Abstract—The cloud radio access network (C-RAN) concept is a strong candidate for next-generation wireless networks due to its flexibility, scalability, increased asset utilization, and energy savings. While resources are centralized in C-RAN architecture, wireless resource virtualization (WRV) can offer substantial capacity increase and efficiency gain. This paper provides solutions for virtualizing C-RANs wireless resources and sharing them between multiple mobile network operators (MNOs). The proposed solutions dynamically allocate wireless resources to users who subscribe to MNOs across the network. In addition, the proposed solutions maintain a high level of isolation between different MNOs, provide efficient and fair resource utilization, enable different scheduling policies, and manage intercell interference (ICI). An optimal solution is formulated as a combinatorial problem, which is computationally expensive. Consequently, two low-complexity suboptimal solutions with comparable performance are provided. The optimal and suboptimal solutions are compared in terms of complexity and performance. The optimal solution outperforms the suboptimal by only 6% at the expense of a significantly higher time complexity.

Keywords—Wireless resource virtualization, C-RAN, ICI management, resource allocation.

I. INTRODUCTION

Mobile network operators (MNOs) are desperate for cost-effective, scalable, and innovative solutions in order to improve network capacity and coverage to handle the ever-growing demand for mobile data transmission. Traditional solutions of deploying more base stations (BSs) and acquiring new spectrum licenses are becoming economically unsustainable. Therefore, cellular wireless networks should develop new techniques to enhance the utilization of the computing and wireless resources. Recent spectrum utilization measurements show that the bandwidth licensed to MNOs is mostly underutilized [1], [2]. According to a recent study by Nokia [3], only 20% of a radio access network’s full capacity is used at any given time with 80% idle waiting for peak hour demand.

Typically, the average traffic load in cellular networks is considered lower than the peak load. In addition, the peak load time depends on the geographical location, because different locations might have different peak load times. For example, peak loads in residential neighborhoods occur at night, whereas peak loads in central office districts happen during business hours. Conventionally, BSs consist of network elements running purpose-built software and dedicated hardware. Therefore, efficient network design becomes very challenging because the network must employ sophisticated scalability techniques to handle a wide range of traffic loads at different times. Resource over-provisioning is costly, while under-provisioning compromises service quality. Moreover, adding new services or modifying existing ones is not a trivial process because it creates an impediment to network scalability, increases the time to market, and obstructs efficient utilization of the infrastructure.

With the growth of cloud-hosted development platforms and services, new promising solutions are on the horizon for MNOs. In 2010, China Mobile introduced the Cloud Radio Access Network (C-RAN) as a new platform for future mobile network infrastructure [4]. In C-RAN, the baseband processing resources are aggregated and virtualized on a cloud data center, which allows better utilization of computing resources and more cooperative sharing of radio resources across wireless nodes. C-RAN technology promises MNOs with significant reductions in capital expenditure (CAPEX) and operating expenditure (OPEX), as well as reduced energy consumption. Using C-RAN is expected to reduce power consumption by about 71% compared to traditional RAN [5].

The main concept behind C-RAN architecture is to decouple the radio function unit of RANs and the processing unit, as shown in Fig. 1; where a low-cost remote radio head (RRH) replaces the radio function unit. The baseband processing functions and resource management are delegated to baseband processing units (BPUs), which are pooled on a cloud in remote data centers. In traditional RAN architectures, the processing resources of BSs cannot be shared or scaled based on the traffic load. Since the processing power in C-RAN is centralized, innovative cloud-based solutions can be applied to improve the utilization of processing resources. Consequently, C-RAN architecture requires fewer BPUs compared to the traditional RAN architecture [6], [7].

Centralizing baseband processing and management enables better coordination across RRHs because cell site information, such as traffic loads, user-channel conditions, and user-traffic requirements, are available across the network. Such information can be effectively exploited to optimize the allocation of radio resources across cell sites, manage intercell interference (ICI), and improve coverage and handover procedures. In addition, sharing information can enhance capacity by facilitating the implementation of new technologies such as ICI coordination (ICIC), self-organizing networks (SON), and coordinated multipoint (CoMP) transmission.
Similar to C-RAN that centralizes the computing resources, wireless resource virtualization (WRV) is a new paradigm aims at centralizing wireless spectrum resources and sharing them between MNOs. WRV enables network operators to create multiple logical networks on a single physical substrate yielding better efficiency in terms of energy consumption and resource utilization. A new study from ABI Research [8] shows that over a period of five years, deploying active infrastructure-sharing worldwide can save up to 60 billion USD in OPEX and CAPEX.

As MNOs need to continually enhance the capacity of mobile networks, sharing radio resources between multiple MNOs facilitates carrier resource aggregation and supports higher peak rates. It also introduces multi-MNO multiplexing gains as a result of increasing the number of users per cell. For instance, in a Rayleigh fading channels, the aggregated capacity of a cell can increase by $\ln(K)$, where $K$ is the number of users in the cell [9]. Furthermore, network sharing facilitates new business models in the wireless market. For example, operators without Long Term Evolution (LTE) licenses or network resources will be able to provide LTE services by renting LTE radio resources from other MNOs.

Combining C-RAN and WRV will help MNOs to overcome the challenges inherent to the current mobile networks. The aim of this paper is to virtualize the wireless resources of cloud-based RANs such that they can be shared between multiple MNOs. An entity called Hypervisor is added on top of the physical network as shown in Fig. 1, which is responsible for allocating the available wireless resources to users subscribed to different MNOs. The allocation of wireless resources is determined by the sharing contract between MNOs, the traffic load at each RRH, and the interference between different RRHs. The rest of the paper is organized as follows: Section II discusses related work in the field and the contribution of this work. Section III presents the system model, and problem formulation. The optimal solution of the problem is presented in Section IV. Low-complexity schedulers are presented in Section V. Section VI presents and discusses simulation results, and Section VII concludes the paper.

