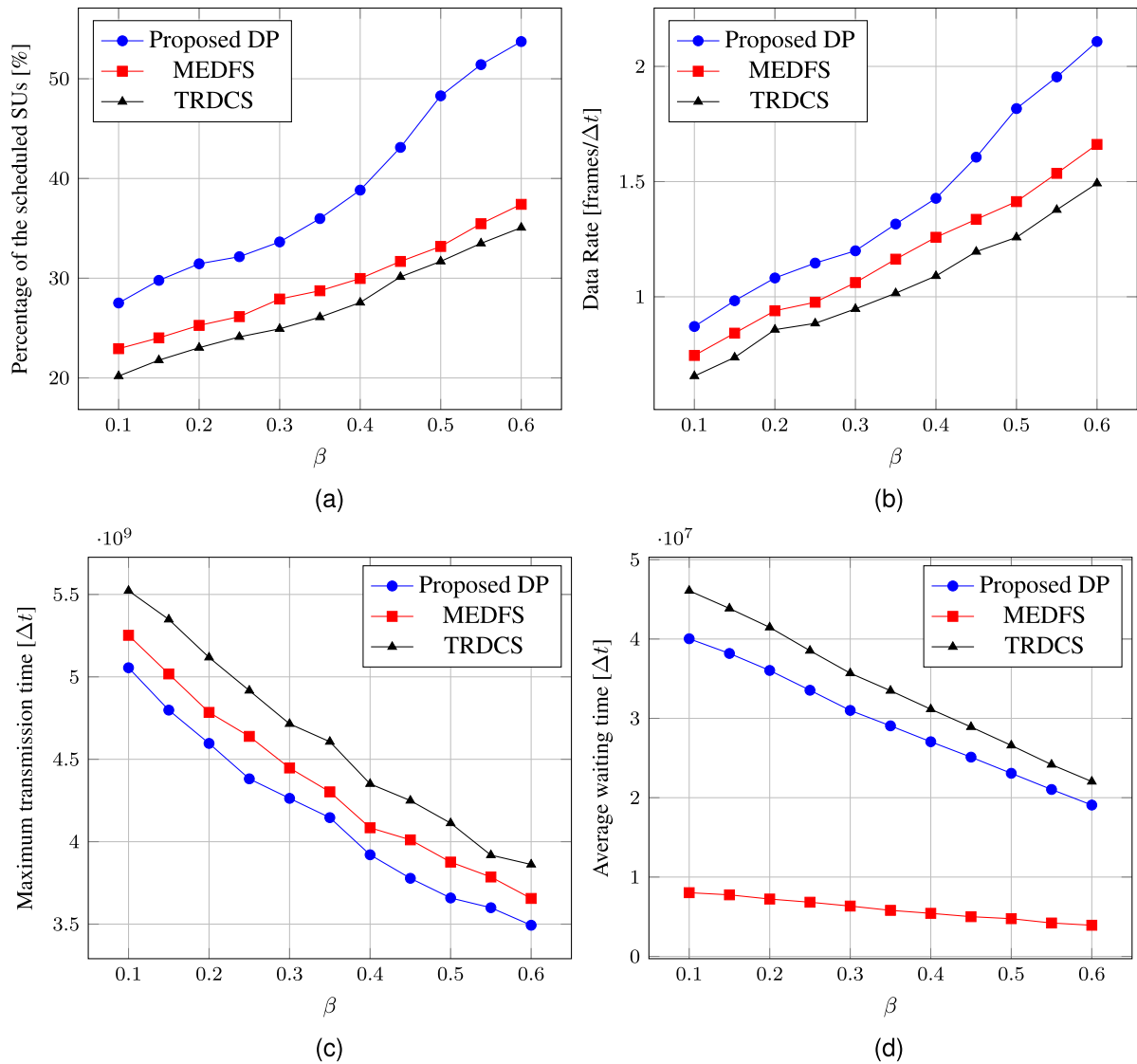


FIGURE 5. Monte Carlo simulation results in the CRAN with respect to β : (a) Percentage of the scheduled SUs, (b) Total data rate, (c) Maximum transmission time, and (d) Average waiting time.

A. SIMULATION SETUP

We consider a CRAN within $1000 \times 1000 \text{ m}^2$ area with multiple RRHs for serving SUs, and a single base station for

serving the PUs. These RRHs and SUs are uniformly and independently distributed within the square area. We assume that the SUs are either static or have low mobility, and

