Traffic Density-Based RRH Selection for Power Saving in C-RAN

Wentao Zhao, *Student Member, IEEE*, and Shaowei Wang, *Senior Member, IEEE*

***Abstract*— Cloud radio access network (C-RAN) is deemed as a promising architecture to meet the exponentially increasing traffic demand in a mobile network. However, it needs a huge amount of transport capacity between the remote radio heads (RRHs) and the baseband unit pool. Though an optical transport network can be employed to support such a massive capacity, it always leads to enormous power consumption that is comparable with the transmission powers of the RRHs, which can significantly decrease the performance of the C-RAN if not well addressed. In this paper, we investigate the power saving issue in the C-RAN, where we attempt to get a power consumption tradeoff between the optical transport network and the RRHs. An RRH selection problem is formulated to achieve this goal, which is based on the traffic density of the service area. Our optimization task is to select a subset of the RRHs to minimize the total power consumption of the C-RAN while satisfying a series of network constraints, including the power and bandwidth budgets of the RRHs, the traffic demands of users, and the spectral efficiency. We develop an efficient local search algorithm, which includes three types of local improvement operations: "add," "open," and "close," to address the intractable optimization task. Numerical results indicate that our proposal can significantly reduce the power consumption of the C-RAN. Moreover, our proposed algorithm converges stably and quickly, indicating that it is promising for applications.**

***Index Terms*— C-RAN, green communications, power saving, RRH selection, traffic density.**

I. Introduction

IN THE evolution of mobile communication system, one of the most important challenges is to mitigate power consumption of the system that always grows rapidly as the increasing of mobile subscribers, which receives significant attention in recent years [1]–[4]. Densely deploying small cells in traditional mobile communication network with macro cells is deemed as a promising approach to meet this goal [5], [6]. By exploiting spatial spectrum reuse gain, the throughput of mobile communication system can be dramatically enhanced in this way. Moreover, since the distance from access points to users decreases, considerable power saving can be obtained. However, resource allocation and interference management rise as new challenges in the densely deployed small cells scenario, as well as the consequential increase of capital expenditure (CAPEX) and operating expense (OPEX) of the network.

Cloud radio access network (C-RAN) has been recently proposed as a novel network architecture to answer these challenges [7]. In the C-RAN, remote radio heads (RRHs) and baseband units (BBUs) are separated to fulfill different functions, where the BBUs are centralized into a pool to create a set of common computing resources for signal processing while the main task of the RRHs is the radio transmission and reception. The C-RAN enables efficient resource allocation and interference management, increases the flexibility in network upgrade, and reduces the cost of network deployment, operation and maintenance. Investigations have also demonstrated the cost effectiveness of the C-RAN based heterogeneous network [8]. Moreover, as the BBUs are co-located in a centralized pool, they can interact with each other with low-latency so that advanced signal processing methods, such as enhanced intercell interference coordination (eICIC) and coordinated multi-point (CoMP), are much easier to be implemented in the C-RAN to tackle the resource dispatch and interference management issues [9], [10].

Since RRHs and BBUs are usually deployed in different locations, huge amounts of raw data should be conveyed between the RRHs and the BBU pool in the C-RAN. In fact, the transport network of the C-RAN not only needs to support such huge transport capacity, but also to meet strict latency requirements. Passive optical network (PON) [11], which uses an optical line terminal (OLT) to connect a set of optical network units (ONUs) through a single optical fiber, can provide cost-effective links between the RRHs and the BBU pool. However, power consumption rises as a non-negligible issue in such an optical transport network [12]. Implementing sleep mode of the ONUs can save the transport network power significantly as investigated in [11]. Moreover, due to the variation of spatial traffic, it is also feasible to switch off part of the RRHs without deteriorating the QoSs of users. In brief, we can design an RRH selection scheme to minimize the required number of active RRHs and the corresponding ONUs to save the powers of the C-RAN. From a practical application point of view, relatively complex RRH selection algorithm can be implemented in the C-RAN since the centralized BBUs provide ample computation resources.