II. RELATED WORK AND CONTRIBUTIONS

Part of the work discussed in this paper is reported in [10], where the wireless resources of a single cell zone are virtualized. However, this work addresses wireless resource-scheduling techniques across multiple cell zones shared between multiple MNOs. In contrast to [10], the present work considers a network-wide virtualization and coordination between interfering cell zones to prevent ICI. Moreover, the current work manages the interference between different cell zones by using graph theory. In addition, more solutions and performance comparisons are provided in this work.

In the recent literature, it can be clearly noticed that there is a growing interest in virtualizing networks’ wireless resources. For example, an LTE air interface virtualization scheme is proposed in [11], where a hypervisor is added on top of the physical resources. The hypervisor is responsible for virtualizing the evolved NodeB (eNB) into a number of virtual eNBs that can be used by different MNOs. It is shown that more capacity can be achieved by sharing spectrum resources between different MNOs.

More practical scenarios that consider load balancing are studied in [12], [13], where the hypervisor manages the sharing process of multiple eNBs among multiple MNOs. Nevertheless, only fixed resource allocation across BSs is considered. The load is balanced between multiple BSs by moving users from high-traffic cells to low-traffic cells. However, transferring users across cells increases handover overhead, and may degrade the system capacity as users may be transferred to BSs further away, which would reduce the quality of the wireless link.

Another framework for wireless network virtualization that separates service providers from a network operator is reported in [14]. The service providers (SPs) are responsible for QoS management, while the network operator is responsible for spectrum management. The interaction between SPs and the network operator is modeled as a stochastic game regulated by the network operator. The role of the SPs is to compete for wireless resources for each subscribed user.

Network virtualization substrate (NVS) [15] is a flow-level solution that divides the wireless resources on a single BSs into different slices. NVS enables customized flow scheduling per slice, where each slice can be seen as a virtual MNO that supports a set of flows. An admission control procedure is assumed to restrict the number of flows established and
to guarantee feasible solutions. However, NVS schedules the resources for each BS independently, which may deteriorate the network resource utilization [2]. In addition, NVS fails to maintain a minimum aggregate resource allocation for each slice across the network. For example, assume that an entity (flow group) reserves 50% of the spectrum resource; irrespective of the demands, then, NVS allocates 50% or fewer resources to the entity at every BS in the network. To address this shortcoming, the same authors of [15] presented NetShare [2], a flexible solution to dynamically reschedule network resources to flow groups, based on both their aggregate resource reservations and demands at each BS. With NetShare, flow groups compute the resource demands for resources generated by their subscribers at each BS and report them to a central controller. Feedback is sent from the central controller to each BS in order to reconfigure its NVS scheduler to meet individual users requirements across the network. However, NetShare is applied to a traditional RAN architecture, where each BS has a fixed number of radio resource blocks (RBs) regardless of its instantaneous traffic load. Thus, when a BS is congested with a high number of users, users requirements may not be satisfied.

In multi-cell systems, the same frequency bands can be assigned to users in different cells, which is referred to as the frequency-reuse (FR) principle, which is used to increase both coverage and capacity. However, to minimize ICI, cells that use the same frequency bands should be separated by a sufficient distance. Several ICIC techniques have been proposed for multi-cell systems as described in [16] and the references listed therein. The most promising is the fractional FR (FFR), which is adopted by 3GPP LTE [17].

The performance of FFR has been extensively studied for traditional cellular networks [18], [19], [20]. For C-RAN architecture, a dynamic FR scheme based on FFR is proposed in [21]. The wireless resources are assigned to cell zones using a graph-coloring approach. Each color represents a certain segment of bandwidth. To minimize ICI, different colors should be assigned to any interfering zones.

A dynamic interference coordination scheme for downlink multi-cell systems is presented in [22]. The allocation problem is divided into two sub-problems, one at the BS level and the other at the central controller. It is assumed that BSs are able to communicate with each other using an X2 interface. At the BS level, each sector potentially allocates bandwidth chunks to its connected users. Then, each sector sends a request to the central controller. The request specifies a list of bandwidth chunks to be restricted at the dominant interfering zones, conflicting requests are resolved by the central controller, which sends a refined list of chunks that should be restricted to each sector.

Minimizing network power consumption of C-RAN is investigated in [23], where the power consumption of the transport network and RRHs is considered. The authors assume that transport links and RRHs can support sleep mode. The problem is formulated as a joint RRH selection and power minimization beamforming problem. The network power consumption is reduced by minimizing the number of active RRHs and reducing their transmit power subject to QoS constraints. Through simulations, the authors show that the network power consumption can be notably reduced. The performance of CoMP transmission schemes in a C-RAN architecture for LTE-A Heterogeneous networks is studied in [24]. With C-RAN architecture, a larger number of RRH can be considered in CoMP transmission, which improves the transmission performance.

The main objective of this paper is to propose a new model that virtualize the wireless resources of cloud-based RANs, then efficiently share them between multiple MNOs’ users. The main contributions of this paper are:

- An optimal model is proposed to enable sharing wireless resources between multiple MNOs and RRHs. The model considers many aspects, including ICI management, the fair distribution of resources between RRHs based on their traffic loads, a high level of isolation between MNOs, the ability for different MNOs to apply different and customizable resource scheduling policies, and efficient resource utilization across the network.
- Low-complexity suboptimal solutions are provided. The time complexity of the suboptimal solutions is considerably lower than the optimal while their throughput and delay performance are comparable. The suboptimal solutions are obtained by dividing the wireless resource allocation problem into sub-problems. The objective of each sub-problem is to allocate one RB to a set of non-interfering RRHs. The allocation per single RB is formulated as a maximum weighted independent set problem, which is solved using binary integer programming (BIP) solvers. In addition, a low-complexity heuristic algorithm that solves the BIP problem is proposed. The performance of each suboptimal solution is compared with the optimal solution as well as with a static sharing solution.