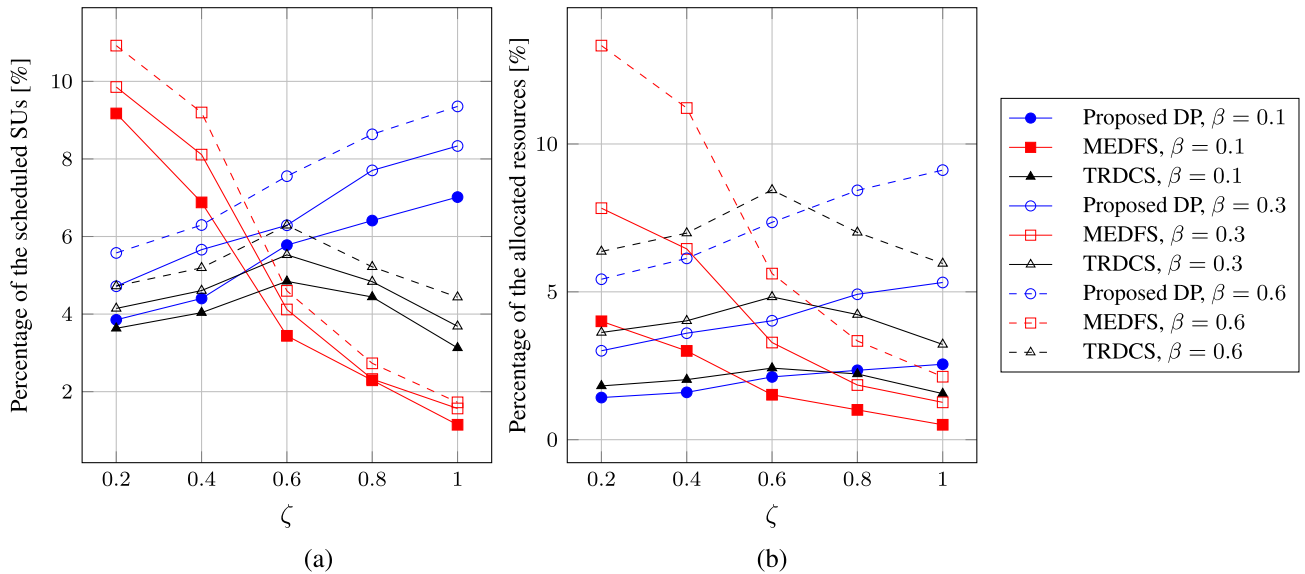


FIGURE 6. Performance of different algorithms with respect to ζ : (a) Percentage of scheduled SUs, and (b) Percentage of allocated resources.

TABLE 4. Values of simulation parameters.

| Parameter | Value |
|------------------|---------------------------|
| SNR Gap Γ | 0 dB [21], [23] |
| $P_{r,u}$ | 33 dBm [64] |
| Δt | 10 μ s [59] |
| s^{\max} | $(5/U) \mathcal{S} $ |
| α_u | $1, \frac{L_u}{\sum L_u}$ |
| a_o | 0.5 dB |

therefore their positions remain almost constant during one scheduling epoch. As a result, large-scale fading and shadowing are almost constant in one scheduling epoch. The proposed DP algorithm does not require knowledge of small-scale fading, but only its statistics are required. Once a RRH-user assignment is made by the algorithm, exact small-scale fading can be obtained online via pilot-based methods to enable the maximum ratio transmission beamforming by the RRHs. It is assumed the capacity of the backhaul and fronthaul links are large enough to support needed data flow of all scheduled users simultaneously [68]. Infinite buffer size is assumed for both RRHs and users. SUs independently make data requests with size L_u frames, which is uniformly distributed in the range $[2^8, 2^{16}]$. The background noise spectral density is equal to $\sigma^2 = -168.60$ dBm/Hz [22].

By assuming an urban environment for the small cells, the distance-dependent path loss of RRHs-SUs channels is given by: $PL[\text{dB}] = PL_o + 36.7 \log_{10} d_{r,u} - a_o$ where $d_{r,u} \geq 1.135$ m and is in meter. Furthermore, our operating frequency equals 900 MHz, corresponding to a wave length of $\frac{1}{3}$ m. Shadowing effect is modeled by the log-normal distribution with variance equal to 8 dB [69]. PL_o is equal to 30.58 dB, and a_o is a correction factor which is used to

take into account different antenna heights at RRHs and SUs. The total bandwidth of the network is assumed to be 20 MHz, with an activity rate 0.4 to 0.9 [70] for PUs. Correspondingly, the relative spectrum pool capacity, β , is in the range 0.1 to 0.6 of the total bandwidth. The free part of this bandwidth is partitioned into spectrum units with $\Delta f = 200$ KHz [71]. The spectrum units are occupied by PUs with an exponential distribution dwell time equal to $10^3 \times \Delta t$. Note that \mathcal{S}' can be formed by using these information. The values for the rest of simulation parameters are presented in Table 4.

B. SINGLE CRAN REALIZATION

An instance realization of the coverage area is shown in Fig. 3a for a CRAN with $U = 20$ and $R = 10$. To ensure a uniform spread of both RRHs and SUs across the overall square area, each pair of SUs is forced to maintain a minimum distances equal to $\sqrt{2} \times 1000\text{m}/20$. Similarly, RRHs are forced to maintain at least a distance of $\sqrt{2} \times 1000\text{m}/10$ to each other. The minimum and maximum waiting times, minimum and maximum possible data transmission times (corresponding to various schedules) for these SUs are shown in Fig. 3b when they make requests with lengths that are shown in Fig. 3c. For illustration purpose, these times are scaled with different factors. The histogram for the number of possible transmission time intervals and resources (see (4)) is plotted versus SU indices in Fig. 3d.

It can be observed that some SUs have no scheduling option because of the constraints in (6) and (8). The possible minimum and maximum service times for these SUs are set to zero in Fig. 3b. Therefore, they are rejected right away and are not given as viable inputs to Algorithm 1. Fig. 4a shows the number of disjunctive subsets. Figs. 4b and 4c show results for the number of the scheduled SUs, and total data rate,