Manuscript received July 31, 2015; revised December 5, 2015 and April 14, 2016; accepted August 3, 2016. Date of publication August 15, 2016; date of current version December 29, 2016. This work was supported in part by the National Natural Science Foundation of China under Grant 61671233, in part by the Jiangsu Science Foundation under Grant BK20151389, and in part by the Open Research Fund of National Mobile Communications Research Laboratory under Grant 2016D08. *(Corresponding author: Shaowei Wang.)*

W. Zhao is with the School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China (e-mail: zhaowt@smail.nju.edu.cn).

S. Wang is with the School of Electronic Science and Engineering, Nanjing University, Nanjing 210023, China, and also with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China (e-mail: wangsw@nju.edu.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSAC.2016.2600414

The RRH selection in the C-RAN is similar to that of turning off part of the base station (BSs) in a traditional cellular system to save energy (including energy cost of power amplifiers, air conditioning etc.), which is termed as BS sleeping or BS switching-on/off approaches [13]–[16]. The general optimization objective of BS sleeping is to minimize the number of active BSs so as to reduce the total energy consumption of the system while satisfying necessary network constraints. In [13], the authors propose a distributed and dynamic switching-on/off approach through a network-impact parameter. In [14], the energy saving problem is formulated as a constrained graphical game and a decentralized iterative algorithm is developed to find the best generalized Nash equilibrium. Different from [13] and [14], centralized approaches are proposed in [15] and [16]. A BS sleeping strategy that considers important quality of service (QoS) metrics is investigated in [15] and a two-stage dynamic programming algorithm is introduced to tackle the formulated problem, where the state of each BS is determined firstly and a radio resource allocation procedure follows. In [16], two approximation algorithms are designed to minimize the number of active BSs while satisfying the rate requirements of users.

As for the C-RAN, note that putting some RRHs into sleep mode would inevitably make the remaining RRHs consume more powers since these active RRHs have to increase their transmission powers to guarantee the traffic demands of users which are served by the RRHs in sleep mode. Balancing power consumption between the transmission links of the RRHs and the transport network in the C-RAN is not investigated in existing works [13]–[16]. It is necessary to explore a trade-off between the transmission powers of the RRHs and the transport network power as discussed in [17] and [18]. In [17], the objective is to minimize the total power consumption of the C-RAN while satisfying the signal-to-interference-plus-noise ratio (SINR) requirement of each user. A high complexity greedy algorithm is proposed, where the RRH that can achieve the maximum power reduction at each iteration is switched off. Two algorithms with relative low complexity are also presented. In [18], joint user association and beamforming design are investigated, where both downlink and uplink transmissions are taken into consideration. A virtual downlink transmission is established to convert the original problem into an equivalent form which can be solved by group-sparse optimization and relaxed-integer programming.

The RRH selection strategies proposed in [17] and [18] are based on the SINR requirements of users. In practical network scenario, the RRHs may be switched on/off frequently due to the mobility of users if employing these methods. However, it is difficult if not impossible to adjust the states of the RRHs according to the instantaneous traffic snapshot of the system. On the other hand, we notice that the traffic density (i.e. traffic load per unit region) keeps approximatively invariable in a relative long time [19] even if the locations of users change in a short time. An RRH selection strategy considering the traffic density of the service area is more reasonable and preferable to that of only focusing on the SINR requirements of users. For the traffic density based RRH selection scheme, the key issue is how to jointly allocate the available power and spectrum resources to improve the performance of the C-RAN, which is not mentioned in [13]–[18].

In this paper, we study the traffic density based RRH selection problem, where power and bandwidth allocation is jointly considered. Our optimization task is to select a subset of RRHs in the C-RAN with the minimum power consumption while satisfying a series of network constraints, such as transmission power budget, spectrum limitation, traffic demand and spectral efficiency requirement. An efficient local search algorithm is developed, which can find promising solutions efficiently. The performance of our proposal is verified by numerical results. The main contributions of this work are summarized as follows:

- We formulate a general RRH selection optimization task to deal with the power saving issue in the C-RAN. We consider link gain, traffic density, power saving, bandwidth allocation and spectral efficiency in our system model, where the active RRHs should guarantee the system performance under general network scenario and worst-case one with spectrum and power resource budget. Theoretical analysis and numerical results show that our proposed RRH selection strategy can reduce the total power consumption of the C-RAN significantly.
- We study a feasibility problem to determine whether a set of active RRHs can satisfy the peak rate requirement of the service area or not. An equivalent form of the problem is derived by introducing a crucial indicator parameter. We employ a fast barrier method to address it by exploiting the special structure of Hessian matrix. It is shown that the storage and computational complexity can be substantially reduced by the fast barrier method compared to standard ones.
- We investigate the power and bandwidth allocation problem in the C-RAN, where the optimization objective is to minimize the total transmission power for a given set of RRHs while satisfying the average rate and spectral efficiency requirement under general network scenario. Based on the optimal values of the power and bandwidth allocation problem, we develop an effective local search algorithm for the RRH selection problem that is NP-hard [20]. Specifically, three types of local improvement operations, including "add", "open" and "close", are introduced to search locally optimal solutions.