### III. System Model

Consider the downlink of a cloud-based RAN architecture shared between $M$ MNOs, where $N$ RRHs are distributed to cover a certain geographical area of interest. The RRHs are connected to a pool of BPUs in remote data centers via transport networks such as optical transport networks. It is assumed that the wireless link quality and expected traffic load are known for each user at the data center, which is responsible for resource allocation decisions. We assume that orthogonal frequency-division multiple access (OFDMA) is used for the downlink transmission. The total number of RBs available in the network is $R$. Each RRH is assumed to be capable of transmitting over any RB. The total number of user equipments (UEs) served by the network is $K$; each UE communicates with a single RRH.

Without loss of generality, we assume that each UE connects to the nearest RRH, and is labeled by a unique index $k \in [1, 2, \ldots, K]$. A table that maps each UE to a particular MNO and an RRH is assumed to be available for BPUs. To facilitate readability, Table I summarizes the notations frequently used throughout the paper.

The number of bits that can be transmitted over one RB is determined by the signal-to-interference plus noise ratio
Radio resources are assumed to be dynamically granted to MNOs based on a contract signed between MNOs; the contract specifies the resource percentage each MNO obtains in every possible load scenario. As wireless resource virtualization is still in its infancy stage, no well-defined sharing models exist yet [15]. Therefore, a general sharing model is assumed such that

\[ \Psi_1 : \Psi_2 : \cdots : \Psi_M = \delta_1 : \delta_2 : \cdots : \delta_M \]

where \( \Psi_m \) is the RB access probability for MNO \( m \) across all RRHs, and \( \sum_{m=1}^{M} \delta_m = 1, \quad 0 \leq \delta_m \leq 1, \quad \forall m \).

In the case of static sharing, MNOs are assumed to distribute their resources such that the frequency reuse factor is maximized while maintaining a proportional fairness criterion such that

\[ \max \sum_{m \in M_n} \Lambda_{m,n} \]

subject to

\[ \Lambda_{m,n} + \sum_{c \in C_n} \Lambda_{m,c} \leq R_m, \quad \forall n \]

\[ \Lambda_{m,1} : \cdots : \Lambda_{m,N_m} = (1 + \alpha)(L_{m,1} : \cdots : L_{m,N_m}) \]

where \( \Lambda_{m,n} \) is number of RBs allocated to MNO \( m \) at RRH \( n \), \( L_{m,n} \) is the load of MNO \( m \) at RRH \( n \), \( N_m \) is the set of RRHs that serves the UEs subscribed to MNO \( m \), and \( \alpha \) is a small constant that relaxes the fairness constraints in (4c) to ensure feasible solutions for the optimization problem. The load can be considered as the number of users or a number of packets queued in buffers for the users. As the fluctuation rate of the load in RRHs is slow compared with the transmission time interval (TTI), which is 1ms in LTE systems, the optimization problem in (4) can be solved at a coarser granularity than TTI. As a matter of fact, other sharing models can be applied here, however, maximizing the frequency reuse factor while considering a fairness criterion is an intuitive target that MNOs...
are looking to achieve.

IV. PROBLEM FORMULATION

In this work, each MNO aims at maximizing its sum weighted data rates, which is a very common optimization problem in wireless systems [22], [28], [29], [30]. The weights are selected by MNOs according to their scheduling policies. Assume that user $k$ is connected to RRH $n$, the scheduling problem can be formulated as

$$
\max_{m=1}^{M} \sum_{n \in \mathcal{N}_m} \left[ \sum_{k \in \mathcal{K}_m} \sum_{r=1}^{R} \bar{w}_{k} u_{r,k} \beta_{r,k} \right] \quad (5a)
$$

subject to

$$
\sum_{c \in \mathcal{C}_n} \left[ \sum_{k \in \mathcal{K}_c} \beta_{r,k} \right] + \sum_{k \in \mathcal{K}_n} \beta_{r,k} \leq 1, \forall n, r \quad (5b)
$$

$$
\sum_{r \in \mathcal{R}_k} T_{r,k} \leq q_k, \forall n, k \quad (5c)
$$

$$\left( \Phi_{m,n} > \Phi_{m}^{kh} \right) \text{ or } \left( \Psi_{m,n} \geq \Lambda_{m,n} \right) \text{ must hold, } \forall (m,n) \quad (5d)
$$

where $u_{r,k}$ is the data rate achieved by assigning RB $r$ to UE $k$, $\bar{w}_{k}$ is the normalized weight for UE $k$, $\mathcal{K}_m = \bigcup_{n=1}^{N} \mathcal{K}_{m,n}$ is the set of UEs connected to RRH $n$, $\mathcal{K}_{m,n}$ is the set of UEs subscribed to MNO $m$ and connect to RRH $n$, $\Phi_{m,n}$ is the number of RBs accessed by MNO $m$ at RRH $n$, $R_k$ is the set of RBs assigned to UE $k$, $\Phi_{m,n}$ and $\Phi_{m,n}^{kh}$ are the service status and service status threshold of MNO $m$ at RRH $n$, and $\beta_{r,k}$ is a binary number indicator defined as

$$
\beta_{r,k} = \begin{cases} 
1, & \text{if RB } r \text{ is assigned to UE } k \\
0, & \text{otherwise.}
\end{cases}
$$