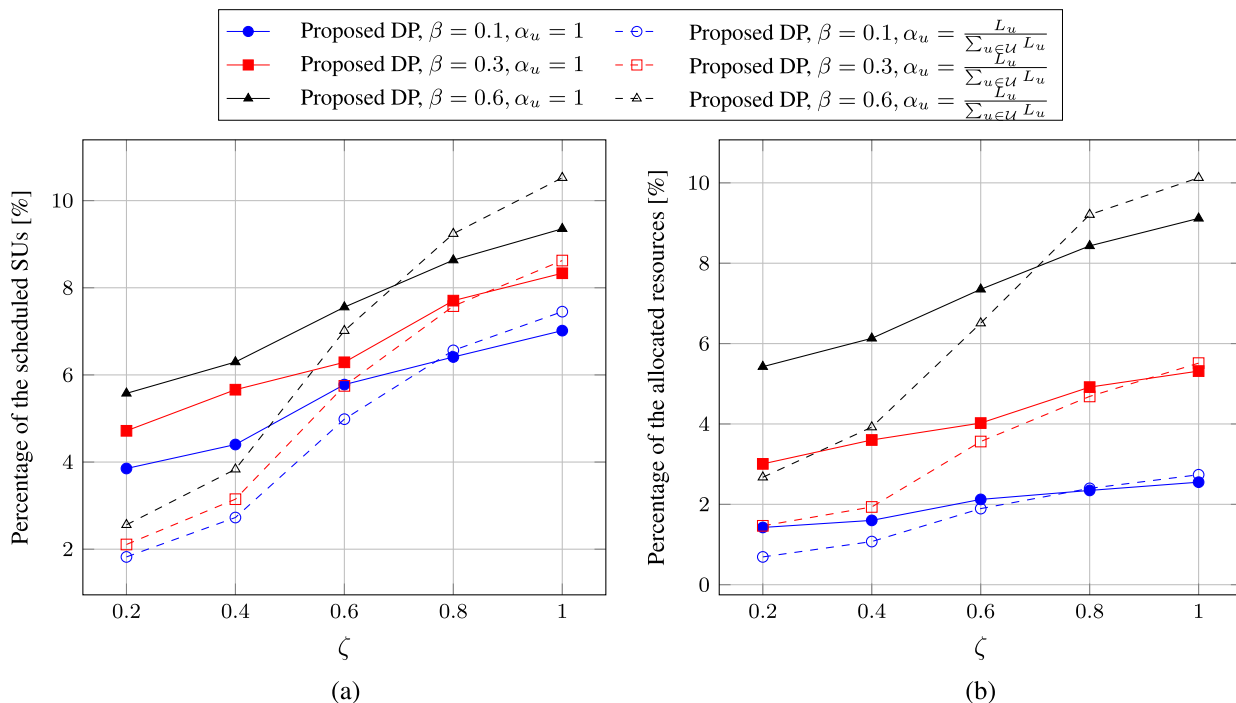


FIGURE 7. Performance of the proposed DP with respect to ζ when α_u equal to 1 and $\frac{L_u}{\sum_{u \in \mathcal{U}} L_u}$: (a) Percentage of scheduled SUs, and (b) Percentage of allocated resources.

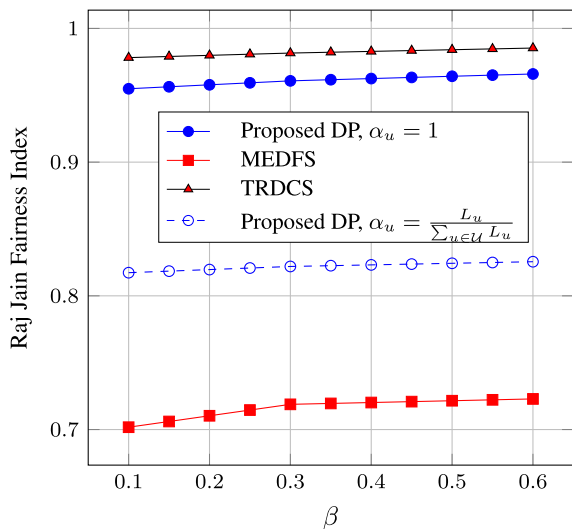


FIGURE 8. Average Raj Jain fairness indices of different algorithms.

respectively. These results clearly show better performance of our proposed DP approach.

C. MONTE CARLO SIMULATION RESULTS

The following simulation results are obtained by averaging over 10^4 randomly-generated CRAN realizations. The percentage of the scheduled SUs, total data rate, maximum transmission time, average waiting time for scheduled SUs in the CRAN with respect to β are illustrated in Fig. 5. The results show that the proposed DP achieves better performance in the percentage of the scheduled SUs and total data

rate with respect to different β , which represents spectrum pool capacity. The average percentages of the scheduled SUs are 38.75%, 29.24%, and 26.97%, for the proposed DP, MEDFS, and TRDCS, respectively. Although the percentages of the scheduled SUs increase with increasing β for all algorithms, the proposed DP enjoys an increasing number of available schedules and achieves a higher performance. The average data rates are 1.41, 1.18, and 1.04 [frames/ Δt] for the proposed DP, MEDFS, and TRDCS, respectively. Besides, due to the presence of (11) in the denominator of the objective function, our proposed DP offers the best result for the maximum transmission times, as shown in Fig 5c. Fig. 5d plots average waiting times of the scheduled SUs. The MEDFS algorithm achieves the minimum average waiting time in comparison with the proposed DP and TRDCS. However, with our proposed DP algorithm the scheduled SUs experience lower waiting times as compared to the TRDCS algorithm. We could use higher values of α_u and lower values of T_u^w for the SUs with real time services, and provide better waiting times for them.

D. BIG DATA

Recall that the main motivation to propose (12) is to favor and support big data transmission in the CRAN. In Fig. 6, the percentage of the scheduled SUs, and the percentage of the allocated resources are demonstrated with respect to ζ , which classifies data sizes into one of five ranges. Curves are plotted for three cases: $\beta = 0.1, 0.3$, and 0.6 . Fig. 6a shows that our proposed DP algorithm schedules more SUs with big data requests. Also, Fig. 6b demonstrates that our proposed approach has better performance in resource allocation for