The remainder of this paper is organized as follows. In Section II, network model and the RRH selection problem are illustrated. In Section III, our proposed algorithms are presented in detail. In Section IV, numerical results are given with discussions. Conclusions are drawn in Section V.

II. Network Model and Problem Formulation

A. Transport Network Model

Consider a region $\mathcal{D} \in \mathbf{R}^2$ served by a C-RAN as illustrated in Fig. 1. There are a set of RRHs denoted by $\mathcal{N} = \{1, 2, \ldots, N\}$ and a BBU pool. For RRH $n \in \mathcal{N}$, the maximum transmission power and the available bandwidth are p_n^{max} and b_n^{max}, respectively. To simplify symbols used in the paper, the transport link between RRH n and BBU pool is also denoted

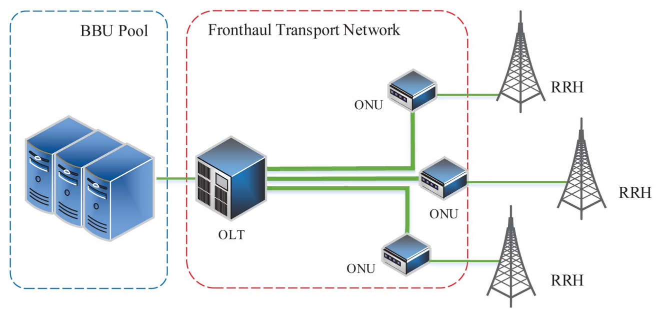

Fig. 1. Illustration of network model.

by n, and $\mathcal{N}$ is also the set of the transport links. We adopt the transport network power model discussed in [11] and [17], which can be written as follows:

$$P_{tn} = P_{olt} + \sum_{n \in \mathcal{N}} P_{tl,n}, \tag{1}$$

where P_{olt} is the power consumption of the OLT and $P_{tl,n}$ is the power consumption of transport link n, that is,

$$P_{tl,n} = \begin{cases} P_{tl,n}^{a} & \text{RRH } n \text{ is active,} \\ P_{tl,n}^{s} & \text{RRH } n \text{ is inactive.} \end{cases} \quad \forall n \in \mathcal{N}. \tag{2}$$

If we switch off RRH n, the power consumption saved on transport link n is

$$P_n = P_{tl,n}^{a} - P_{tl,n}^{s}. \tag{3}$$

Since $P_{tl,n}^{s}$ is generally much less than $P_{tl,n}^{a}$, we can put some RRHs into sleep mode to reduce the power consumption of the C-RAN. This is the motivation of this work.

Binary variable z_n indicates whether BBU n is active or not, that is,

$$z_n = \begin{cases} 1 & \text{RRH } n \text{ is active,} \\ 0 & \text{RRH } n \text{ is inactive.} \end{cases} \quad \forall n \in \mathcal{N}. \tag{4}$$

Then we can rewrite P_{tn} as follows:

$$\begin{aligned} P_{tn} &= P_{olt} + \sum_{n \in \mathcal{N}} (P_{tl,n}^{a} - P_{tl,n}^{s}) z_n + \sum_{n \in \mathcal{N}} P_{tl,n}^{s} \\ &= P_{olt} + \sum_{n \in \mathcal{N}} z_n P_n + \sum_{n \in \mathcal{N}} P_{tl,n}^{s}. \end{aligned} \tag{5}$$

Note that P_{olt} and $\sum_{n \in \mathcal{N}} P_{tl,n}^{s}$ are fixed for variables z_n.

B. Traffic Model

An important issue of our RRH selection scheme is how to meet the traffic demand with limited spectrum and power resources. Generally speaking, users keep on moving among different regions in the service area, e.g., residential and office, resulting that the traffic demand distribution of the service area varies throughout a day. However, the traffic density of the service area can keep approximatively invariable even in a long time span, e.g., several seconds or minutes. Assume that the traffic density of the area is fixed within a given time interval, it can be equivalent to the density of rate requirement (in bps/km^2). Then our RRH selection algorithm can be triggered at this time scale so that the RRH with light load can be switched off timely to reduce the power consumption of the system.