Constraint (5b) represents the exclusive constraint which ensures that (i) each RB is assigned to one UE (at most) at each RRH, and (ii) orthogonal sets of RBs are allocated to RRHs that may interfere with each other. It is assumed that the interference is avoided if interfering RRHs are granted orthogonal sets of RBs. Constraint (5c) ensures that the transport block size for every UE is less than its unserved data size, where $\mathcal{R}_k$ is the RB set that is assigned to user $k$. Constraint (5d) specifies whether the service status of MNO $m$ at RRH $n$ is higher than a certain threshold or, if that is not the case, MNO $m$ should access at least $\Lambda_{m,n}$ RBs. This constraint ensures isolation between MNOs such that MNOs are either satisfied, or can access at least the same number of RBs in case of static sharing. It is noteworthy that constraint (5d) can be split into two constraints by introducing a binary variable $y_{m,n}$ and a sufficiently large upper bound $B_m$ so that

$$\Phi_{m,n} > \Phi_{m}^{kh} - B_m y_{m,n} \quad (6a)$$

$$\Psi_{m,n} \geq \Lambda_{m,n} - B_m (1 - y_{m,n}) \quad (6b)$$

When $y_{m,n} = 0$, constraint (6a) holds, whereas constraint (6b) becomes $\Psi_{m,n} \geq \Lambda_{m,n} - B_m$, which is always satisfied if $B_m$ is large enough. Note that the constraint $\Psi_{m,n} \geq \Lambda_{m,n}$ may still be satisfied when $y_{m,n} = 1$, only constraint (6b) holds. Consequently, one constraint holds, and the other one may be satisfied.

The formulation in (5) allows MNOs to apply different scheduling policies by weighting their UEs differently. In addition, it guarantees that MNOs use their share of RBs at the overloaded RRH. However, if an MNO is underloaded at a specific RRH, its share of RBs can be granted to other MNOs that are overloaded.

A. Radio Resource Scheduling Policies

In wireless networks, radio resource scheduling plays a vital role in achieving maximum spectrum utilization, QoS satisfaction, and fairness between UEs. To achieve such goals, users are weighted according to the UEs service status and the scheduling policy. The service status of a user can be related to aspects such as queue length, traffic priority, and past performance levels achieved. The product of the weight and the data rate achieved using a specific RB interprets the users priority of using the RB. It is worth mentioning that the users’ weights may vary with time.

Various scheduling policies are proposed for LTE networks [31], including channel-aware policies, such as Maximum Throughput (MT), Proportional Fair (PF), and Generalized PF (GPF); channel-aware and QoS-aware policies, such as Modified Largest Weighted Delay First (M-LWDF) and LOG rule; and energy-aware policies [32]. Table III illustrates examples of scheduling policies along with their types, targets, and weight definitions.

B. Complexity of BIP Solution

The scheduling problem given in (5) is a BIP optimization problem, which is NP-hard. The complexity of solving such optimization problems is considerably high and grows exponentially with number of users, MNOs, and RBs. Therefore, obtaining the optimal solution is computationally prohibitive even for a single MNO [22], [26].

In order to give a glimpse of the complexity of the BIP problem, one part of the problem is considered, which involves assigning RBs to RRHs. Consider a small scenario where two MNOs share three RRHs that interfere with each other. Assume that the total number of RBs in the network is 60. As the RRHs interfere with each other, orthogonal sets of RBs should be assigned to RRHs. If each RRH receives 20 RBs, the number of combinations of 20 RBs chosen from 60 RBs for one MNO is

$$\frac{60!}{20!(60-20)!} = 4.1918 \times 10^{15}.$$

For example, if the finest scheduling granularity is chosen, which is one subframe (1ms) for LTE systems, the optimization problem should be solved in less than 1ms; which is practically infeasible. Therefore, low-complexity algorithms should be used as alternatives for solving the problem.
TABLE III. EXAMPLES OF DIFFERENT SCHEDULERS [31]

| Scheduler | Type                  | Target                  | Weight                                                                 |
|-----------|-----------------------|-------------------------|------------------------------------------------------------------------|
| MT        | channel-aware, QoS-unaware | fairness & throughput   | $w_k = 1/T_k$, where $T_k$ is the past average throughput for UE $k$. |
| FF        | channel-aware, QoS-unaware | fairness & throughput   | $w_k = (u_k)\eta_1/(T_k)^{\eta_2}$, where $\eta_1$, $\eta_2$ are constants tuned by the scheduler. |
| GPF       | channel-aware, QoS-unaware | fairness & throughput   | $w_k = -D_k log(\alpha_k)/D_k^{max}$, where $\alpha_k$ is the acceptable delay violation probability, $D_k^{max}$ is delay threshold, and $D_k$ is the head-of-line packet delay. |
| M-LWDF    | channel-aware, QoS-aware | queue stability         | $w_k = b_k log(c_k + \alpha_k D_k)$, where $\alpha_k$, $b_k$, and $c_k$ are constants tuned by the scheduler. |
| LOG rule  | channel-aware, QoS-aware | queue stability         | $w_k = b_k log(c_k + \alpha_k D_k)$, where $\alpha_k$, $b_k$, and $c_k$ are constants tuned by the scheduler. |

V. LOW-COMPLEXITY SOLUTIONS

The optimal solution in (5) is achieved by jointly allocating RBs to users subscribed to different MNOs across all the available RRHs. However, it is computationally expensive to take all RBs in the scheduling problem into consideration, as seen in Section IV-B. A possible approach to reduce the complexity is to allocate RBs sequentially. In this section, two iterative low-complexity solutions are proposed, each of which allocates a single RB at each iteration. The basic concept is that, if RB $r$ is assigned to RRH $n$, it should be assigned to a user who is subscribed to the least satisfied MNO (LSM). As the objective is to maximize weighted sum utility, the assigned RB to the RRH is granted to the user who can maximize the sum-utility. To allocate RB $r$, the set of LSM at every RRH is denoted by $1 \times N$ vector $s_r = [s_{1,r}, \cdots, s_{N,r}]$, where $s_{n,r}$ is the index of the LSM at RRH $n$. The vectors $z_r = [z_1, \cdots, z_{N,r}]$ and $j_r = [j_{1,r}, \cdots, j_{N,r}]$ denote the utilities and indices, respectively, of the UEs subscribed to $s_r$ and who have maximum utility, where

$$z_{n,r} = \max_{k \in [s_{n,r}]} u_{k,r}$$

and $u_{k,r}$ is the utility of UE $k$.