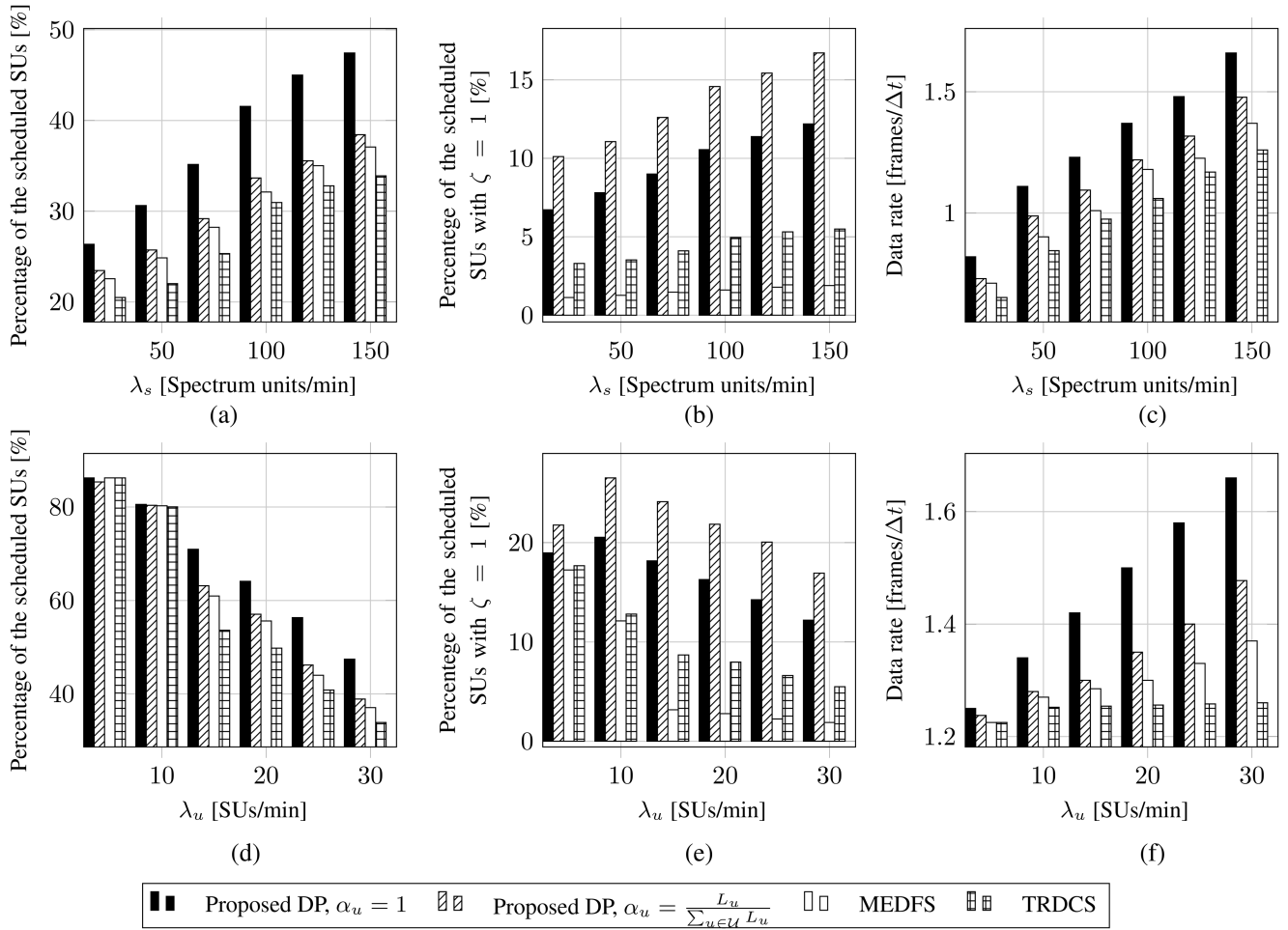


FIGURE 9. Long-term performance evaluation of different algorithms versus arrival rate of spectrum units: (a) Percentage of scheduled SUs, (b) Percentage of scheduled SUs with $\zeta = 1$; Long-term performance evaluation of different algorithms versus arrival rate of SUs: (c) Data rate, (d) Percentage of scheduled SUs, (e) Percentage of scheduled SUs with $\zeta = 1$, (f) Data rate.

the big volumes of data requests. This better performance is a direct consequence of using (11) in the objective function, and is achieved in addition to the higher data rate for CRAN, and higher percentage of the scheduled SUs (as studied in Subsection VI-C).

Fig. 7 illustrates the results obtained by setting priority weights of users to $\alpha_u = \frac{L_u}{\sum_{u \in \mathcal{U}} L_u}$ to favor big data users even more. We observe that better performance for big data users is achieved. However, this performance improvement comes at the cost of reduced fairness between the requests with different sizes. This fact is illustrated in Fig. 8 where the average Raj Jain fairness index of our proposed DP method is 96.10% and 82.17% for $\alpha_u = 1$ and $\alpha_u = \frac{L_u}{\sum_{u \in \mathcal{U}} L_u}$, respectively. However, the proposed DP method achieves a better fairness than the MEDFS method (with average Raj Jain fairness index of 71.59%) and lower fairness than the TRDCS method (with average Raj Jain fairness of 97.82%).

E. LONG TERM REALIZATION

A long-term realization of the CRAN is evaluated for 1 hour. We assume SUs arriving to the CRAN according to a Poisson

distribution with rate equal to λ_u [SUs/min], which is coupled with the data request rate. Also, we consider the rate λ_s [spectrum units/min] for the Poisson process that simulates arriving spectrum units in the spectrum pool. To have different conditions, SUs arrive in independent and uniform random locations of the CRAN.

Figs. 9a-c show the long-term performance of different algorithms versus λ_s , with $\lambda_u = 30$ [SUs/min] and $\beta \geq 0.1$. Fig. 9a shows that our proposed DP algorithm achieves a higher percentage of the scheduled SUs. By incorporating nonuniform priority in our approach, the percentages of the scheduled SUs are decreased, but still these percentages are higher in comparison with those of other two algorithms. Fig. 9b shows that by using this priority flexibility the BBU pool serves highest percentages of SUs with two upper deciles of data request sizes. Fig. 9c demonstrates that the BBU pool handles higher rates of data flow using our proposed DP method.