Let $x \in \mathcal{D}$ be a location in the area. We use $\zeta_a(x)$ and $\zeta_p(x)$ to represent the functions of the density of average rate requirements and the density of peak rate requirements, respectively. Assume that the total average rate requirement in the service area is R_D^{avg}, and the total peak rate requirement in the service area is R_D^{peak}, we have

$$\iint_{\mathcal{D}} \zeta_a(x) d\sigma = R_D^{avg}, \tag{6}$$

and

$$\iint_{\mathcal{D}} \zeta_p(x) d\sigma = R_D^{peak}. \tag{7}$$

Similar to the methods proposed in [21] and [22], we discretize the service area into K traffic demand areas (TDAs) denoted by $\mathcal{K}$. The service region of TDA k is $\mathcal{D}_k$. R_k^{avg} and R_k^{peak} are the average rate and the peak rate requirements of TDA k, where

$$R_k^{avg} = \iint_{\mathcal{D}_k} \zeta_a(x) d\sigma, \tag{8}$$

and

$$R_k^{peak} = \iint_{\mathcal{D}_k} \zeta_p(x) d\sigma. \tag{9}$$

Denote $h_{k,n}$ as the channel gain[1] between RRH n and TDA k. Define $b_{k,n}$ and $p_{k,n}$ as the bandwidth and power of RRH n allocated to TDA k, the achievable rate for RRH n to serve TDA k can be calculated as

$$r_{k,n} = b_{k,n} \log_2 \left[1 + \frac{p_{k,n} |h_{k,n}|^2}{b_{k,n}(N_0 + I_{k,n})} \right], \tag{10}$$

where N_0 and $I_{k,n}$ are noise power and the interference introduced by adjacent RRHs with unit bandwidth, respectively.

As discussed in [9] and [10], interference management techniques, such as CoMP and eICIC, can be employed to reduce intercell interference. Benefiting from the centralized computation resources in the BBU pool, these advanced signal processing techniques are much easier to be implemented in the C-RAN. Moreover, due to the low transmission power of RRH, large propagation loss of high frequency band and penetration loss, intercell interference could be slight without complex system design. On the other hand, for the case of densely deployed RRHs, clustering-based interference management can be employed [23]. In brief, we can put aside the inter-cell interference issue, that is, $I_{k,n} \approx 0$. The achievable rate from RRH n to TDA k is

$$r_{k,n} = b_{k,n} \log_2 \left(1 + \frac{p_{k,n} H_{k,n}}{b_{k,n}} \right), \tag{11}$$

where $H_{k,n} = |h_{k,n}|^2 / N_0$.

The frequently used notations is listed in Table I. For simplifying symbols, we also collect the variables z_n's, $b_{k,n}$'s, $p_{k,n}$'s and $r_{k,n}$'s into vectors $\bar{z}$, $\bar{b}$, $\bar{p}$ and $\bar{r}$, respectively.

[1]For simplicity, we take the power gain between RRH n and the center point of TDA k as the link gain $h_{k,n}$.

TABLE I
NOTATIONS

Symbol	Semantics
η_n	Drain efficiency of radio frequency power amplifier
$\zeta_a(x)$	Function of the density of average rate requirements
$\zeta_p(x)$	Function of the density of peak rate requirements
$b_{k,n}$	Bandwidth of RRH n allocated to TDA k
b_n^{max}	Total available bandwidth of RRH n
$\mathcal{D}_k$	Service region of TDA k
$h_{k,n}$	Channel gain between RRH n to TDA k
$I_{k,n}$	Interference throned by adjacent RRHs with unit bandwidth for TDA k and RRH n
$\mathcal{K}$	Set of TDAs
K	Number of TDAs
$\mathcal{N}$	The set of RRHs
N	Number of RRHs
N_0	Noise power with unit bandwidth
p_n^{max}	Maximum transmission power of RRH n
$p_{k,n}$	Power of RRH n allocated to TDA k
P_n	Power saved by switching off RRH n on the nth transport link
P_{olt}	Power consumption of OLT
$P_{tl,n}$	Power consumed by transport link n
$P_{tl,n}^a$	Power consumption of transport link n when RRH n is active
$P_{tl,n}^s$	Power consumption of transport link n when RRH n is unactive
$r_{k,n}$	Achievable rate from RRH n to TDA k
R_D^{avg}	Total average rate requirement of D
R_D^{peak}	Total peak rate requirement of D
R_k^{avg}	Average rate requirement of TDA k
R_k^{peak}	Peak rate requirement of TDA k
$S_{k,n}$	Spectral efficiency of RRH n to TDA k
S_k^{min}	Required spectral efficiency of TDA k
z_n	Selection variable of RRH n