The optimization problem per RB can be seen as the maximum weighted independent set (MWIS) of a graph path. Each RRH represents a vertex; an edge (line) is drawn between two vertices if they interfere with each other. A graph can be described by the pair $G = (V, E)$, where the set $V$ is the vertices of $G$, and the set $E$ is the edges of $G$. The MWIS is the subset of vertices that has maximum weighted sum such that no two vertices are connected with an edge. Fig. 2 shows an example of a graph and its MWIS.

The MWIS $I_S$ can be found by solving the following optimization problem

$$I_S = \max \sum_{n=1}^{N} u_n \beta_n$$  \hspace{1cm} (7a)

subject to

$$\sum_{c \in E_n} \beta_c \leq 1, \forall n$$  \hspace{1cm} (7b)

where $\beta_b$ is a binary variable, equal to one if $n \in I_S$ and zero otherwise, and $v_{n,r}$ is the weight of vertex (RRH) $n$. In order to bias the scheduler towards allocating RBs in favour of highly loaded RRHs, the weights are chosen such that

$$v_{n,r} = \begin{cases} z_{n,r} exp(\Omega_{s_{n,n}}), & \text{if } \Phi_{s_{n,n}} < \Phi^{thr} \\ 0, & \text{otherwise.} \end{cases}$$

where $\Omega_{m,n}$ is the number of RBs required by MNO $m$ at RRH $n$. Table (IV) shows the pseudo code of an iterative low-complexity algorithm that solves (7) by using a BIP solver. At each iteration, one RB is assigned to the MWIS of RRHs that maximizes the sum-weighted utility. Lines 5-8 find the LSM ($s_n$) at every RRH. To avoid assigning RBs to MNOs that have no data to transmit, the service status of the LSM has to be lower than the threshold $\Phi_{m,n} < \Phi^{thr}$. Lines 9-10 find the index and the utility value of the user who subscribes to MNO $s_n$ and maximizes the sum utility. In line 12, the algorithm solves the MWIS optimization problem (7) and finds the subset $I_S$. The RB is assigned to users in line 14. The number of RBs required is decremented for each RRH that
TABLE IV. PER RB OPTIMAL ALLOCATION ALGORITHM

1: input: \( \Lambda_{m,n}, \bar{\Omega}_{m,n}, \forall m,n \)
2: \( R_n = R, \forall n \)
3: for \( r = 1 : R \) do
4: for \( n = 1 : N \) do
5: \( v_{n,r} = 0 \)
6: \( \Delta \Omega_{m,n} = \Lambda_{m,n} - \bar{\Omega}_{m,n}, \forall m \)
7: \( \Omega_{m,n} = \Lambda_{m,n} + \Delta \Omega_{m,n}, \forall m \)
8: \( s_n = \arg \max \Omega_{m,n}, \forall m \) such that \( \Phi_{m,n} < \Phi^{th}_m \)
9: \( z_{n,r} = \max_{k \in \mathcal{K}_{s_n,n}} u_{k,r} \)
10: \( j_{n,r} = \arg \max_{k \in \mathcal{K}_{s_n,n}} u_{k,r} \)
11: end for
12: solve the BIP problem in (7)
13: assign RB \( r \) to UE \( j_{n,r}, \forall n \in I_S \)
14: update \( \bar{\Omega}_{s_n,n} \) and \( \Phi_{s_n,n}, \forall n \in I_S \)
15: end for
16: end for

belongs to \( I_S \) in line 13, whereas \( \bar{\Omega}_{s_n,n}, \Phi_{s_n,n} \) are updated in line 20. The algorithm runs until all RBs have been assigned.

Although the algorithm solves the BIP problem \( R \) times each TTI, the complexity of the algorithm is relatively low as compared to (5) because the size of the BIP is small. In particular, the BIP has \( N \) decision variables. However, for a large number of RRHs, it might be computationally expensive to solve the BIP problem shown in (7) \( R \) times for each TTI. Therefore, a low-complexity heuristic algorithm that solves the BIP is presented in Table V. The heuristic algorithm is greedy in the sense that its assigns an RB to the RRH that has the maximum weight \( v_{n,r} \) then excludes its interfering RRHs from the allocation process. The first 11 lines in the heuristic algorithm are similar to the algorithm shown in Table IV, where the LSMs and their users who maximize the sum utility are specified. The RRH index that has the maximum weighted utility \( n^* \) is found in line 14 and is added to the subset \( I_S \) in line 15. The RRH \( n^* \) and its interfering RRHs indices are deleted from the potential set of RRHs \( S_{ind} \). Consequently, interfering RRHs are not assigned assigned the same RBs, thereby eliminating interference. The RB is assigned to \( j_{n,r}, \forall n \in I_S \) in line 18, and \( \Omega_{s_n,n}, \bar{\Omega}_{s_n,n}, \Phi_{s_n,n} \) are updated in lines 19-20.

A. The complexity of the heuristic algorithm

For every TTI, the heuristic algorithm runs \( R \) major iterations (lines 3-21). Each major iteration finds the LSM and a candidate user \((j_{n,r})\) for each RRH. Finding the LSM at RRH requires \( M \) operations, whereas finding the user \((j_{n,r})\) requires \( K_{s_n,n} \) operations. The number of MNOs is usually much larger than the number of UEs, which makes finding \( j_{n,r} \) the dominating operation. Assigning each RB to the subset \( I_S \) requires at most \( N \) operations, assuming that no RRHs interfere with each other. Therefore, the worst-case complexity is \( O(R \times (N + K^{max})) \), where \( K^{max} = \max_{n,m} K_{n,m,n} \) is the maximum number of users that connect to an RRH and subscribe to one MNO.