Figs. 9d-f illustrate long-term performances of different algorithms versus λ_u , with $\lambda_s = 150$ [Spectrum units/min]. In Fig. 9d, by increasing λ_u , all algorithms have more scheduling options. However, because of the fixed spectrum

pool capacity, percentages of the scheduled SUs decrease with increasing λ_u for all algorithms. Also, for the cases $\lambda_u = 5$ and 10, the BBU pool has large spectrum pool capacity and hence different algorithms produce the same results. Our proposed DP algorithm achieves higher percentages of the scheduled SUs in all values of λ_u . In Fig. 9e the highest percentages of the scheduled SUs with two upper deciles of data request sizes are for DP with $\alpha_u = \frac{L_u}{\sum_{u \in \mathcal{U}} L_u}$, $\alpha_u = 1$, TRDCS, and MEDFS, respectively. Finally, Fig. 9f shows that by using the proposed DP algorithm, the BBU pool achieves higher data rates in all values of λ_u .

VII. CONCLUSION AND FUTURE DIRECTIONS

We have formulated and solved a joint SU selection, RRHs and spectrum allocation, and deadline-aware time scheduling problem for efficient downlink data transmission of SUs in a cognitive CRAN. The problem is formulated in such a way to give a higher priority to big data requests. The problem was shown to be NP-hard. We then proposed a DP-based approach to reduce the complexity in finding the optimum solution. The proposed solution approach has two phases. In the first phase, all feasible solutions are obtained by Algorithm 1, and in the second phase, optimum schedules are obtained by Algorithm 2. Simulation results demonstrate that significant improvements in total data rate, percentage of the scheduled SUs, and most importantly, higher priority for big data transmissions are achieved by the proposed algorithm in comparison with two baseline scheduling algorithms which are referred to as MEDFS and TRDCS algorithms.

As future research directions, this work can be extended to the scenarios where RRH and/or SUs have multiple antennas, and/or when the BBU pool has imperfect knowledge of RRH-SU links. This imperfect knowledge can be in the form of statistical CSI, quantized CSI, or estimated CSI. In addition, techniques such as NOMA and/or beamforming may be utilized concurrently to simultaneously allocate a vacant spectrum unit to more than one SU. Moreover, the maximum number of spectrum units that can be used by each SU (s^{\max}) and the maximum number of SUs that can be simultaneously served by each RRH (u^{\max}) could be designed in an optimal fashion based on the requested data types and CSI availability.

REFERENCES

- [1] Cisco Annual Internet Report (2018–2023), White Paper, Cisco, San Jose, CA, USA, 2020.
- [2] M. Agiwal, A. Roy, and N. Saxena, "Next generation 5G wireless networks: A comprehensive survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 1617–1655, 3rd Quart., 2016.
- [3] J. Navarro-Ortiz, P. Romero-Diaz, S. Sendra, P. Ameigeiras, J. J. Ramos-Munoz, and J. M. Lopez-Soler, "A survey on 5G usage scenarios and traffic models," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 2, pp. 905–929, 2nd Quart., 2020.
- [4] H. Viswanathan and P. E. Mogensen, "Communications in the 6G era," *IEEE Access*, vol. 8, pp. 57063–57074, 2020.
- [5] W. S. H. M. W. Ahmad, N. A. M. Radzi, F. S. Samidi, A. Ismail, F. Abdullah, M. Z. Jamaludin, and M. N. Zakaria, "5G technology: Towards dynamic spectrum sharing using cognitive radio networks," *IEEE Access*, vol. 8, pp. 14460–14488, 2020.
- [6] N. Zhang, P. Yang, J. Ren, D. Chen, Y. Li, and X. Shen, "Synergy of big data and 5G wireless networks: Opportunities, approaches, and challenges," *IEEE Wireless Commun.*, vol. 25, no. 1, pp. 12–18, Feb. 2018.
- [7] A. Ghosh, A. Maeder, M. Baker, and D. Chandramouli, "5G evolution: A view on 5G cellular technology beyond 3GPP release 15," *IEEE Access*, vol. 7, pp. 127639–127651, 2019.
- [8] *5G; System Architecture for the 5G System (5GS)*, document TS 23.501, Rev. 16.7.0 Release 16, 3GPP, 2021.
- [9] J. Gozalvez, "5G worldwide developments [mobile radio]," *IEEE Veh. Technol. Mag.*, vol. 12, no. 1, pp. 4–11, Mar. 2017.
- [10] (Jan. 2020). *5G evolution and 6G*. NTT DoCoMo, INC. [Online]. Available: https://www.nttdocomo.co.jp/english/binary/pdf/corporate/technology/whitepaper_6g/DOCOMO_6G_White_PaperEN_20200124.pdf
- [11] K. Wang, Y. Wang, X. Hu, Y. Sun, D. Deng, A. Vinea, and Y. Zhang, "Wireless big data computing in smart grid," *IEEE Wireless Commun.*, vol. 24, no. 2, pp. 58–64, Apr. 2017.
- [12] X. Yang, T. Wang, X. Ren, and W. Yu, "Survey on improving data utility in differentially private sequential data publishing," *IEEE Trans. Big Data*, vol. 7, no. 4, pp. 729–749, Oct. 2021.
- [13] M. A. Habibi, M. Nasimi, B. Han, and H. D. Schotten, "A comprehensive survey of RAN architectures toward 5G mobile communication system," *IEEE Access*, vol. 7, pp. 70371–70421, 2019.
- [14] *C-RAN: The Road Towards Green RAN*, China Mobile Res. Inst., Beijing, China, Oct. 2011.
- [15] D. Harutyunyan and R. Riggio, "How to migrate from operational LTE/LTE-A networks to C-RAN with minimal investment?" *IEEE Trans. Netw. Service Manage.*, vol. 15, no. 4, pp. 1503–1515, Dec. 2018.
- [16] I. A. Alimi, A. L. Teixeira, and P. P. Monteiro, "Toward an efficient C-RAN optical fronthaul for the future networks: A tutorial on technologies, requirements, challenges, and solutions," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 1, pp. 708–769, Nov. 2018.
- [17] *IEEE Standard for Packet-Based Fronthaul Transport Networks*, Standard 1914.1-2019, IEEE, 2020, pp. 1–94.
- [18] M. Peng, Y. Sun, X. Li, Z. Mao, and C. Wang, "Recent advances in cloud radio access networks: System architectures, key techniques, and open issues," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 3, pp. 2282–2308, Aug. 2016.
- [19] B. Niu, Y. Zhou, H. Shah-Mansouri, and V. W. Wong, "A dynamic resource sharing mechanism for cloud radio access networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 12, pp. 8325–8338, Dec. 2016.
- [20] A. Douik, H. Dahrouj, T. Y. Al-Naffouri, and M.-S. Alouini, "Coordinated scheduling for the downlink of cloud-radio-access networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Jun. 2015, pp. 2906–2911.
- [21] A. Douik, H. Dahrouj, T. Y. Al-Naffouri, and M.-S. Alouini, "Coordinated scheduling and power control in cloud-radio access networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 4, pp. 2523–2536, Apr. 2016.
- [22] A. Douik, H. Dahrouj, T. Y. Al-Naffouri, and M.-S. Alouini, "Low-complexity scheduling and power adaptation for coordinated cloud-radio access networks," *IEEE Commun. Lett.*, vol. 21, no. 10, pp. 2298–2301, Oct. 2017.
- [23] A. Douik, H. Dahrouj, T. Y. Al-Naffouri, and M.-S. Alouini, "Distributed hybrid scheduling in multi-cloud networks using conflict graphs," *IEEE Trans. Commun.*, vol. 66, no. 1, pp. 209–224, Jan. 2018.
- [24] A. Douik, H. Dahrouj, T. Y. Al-Naffouri, and M.-S. Alouini, "Joint scheduling and beamforming via cloud-radio access networks coordination," in *Proc. IEEE 88th Veh. Technol. Conf. (VTC-Fall)*, Aug. 2018, pp. 1–5.
- [25] M. Awais, A. Ahmed, M. Naeem, M. Iqbal, W. Ejaz, A. Anpalagan, and S. and H Kim, "Efficient joint user association and resource allocation for cloud radio access networks," *IEEE Access*, vol. 5, pp. 1439–1448, 2017.
- [26] L. Ferdouse, W. Ejaz, A. Anpalagan, and A. M. Khattak, "Joint workload scheduling and BBU allocation in cloud-RAN for 5G networks," in *Proc. Symp. Appl. Comput.*, New York, NY, USA, Apr. 2017, pp. 621–627, doi: 10.1145/3019612.3019770.
- [27] Y. Li, M. C. Gursoy, and S. Velipasalar, "Intercell interference-aware scheduling for delay sensitive applications in C-RAN," in *Proc. IEEE 86th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2017, pp. 1–5.
- [28] W. Xia, T. Q. S. Quek, J. Zhang, S. Jin, and H. Zhu, "Programmable hierarchical C-RAN: From task scheduling to resource allocation," *IEEE Trans. Wireless Commun.*, vol. 18, no. 3, pp. 2003–2016, Mar. 2019.
- [29] H. M. Soliman and A. Leon-Garcia, "Fully distributed scheduling in cloud-RAN systems," in *Proc. IEEE Wireless Commun. Netw. Conf.*, Apr. 2016, pp. 1–6.