C. Network Constraints

The following network constraints should be considered in the RRH selection procedure:

1) Power Budget: The transmission power of each RRH is limited, that is,

$$\sum_{k\in\mathcal{K}} p_{k,n} \le z_n p_n^{max}, \quad \forall n \in \mathcal{N}. \tag{12}$$

2) Bandwidth Budget: The allocated bandwidth of each RRH is also limited:

$$\sum_{k\in\mathcal{K}} b_{k,n} \le z_n b_n^{max}, \quad \forall n \in \mathcal{N}. \tag{13}$$

3) System Performance Guarantee: When we put some RRHs into sleep mode to reduce power consumption of the system, we also need to ensure that the remaining active RRHs can guarantee the system performance under different network scenarios.

General scenario: The selected RRHs should meet the average rate requirements of all TDAs, that is,

$$\sum_{n\in\mathcal{N}} r_{k,n} \ge R_k^{avg}, \quad \forall k \in \mathcal{K}. \tag{14}$$

It is also necessary to take spectral efficiency into consideration for the general scenario. Define spectral efficiency function $S_{k,n}$ for $k \in \mathcal{K}, n \in \mathcal{N}$ as follows:

$$S_{k,n} = \frac{r_{k,n}}{b_{k,n}} = \log_2\left(1 + \frac{p_{k,n}H_{k,n}}{b_{k,n}}\right). \tag{15}$$

A minimum spectral efficiency requirement, denoted by S_k^{min}, is required when RRH n is assigned to serve TDA n. If $b_{k,n} > 0$, it requires $S_{k,n} \ge S_k^{min}$. With simple mathematical operations, spectral efficiency constraint can be rewritten as follows:

$$p_{k,n} \ge \Delta_{k,n} b_{k,n}, \tag{16}$$

where $\Delta_{k,n} = (2^{S_k^{min}} - 1)/H_{k,n}$. Note that we usually take the cell-edge user spectral efficiency as reference.

Worst-case scenario: Under this scenario, the active RRHs should satisfy the peak rate requirement of the considered region. The spectral efficiency of each TDA is not required because our priority is to meet the peak traffic load with acceptable performance loss in this scenario. In other words, we need to solve the following feasibility problem for a given $\bar{z}$:

$$\begin{aligned} &\text{find} \quad \bar{b}, \bar{p} \\ &s.t. \; C_1: \sum_{k\in\mathcal{K}} p_{k,n} \le z_n p_n^{max}, \quad \forall n \in \mathcal{N}, \\ &\quad\;\; C_2: \sum_{k\in\mathcal{K}} b_{k,n} \le z_n b_n^{max}, \quad \forall n \in \mathcal{N}, \\ &\quad\;\; C_3: \sum_{n\in\mathcal{N}} r_{k,n} = R_k^{peak}, \quad \forall k \in \mathcal{K}, \\ &\quad\;\; C_4: \bar{b} \in \mathbf{R}_+^{KN}, \; \bar{p} \in \mathbf{R}_+^{KN}. \end{aligned} \tag{17}$$

We define

$$\mathbf{Z} = \{\bar{z} | z_n \in \{0, 1\}, \quad \forall n, (17) \text{ is feasible with } \bar{z}\}. \tag{18}$$

D. Problem Formulation

As discussed above, the power consumption of the C-RAN involves two parts: transport network power consumption and the transmission power of the RRHs, which can be written as follows:

$$\begin{aligned} P_{total} &= P_{tn} + \sum_{n\in\mathcal{N}} \frac{z_n}{\eta_n} \sum_{k\in\mathcal{K}} p_{k,n} \\ &= \sum_{n\in\mathcal{N}} z_n P_n + \sum_{n\in\mathcal{N}} \frac{z_n}{\eta_n} \sum_{k\in\mathcal{K}} p_{k,n} + P_{fixed}, \end{aligned} \tag{19}$$

where $P_{fixed} = P_{olt} + \sum_{n\in\mathcal{N}} P_{tl,n}^s$ and η_n is the drain efficiency of radio frequency power amplifier.