TABLE V. HEURISTIC ALGORITHM

1: input: \( \Lambda_{m,n}, \bar{\Omega}_{m,n}, \forall m,n \)
2: \( R_n = R, \forall n \)
3: for \( r = 1 : R \) do
4: for \( n = 1 : N \) do
5: \( v_{n,r} = 0 \)
6: \( \Delta \Omega_{m,n} = \Lambda_{m,n} - \bar{\Omega}_{m,n}, \forall m \)
7: \( \Omega_{m,n} = \Lambda_{m,n} + \Delta \Omega_{m,n}, \forall m \)
8: \( s_n = \arg \max \Omega_{m,n}, \forall m \) such that \( \Phi_{m,n} < \Phi^{th}_m \)
9: \( z_{n,r} = \max_{k \in \mathcal{K}_{s_n,n}} u_{k,r} \)
10: \( j_{n,r} = \arg \max_{k \in \mathcal{K}_{s_n,n}} u_{k,r} \)
11: end for
12: assign RB \( r \) to UE \( j_{n,r}, \forall n \in I_S \)
13: update \( \bar{\Omega}_{s_n,n} \) and \( \Phi_{s_n,n}, \forall n \in I_S \)
14: end for

VI. SIMULATION MODEL AND NUMERICAL RESULTS

A layout made of a total of 22 hexagonal cells, as shown in Fig. 3, is considered. Two scenarios are considered; small-size and large-size. In the small-size scenario, only RRHs 1, 3 and 4 are considered. Comparisons are shown for the optimal, BIP, heuristic, and static sharing schemes. The large-size scenario considers the 22 RRHs. Because of the high complexity of the optimal solution for the large-scale scenario, results are shown only for BIP, heuristic, and static sharing schemes. UEs are distributed uniformly across RRHs such that the same number of UEs are connected to each RRH. The channel is modeled as a quasi-static frequency-flat Rayleigh fading channels. Each channel is assumed to be fixed during one subframe in time domain and over one RB in frequency domain, but changes independently over different subframes, different RBs, and different users.

A. Experiment 1

In this experiment, throughput and complexity comparisons are shown for the small-size scenario. Three RRHs are considered each of which serves the same number of UEs. The network is shared between two MNOs, each of which owns 18 RBs and serves a total number of 9 UEs. The MT switching policy is assumed for both MNOs. To show the worst-case scenario, a backlogged traffic model is assumed: users always have data to transmit, and the probability that one MNO is underloaded is zero. Therefore, MNOs receive the same number of RBs as in the static sharing scenario. The service status thresholds are \( \Phi_{m,n} = \infty, \forall m,n \).

Fig. 4 compares the average aggregate throughput of the MNOs for different average SINRs. As MNOs undergo similar
conditions in terms of the number of UEs, the average SINR, the traffic model, the scheduling policy, and the number of RBs, their performance is similar. Therefore, showing aggregate throughput of both MNOs is a sufficient performance indicator. As average SINR increases, aggregate throughput increases for all solutions. The optimal scheme outperforms the other solutions. The BIP and the heuristic solutions are comparable, performing only 6% lower than the optimal solution. The performance of the static sharing solution is the worst, 20% lower than the optimal solution. In the static sharing scheme, each RRH can only use a certain part of the spectrum, which lowers the chances of assigning an RB with a high SINR for a large number of UEs. In contrast, in other three schemes RRHs have access to all parts of spectrum.

The normalized running times of the four solutions are compared in Fig. 5. The optimal solution takes $1.9386 \times 10^9$ times longer than the static sharing solution. The running time for the suboptimal solutions (BIP and heuristic) is considerably lower than the optimal solution. The BIP scheme takes about 43.7 times more to solve the assignment problem compared to the heuristic scheme as it requires solving the BIP optimization problem. It is worth mentioning that the complexity of the optimal solution exponentially increases with the number of UEs, RBs, and RRHs. However, the complexity of the BIP problem involved in the BIP scheme exponentially increases only with the number of RRHs; its complexity increases linearly with the number of RBs, and does not depend on the number of UEs.

\section*{B. Experiment 2}

The whole layout, which consists of 22 RRHs, is considered in this experiment. The performance of the suboptimal and the static sharing solutions is compared for the worst-case scenario, which assumes a backlogged traffic model. Complexity, average throughput, and throughput for the worst and best users are shown for different numbers of UEs per RRH. Both MNOs apply the MT scheduling policy. The average SINR is set to 10 dB for all users.

Since both MNOs have similar profiles in terms of the number of RBs, scheduling policy, number of UEs, and average SINR, an average result for both MNOs is shown. Fig. 6 illustrates the average throughput per UE. As the number of RBs is fixed, the competition for resources grows as the number of UEs increases. Consequently, the average number of RBs assigned to any single user is lower. Therefore, the average throughput per UE decreases as the number of UEs increases in the network for all schemes. In this case, the BIP scheme outperforms the heuristic scheme by only 3%. A performance degradation of 20% is noticed for the static sharing scheme as compared to the BIP scheme. As the worst-case scenario is shown, using WRV adds an extra 20%
throughput in this experiment.

The average aggregate throughput across the network is shown in Fig. 7. As the number of UEs increases, multiplexing gain improves, which increases the aggregate throughput of the network.

The average throughput for the worst and the best users are exhibited in Figs. 8 and 9, respectively. The worst and the best users have similar trend: the BIP solution performs slightly outperforms the heuristic scheme, but much better than the static sharing solution. The key point is that the proposed suboptimal schemes maintain a high level of isolation between users, RRHs, and MNOs, even though wireless resources are totally shared across the network.

The complexity of the schemes is measured and pointed out in Fig. 10, which is closely analogous to the small-size scenario. The static sharing scheme is the least complex solution, since the RB allocation is independent for each RRH. The heuristic solution is solved considerably slower than static sharing but faster than the BIP scheme.