- [30] K. Lin, W. Wang, Y. Zhang, and L. Peng, "Green spectrum assignment in secure cloud radio network with cluster formation," *IEEE Trans. Sustain. Comput.*, vol. 4, no. 2, pp. 191–203, Apr. 2019.
- [31] L. You and D. Yuan, "User-centric performance optimization with remote radio head cooperation in C-RAN," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 340–353, Jan. 2020.
- [32] Y. Zhou, S. Ci, and Y. Yang, "Energy-aware joint clustering and scheduling for multicast beamforming in cloud-RAN downlink," *IEEE Wireless Commun. Lett.*, vol. 9, no. 4, pp. 461–464, Apr. 2020.
- [33] C.-H. Fang, P.-R. Li, and K.-T. Feng, "Joint interference cancellation and resource allocation for full-duplex cloud radio access networks," *IEEE Trans. Wireless Commun.*, vol. 18, no. 6, pp. 3019–3033, Jun. 2019.
- [34] Z. Lin and Y. Liu, "Joint uplink and downlink transmissions in user-centric OFDMA cloud-RAN," *IEEE Trans. Veh. Technol.*, vol. 68, no. 8, pp. 7776–7788, Aug. 2019.
- [35] S. Jabeen and P.-H. Ho, "Joint user scheduling and RRU association for OFDMA-based networks with inter-BBU coordination," *IEEE Trans. Commun.*, vol. 67, no. 5, pp. 3638–3650, May 2019.
- [36] W. Tang and S. Feng, "User selection and power minimization in full-duplex cloud radio access networks," *IEEE Trans. Signal Process.*, vol. 67, no. 9, pp. 2426–2438, May 2019.
- [37] A. Kumar, K. Kumar, M. S. Gupta, and S. Kumar, "A survey on NOMA techniques for 5G scenario," in *Proc. Int. Conf. Adv. Electron., Elect. Comput. Intell. (ICAEEC)*, Apr. 2020.
- [38] A. Kumar and K. Kumar, "Multiple access schemes for cognitive radio networks: A survey," *Phys. Commun.*, vol. 38, Feb. 2020, Art. no. 100953, doi: [10.1016/j.phycom.2019.100953](https://doi.org/10.1016/j.phycom.2019.100953).
- [39] A. Kumar and K. Kumar, "A game theory based hybrid NOMA for efficient resource optimization in cognitive radio networks," *IEEE Trans. Netw. Sci. Eng.*, vol. 8, no. 4, pp. 3501–3514, Oct. 2021.
- [40] H. Dai, Y. Huang, J. Wang, and L. Yang, "Resource optimization in heterogeneous cloud radio access networks," *IEEE Commun. Lett.*, vol. 22, no. 3, pp. 494–497, Mar. 2018.
- [41] Q. Liu, S. Sun, and H. Gao, "Joint user-centric clustering and frequency allocation in ultra-dense C-RAN," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, May 2020, pp. 1–6.
- [42] A. Younis, T. X. Tran, and D. Pompili, "Bandwidth and energy-aware resource allocation for cloud radio access networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 10, pp. 6487–6500, Oct. 2018.
- [43] X. Wang, K. Wang, S. Wu, S. Di, H. Jin, S. Ou, and K. Yang, "Dynamic resource scheduling in mobile edge cloud with cloud radio access network," *IEEE Trans. Parallel Distrib. Syst.*, vol. 29, no. 11, pp. 2429–2445, Nov. 2018.
- [44] H. Khedher, S. Hoteit, P. Brown, V. Veque, R. Krishnaswamy, W. Diego, and M. Hadji, "Real traffic-aware scheduling of computing resources in cloud-RAN," in *Proc. Int. Conf. Comput., Netw. Commun. (ICNC)*, Feb. 2020, pp. 422–427.
- [45] S. D'Oro, M. A. Marotta, C. B. Both, L. DaSilva, and S. Palazzo, "Power-efficient resource allocation in C-RANs with SINR constraints and deadlines," *IEEE Trans. Veh. Technol.*, vol. 68, no. 6, pp. 6099–6113, Jun. 2019.
- [46] C. L. Liu and J. W. Layland, "Scheduling algorithms for multiprogramming in a hard-real-time environment," *J. ACM*, vol. 20, no. 1, pp. 46–61, 1973, doi: [10.1145/321738.321743](https://doi.org/10.1145/321738.321743).
- [47] D. Casini, A. Biondi, and G. Buttazzo, "Handling transients of dynamic real-time workload under EDF scheduling," *IEEE Trans. Comput.*, vol. 68, no. 6, pp. 820–835, Jun. 2019.
- [48] K. Wang and Y. Cen, "Real-time partitioned scheduling in cloud-RAN with hard deadline constraint," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2017, pp. 1–6.
- [49] R. K. Saha, S. Nanba, and K. Nishimura, "A technique for cloud based clustering and spatial resource reuse and scheduling of 3D in-building small cells using CoMP for high capacity CRAN," *IEEE Access*, vol. 6, pp. 71602–71621, 2018.
- [50] X. Yi, F. Liu, J. Liu, and H. Jin, "Building a network highway for big data: Architecture and challenges," *IEEE Netw.*, vol. 28, no. 4, pp. 5–13, Jul. 2014.
- [51] W.-J. Wang, H.-C. Yang, and M.-S. Alouini, "Energy consumption analysis for adaptive transmission of big data over fading channels: A statistical characterization," *IEEE Trans. Green Commun. Netw.*, vol. 4, no. 2, pp. 365–374, Jun. 2020.
- [52] F. Al-Turjman, L. Mostarda, E. Ever, A. Darwish, and N. S. Khalil, "Network experience scheduling and routing approach for big data transmission in the Internet of Things," *IEEE Access*, vol. 7, pp. 14501–14512, 2019.
- [53] X. Zheng and Z. Cai, "Real-time big data delivery in wireless networks: A case study on video delivery," *IEEE Trans. Ind. Informat.*, vol. 13, no. 4, pp. 2048–2057, Aug. 2017.
- [54] R. Li and H. Asaeda, "MWBS: An efficient Many-to-Many wireless big data delivery scheme," *IEEE Trans. Big Data*, vol. 6, no. 2, pp. 233–247, Jun. 2020.
- [55] Z. Su and Q. Xu, "Security-aware resource allocation for mobile social big data: A matching-coalitional game solution," *IEEE Trans. Big Data*, vol. 7, no. 4, pp. 632–642, Oct. 2021.
- [56] S. M. Srinivasan, T. Truong-Huu, and M. Gurusamy, "Deadline-aware scheduling and flexible bandwidth allocation for big-data transfers," *IEEE Access*, vol. 6, pp. 74400–74415, 2018.
- [57] R. van Bevern, M. Mnich, R. Niedermeier, and M. Weller, "Interval scheduling and colorful independent sets," *J. Scheduling*, vol. 18, no. 5, pp. 449–469, Oct. 2015.
- [58] *IEEE Standard for Information Technology—Telecommunications and Information Exchange Between Systems Local and Metropolitan Area Networks—Specific Requirements—Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, Standard 802.11-2016 (Revision IEEE Std 802.11-2012), IEEE, 2016, pp. 2321–2582.
- [59] *5G; NR; Physical Layer Procedures for Control*, document TS 38.213, Rev. 16.4.0 Release 16, 3GPP, 2021.
- [60] S. Park, O. Simeone, and S. Shamai, "Multi-tenant C-RAN with spectrum pooling: Downlink optimization under privacy constraints," *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 10492–10503, Nov. 2018.
- [61] T. X. Tran, D. V. Le, G. Yue, and D. Pompili, "Cooperative hierarchical caching and request scheduling in a cloud radio access network," *IEEE Trans. Mobile Comput.*, vol. 17, no. 12, pp. 2729–2743, Dec. 2018.
- [62] A. W. Kolen, J. K. Lenstra, C. H. Papadimitriou, and F. C. Spieksma, "Interval scheduling: A survey," *Naval Res. Logistics*, vol. 54, no. 5, pp. 530–543, 2007.
- [63] A. Biswas, V. Raman, and S. Saurabh, *Solving Group Interval Scheduling Efficiently* (Lecture Notes in Computer Science), vol. 11638. Springer, 2019, doi: [10.1007/978-3-030-25005-8_9](https://doi.org/10.1007/978-3-030-25005-8_9).
- [64] *Electromagnetic Field Compliance Assessments for 5G Wireless Networks*, Electromagn. Field Compliance Assessments 5G wireless Netw., Int. Telecommun. Union, Geneva, Switzerland, 2019.
- [65] A. B. Sediq, R. H. Gohary, R. Schoenen, and H. Yanikomeroglu, "Optimal tradeoff between sum-rate efficiency and Jain's fairness index in resource allocation," *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 3496–3509, Jul. 2013.
- [66] M.-S. Chen, C.-P. Hwang, and H.-F. Wang, "Analysis of the queue service probability for the EDF scheduling algorithm," in *Proc. 30th Int. Conf. Adv. Inf. Netw. Appl. Workshops (WAINA)*, Mar. 2016, pp. 960–963.
- [67] F. C. R. Spieksma, "On the approximability of an interval scheduling problem," *J. Scheduling*, vol. 2, pp. 215–227, Sep. 1999, doi: [10.1002/\(SICI\)1099-1425\(199909/10\)2:5<215::AID-JOS27>3.0.CO;2-Y](https://doi.org/10.1002/(SICI)1099-1425(199909/10)2:5<215::AID-JOS27>3.0.CO;2-Y).
- [68] S. Gonzalez-Diaz, A. Garcia-Saavedra, A. de la Oliva, X. Costa-Perez, R. Gazda, A. Mourad, T. Deiss, J. Manges-Bafalluy, P. Iovanna, S. Stracca, and P. Leithead, "Integrating fronthaul and backhaul networks: Transport challenges and feasibility results," *IEEE Trans. Mobile Comput.*, vol. 20, no. 2, pp. 533–549, Feb. 2021.
- [69] S. Sun, T. S. Rappaport, S. Rangan, T. A. Thomas, A. Ghosh, I. Z. Kovacs, I. Rodriguez, O. Koymen, A. Partyka, and J. Jarvelainen, "Propagation path loss models for 5G urban micro- and macro-cellular scenarios," in *Proc. IEEE 83rd Veh. Technol. Conf. (VTC Spring)*, May 2016, pp. 1–6.
- [70] W. Ning, X. Huang, K. Yang, F. Wu, and S. Leng, "Reinforcement learning enabled cooperative spectrum sensing in cognitive radio networks," *J. Commun. Netw.*, vol. 22, no. 1, pp. 12–22, Feb. 2020.
- [71] *5G; NR; Physical Channels and Modulation*, document TS 38.211, Rev. 16.4.0 Release 16, 3GPP, 2021.