C. Experiment 3

In contrast to experiments 1 and 2, which investigate network performance for the worst-case sharing scenario, experiment 3 studies the WRV gain in a non-saturated network, at least for one MNO. The simulation variables are similar to experiment 2 in terms of the average SINR, the number of RRHs, MNOs, UEs, and RBs. Nevertheless, MNO1 applies the M-LWDF scheduling policy, while MNO2 applies the MT scheduling policy. The data traffic for UEs subscribed to MNO1 are modelled by Poisson traffic model with an average packet arrival rate \( \lambda \) [22], [30], [33]. Fixed-sized packet of 4000 bytes arrives to each user’s buffer. UEs subscribed to MNO2 are assumed to be greedy and always have data to transmit.

Fig. 11 shows the average head-of-line (HoL) packet delay for different values of \( \lambda \). As the value of \( \lambda \) increases, the average data arrivals deceases. Consequently, MNO1 becomes further underloaded. The average HoL delay for the BIP and the heuristic schemes are similar, but much less than that for the static sharing scheme. The dynamic sharing schemes maintain an average delay which is 100 times less than the static scheme.

The average aggregate throughput for MNO1 and MNO2 is illustrated in Figs. 12 and 13, respectively. As the value of \( \lambda \) increases, the average aggregate throughput of MNO1 decreases because less data arrives to the UEs’ buffers. In the static sharing scenario, RBs assigned to MNO1 are not accessible by MNO2. Therefore, average aggregate throughput of MNO2 is not affected by the variation in the traffic load of MNO1. For the BIP and the heuristic schemes, the throughput of MNO2 grows as the load of MNO1 becomes lighter. Reducing the load of MNO1 allows MNO2 to access more
RBs that would be granted to MNO1 if it is overloaded.

To sum up, using the proposed BIP and heuristic schemes would improve the performance of the network; lower the HoL delay for MNOs’ users, and increase the aggregate throughput of MNO2. These improvements result from the cooperation between MNOs; in cases such that one MNO is overloaded, the unloaded MNOs help to serve users who subscribe to the overloaded MNO.

VII. CONCLUSION

In this paper, we presented wireless resource virtualization schemes for cloud-based radio access networks (C-RANs). The proposed schemes dynamically share wireless resources between multiple mobile network operators (MNOs). Optimal and suboptimal solutions are provided and compared with each other as well as with a static sharing scheme, where each MNO is assigned a fixed set of radio RBs. The optimal solution is formulated as a binary integer programming optimization problem, which is known to be computationally expensive. Consequently, to reduce the complexity of the optimal formulation, two low-complexity suboptimal schemes are derived. The performance of the suboptimal solutions is slightly lower than the optimal solution at the benefit of a significant lower running time. The performance of the proposed schemes are compared in terms of throughput, delay, and time complexity. The simulation results show that the proposed schemes outperform static sharing in terms of both aggregate throughput and delay.

REFERENCES

[1] U. Paul, A. Subramanian, M. Buddhikot, and S. Das, “Understanding traffic dynamics in cellular data networks,” in IEEE Int. Conf. Comput. Commun. (INFOCOM), Apr. 2011, pp. 882–890.
| Average packet arrival rate (λ) | Average aggregate throughput for MNO2 | BIP | Heuristic | No sharing |
|---------------------------------|--------------------------------------|-----|-----------|------------|
| 20                              | 800                                  | 26  | 32        | 30         |
| 22                              | 1000                                 | 24  | 38        | 34         |
| 24                              | 1200                                 | 28  | 36        | 32         |
| 26                              | 1400                                 | 32  | 30        | 38         |
| 28                              | 1600                                 | 34  | 34        | 28         |
| 30                              | 1800                                 | 36  | 36        | 36         |
| 32                              | 2000                                 | 38  | 38        | 38         |
| 34                              | 2200                                 | 40  | 40        | 40         |

Fig. 13. Average aggregate throughput of MNO2’s UEs.

[2] R. Mahinda, M. Khojastepour, H. Zhang, and S. Rangarajan, “Radio access network sharing in cellular networks,” in *IEEE Int. Conf. Netw. Protocols (ICNP)*, Oct. 2013, pp. 1–10.

[3] Auri Aittokallio, “Nokia unveils radio cloud architecture.” [Online]. Available: http://telecoms.com/399321/nokia-unveils-radio-cloud-architecture/

[4] “C-RAN: the road towards green radio access network, presentation,” China Mobile Research Institute, Arlington, VA, USA, Tech. Rep., 2011. [Online]. Available: http://labs.chinamobile.com/cran/wp-content/uploads/2014/06/20140613-C-RAN-WP-3.0.pdf

[5] A. Checko, H. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. Berger, and L. Dittmann, “Cloud ran for mobile networks-a technology overview,” *IEEE Commun. Surveys Tuts.*, vol. PP, no. 99, pp. 1–1, Sept. 2014.

[6] S. Namba, T. Matsunaka, T. Warabino, S. Kaneko, and Y. Kishi, “Colony-ran architecture for future cellular network,” in *Future Network Mobile Summit (FutureNet)*, 2012. July 2012, pp. 1–8.

[7] M. Madhavan, P. Gupta, and M. Chetlur, “Quantifying multiplexing gains in a wireless network cloud,” in *IEEE Int. Conf. Commun. Conf. (ICC)*, June 2012, pp. 3212–3216.

[8] R. Kokku, R. Mahinda, H. Zhang, and S. Rangarajan, “Cellslice: Cellular wireless resource slicing for active ran sharing,” in *Int. Conf. Commun. Syst. and Netw. (COMSNETS)*, Jan 2013.

[9] G. Song and Y. Li, “Cross-layer optimization for OFDM wireless networks-part I: theoretical framework,” *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 614–624, Mar. 2005.