MOHAMMAD BIGDELI received the B.Sc. degree in electrical engineering and computer engineering and the M.Sc. degree (Hons.) in telecommunications engineering from the Iran University of Science and Technology, Tehran, Iran, in 2009 and 2012, respectively, where he is currently pursuing the Ph.D. degree with the School of Electrical Engineering. His current research interests include big data, information theory, channel coding, signal processing for communications, wireless networking, and cooperative communications. He has served as a reviewer for several IEEE journals and major conferences.



SHAHROKH FARAHMAND was born in Tehran, Iran, in 1980. He received the B.Sc. degree in electrical engineering from the Sharif University of Technology, in 2003, and the M.Sc. and Ph.D. degrees in the field of communications and signal processing from the University of Minnesota (UMN) at Twin-Cities, in 2006 and 2011, respectively. From 2011 to 2014, he was with the Iran Research Organization for Science and Technology (IROST) where he held a research faculty position. Since 2018, he has been with the Electrical Engineering Department, Iran University of Science and Technology (IUST), where he is currently an Assistant Professor. His research interests include applications of statistical signal processing, optimization, and machine learning in communications and networking. His current focus is on the Internet of Things (IoT), intelligent reflecting surfaces (IRS), massive MIMO, and ultra-wideband impulse radio (UWB-IR).



BAHMAN ABOLHASSANI was born in Tehran, Iran. He received the B.S. degree in electrical engineering from the Iran University of Science and Technology (IUST), Tehran, and the M.S. and Ph.D. degrees in electrical engineering from the University of Saskatchewan, Saskatoon, SK, Canada. He was an Instrumentation Engineer with the College of Water and Power Technology, Iranian Ministry of Energy, for three years. Then, he worked as a Communication System Engineer in a number of private and government companies. He joined the School of Electrical Engineering, IUST, where he is currently an Associate Professor. He worked as the Dean of the School of Electrical Engineering and an Associate Dean for Research. He also worked as a Sessional Lecturer at the University of Saskatchewan. His research interests include wireless communication systems, network planning, spread spectrum, cognitive radio networks, resource allocation, VANETs, and optimization of large systems.



HA H. NGUYEN (Senior Member, IEEE) received the B.Eng. degree in electrical engineering from the Hanoi University of Technology (HUT), Hanoi, Vietnam, in 1995, the M.Eng. degree in electrical engineering from the Asian Institute of Technology (AIT), Bangkok, Thailand, in 1997, and the Ph.D. degree in electrical engineering from the University of Manitoba, Winnipeg, MB, Canada, in 2001. He joined the Department of Electrical and Computer Engineering, University of Saskatchewan, Saskatoon, SK, Canada, in 2001, and became a Full Professor, in 2007. He currently holds the position of a NSERC/Cisco Industrial Research Chair in Low-Power Wireless Access for Sensor Networks. He is the coauthor, with Ed Shwedyk, of the textbook “*A First Course in Digital Communications*” (published by Cambridge University Press). His research interests include broad areas of communication theory, wireless communications, and statistical signal processing. He is a fellow of the Engineering Institute of Canada (EIC) and a Registered Member of the Association of Professional Engineers and Geoscientists of Saskatchewan (APEGS). He served as a technical program chair for numerous IEEE events. He was an Associate Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and the IEEE WIRELESS COMMUNICATIONS LETTERS, during 2007–2011 and during 2011–2016, respectively. He also serves as an Associate Editor for the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY.

• • •