[10] M. Kalil, A. Shami, and Y. Yinghua, “Wireless resources virtualization in LTE systems,” in *IEEE Int. Conf. Comput. Commun. (INFOCOM)*, Workshop Mobile Cloud Comput., May 2014.

[11] Y. Zaki, L. Zhao, C. Goerg, and A. Timm-Giel, “LTE wireless virtualization and spectrum management,” in *Wireless and Mobile Netw. Conf. (WMNC)*, Oct. 2010.

[12] L. Zhao, M. Li, Y. Zaki, A. Timm-Giel, and C. Gorg, “LTE virtualization: From theoretical gain to practical solution,” in *Int. Teletraffic Congr. (ITC)*, Sept. 2011.

[13] M. Li, L. Zhao, X. Li, X. Li, Y. Zaki, A. Timm-Giel, and C. Gorg, “Investigation of network virtualization and load balancing techniques in LTE networks,” in *IEEE Veh. Technol. Con. (VTC)*, May 2012.

[14] F. Fu and U. Kozat, “Stochastic game for wireless network virtualization,” *IEEE/ACM Trans. Netw.*, vol. 21, no. 1, pp. 84–97, Feb. 2013.

[15] R. Kokku, R. Mahinda, H. Zhang, and S. Rangarajan, “NVS: A substrate for virtualizing wireless resources in cellular networks,” *IEEE/ACM Trans. Netw.*, vol. 20, no. 5, pp. 1333–1346, Oct. 2012.

[16] A. Hanza, S. Khalifa, H. Hamza, and K. Elsayed, “A survey on inter-cell interference coordination techniques in OFDMA-based cellular networks,” *IEEE Commun. Surveys Tuts.*, vol. 15, no. 4, pp. 1642–1670, Nov. 2013.

[17] 3GPP TR 36.921 , “Home eNode B (HeNB) Radio Frequency (RF) requirements analysis,” 2011.

[18] S. Ali and V. Leung, “Dynamic frequency allocation in fractional frequency reused OFDMA networks,” *IEEE Trans. Wireless Commun.*, vol. 8, no. 8, pp. 4286–4295, Aug. 2009.

[19] G. Boudreau, J. Panicker, N. Guo, R. Chang, N. Wang, and S. Vrzić, “Interference coordination and cancellation for 4g networks,” *IEEE Commun. Magazine*, vol. 47, no. 4, pp. 74–81, Apr. 2009.

[20] T. Novlan, R. Ganti, A. Ghosh, and J. Andrews, “Analytical evaluation of fractional frequency reuse for OFDMA cellular networks,” *IEEE Trans. Wireless Commun.*, vol. 10, no. 12, pp. 4294–4305, Dec. 2011.

[21] K. Wang, M. Zhao, and W. Zhou, “Graph-based dynamic frequency reuse in cloud-ran,” in *IEEE Wireless Commun. and Netw. Conf. (WCNC)*, Apr. 2014, pp. 105–110.

[22] M. Rahman and H. Yanikomeroglu, “Enhancing cell-edge performance: a downlink dynamic interference avoidance scheme with inter-cell coordination,” *IEEE Trans. Wireless Commun.*, vol. 9, no. 4, pp. 1414–1425, Apr. 2010.

[23] Y. Shi, J. Zhang, and K. Letaief, “Group sparse beamforming for green cloud-ran,” *IEEE Trans. Wireless Commun.*, vol. 13, no. 5, pp. 2809–2823, May 2014.

[24] A. Davydov, G. Morozov, I. Bolotin, and A. Papathansasiou, “Evaluation of joint transmission CoMP in C-RAN based LTE-A HetNets with large coordination areas,” in *IEEE Globecom Workshops*, Dec. 2013, pp. 801–806.

[25] V. Chandrasekhar and J. Andrews, “Spectrum allocation in tiered cellular networks,” *IEEE Trans. Wireless Commun.*, vol. 57, no. 10, pp. 3059–3068, Oct. 2009.

[26] A. Leith, M.-S. Alouini, D. I. Kim, X. Shen, and Z. Wu, “Flexible proportional-rate scheduling for OFDMA system,” *IEEE Trans. Mobile Comput.*, vol. 12, no. 10, pp. 1907–1919, Oct. 2013.

[27] 3GPP/TS.36.213 11.0.0, “LTE; Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures,” 2012.

[28] G. Song and Y. Li, “Cross-layer optimization for ofdm wireless networks-part ii: algorithm development,” *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, pp. 625–634, Mar. 2005.

[29] G. Song, Y. Li, and L. Cimini, “Joint channel-and queue-aware scheduling for multiuser diversity in wireless OFDMA networks,” *IEEE Trans. Wireless Commun.*, vol. 57, no. 7, pp. 2109–2121, July 2009.

[30] M. Katoozian, K. Navaie, and H. Yanikomeroglu, “Utility-based adaptive radio resource allocation in OFDM wireless networks with traffic prioritization,” *IEEE Trans. Wireless Commun.*, vol. 8, no. 1, pp. 66–71, Jan 2009.

[31] F. Capozzi, G. Piro, L. Greco, G. Boggia, and P. Camarda, “Downlink packet scheduling in lte cellular networks: Key design issues and a survey,” *IEEE Commun. Surveys Tuts.*, vol. 15, no. 2, pp. 678–700, May 2013.

[32] M. Kalil, A. Shami, and A. Al-Dweik, “Qos-aware power-efficient scheduler for LTE uplink,” *IEEE Trans. Mobile Comput.*, vol. PP, no. 99, pp. 1–1, Oct. 2014.

[33] M. Kalil, A. Shami, A. Al-Dweik, and S. Muhaidat, “Low-complexity power-efficient schedulers for LTE uplink with delay-sensitive traffic,” *IEEE Trans. Veh. Technol.*, vol. PP, no. 99, pp. 1–1, Nov. 2014.