The optimization objective of the RRH selection task is to select a subset of $\mathcal{N}$ to minimize the total power consumption of the C-RAN while satisfying the network constraints mentioned above. It can be mathematically formulated as follows:

$$\begin{aligned}
\underset{\bar{z},\bar{b},\bar{p}}{\text{minimize}} \quad & \sum_{n\in\mathcal{N}} P_n z_n + \sum_{n\in\mathcal{N}} \frac{1}{\eta_n} \sum_{k\in\mathcal{K}} p_{k,n} \\
s.t.\ C_1: \quad & \sum_{k\in\mathcal{K}} p_{k,n} \le z_n p_n^{max}, \quad \forall n \in \mathcal{N}, \\
C_2: \quad & \sum_{k\in\mathcal{K}} b_{k,n} \le z_n b_n^{max}, \quad \forall n \in \mathcal{N}, \\
C_3: \quad & \sum_{n\in\mathcal{N}} r_{k,n} \ge R_k^{avg}, \quad \forall k \in \mathcal{K}, \\
C_4: \quad & p_{k,n} \ge \Delta_{k,n} b_{k,n}, \quad \forall n \in \mathcal{N},\ k \in \mathcal{K}, \\
C_5: \quad & \bar{z} \in \mathbf{Z},\ \bar{b} \in \mathbf{R}_+^{KN},\ \bar{p} \in \mathbf{R}_+^{KN}.
\end{aligned} \tag{20}$$

P_{fixed} is dropped in the objective function since it is constant. In (20), C_1 and C_2 are the transmission power limits and available bandwidth budgets for the selected RRHs. C_3 and C_4 ensure the average rate demand and spectral efficiency requirement of each TDA. C_5 is intuitive.

Eq. (20) defines a mixed integer programming problem since it involves both binary variables z_n's and real variables $b_{k,n}$'s, $p_{k,n}$'s. Such a problem is generally NP-hard that is difficult to solve. It is worth noting that (20) is a generalized form of many classic optimization problems in cellular networks. If we only minimize the transport network power consumption, (20) is similar to the cost cell planning problem studied in [24]–[26]; if we ignore the transport network power consumption, (20) becomes a resource allocation problem in the cellular system [27], [28].

On the other hand, the RRH selection is also a generalized form of capacitated facility location problem [29], [30]. Some approximation algorithms are proposed to address this kind of problems. In [29], a multi-exchange local search algorithm is developed. Four types of local improvement operations, named as "add", "open", "close", "multi", are introduced. It is proved that the proposed algorithm can achieve an approximation ratio of $3 + 2\sqrt{2}$ to the optimal solution. Modified "open" and "close" operations are proposed in [30] to further improve the performance. It shows that the modified multi-exchange local search algorithm can improve the performance guarantee to 5.

III. Our Proposed Algorithm

To address the RRH selection problem, we answer the following questions at first: Given a subset of RRHs, is it possible for the RRHs to meet all the TDAs' peak rate requirements with given bandwidth and power budgets? Moreover, if it is possible, how can we minimize the transmission power with the given subset of RRHs while satisfying practical network constraints? After address these two problems, we design an efficient local search algorithm to solve the RRH selection problem.

A. Feasibility of Eq. (17)

Denote $\mathcal{N}_s = \{n|z_n = 1\}$ as the selected RRHs. A preparatory procedure for the RRH selection is necessary to obtain a feasible solution to (17) or to prove the feasible solution does not exist. It is equivalent to solve the following minimization problem by introducing a crucial indicator parameter ξ:

$$\begin{aligned}
\underset{\bar{b},\bar{p},\xi}{\text{minimize}} \quad & \xi \\
s.t.\ C_1: \quad & \sum_{k\in\mathcal{K}} p_{k,n} \le p_n^{max}, \quad \forall n \in \mathcal{N}_s, \\
C_2: \quad & \sum_{k\in\mathcal{K}} b_{k,n} \le b_n^{max} + \xi, \quad \forall n \in \mathcal{N}_s, \\
C_3: \quad & \sum_{n\in\mathcal{N}_s} r_{k,n} \ge R_k^{peak}, \quad \forall k \in \mathcal{K}, \\
C_4: \quad & \bar{b} \in \mathbf{R}_+^{KN_s},\ \bar{p} \in \mathbf{R}_+^{KN_s}.
\end{aligned} \tag{21}$$

Variable ξ can be treated as an upper bound of the maximum infeasibility of the inequalities, as can be observed from the constraints in C_2. If the optimal value is less than 0, (17) is obviously feasible; otherwise, no feasible solution to (17) exists.

Eq. (21) is a convex optimization problem and can be solved by standard convex optimization algorithm. Barrier method is one of these algorithms which has promising convergence rate [31]. The main disadvantage of the barrier method is that the cost of storing Hessian and computing Newton step is high. For the problem defined by (21), the storage complexity and computational complexity of standard barrier method are $O(K^2N_s^2)$ and $O(K^3N_s^3)$, respectively. We introduce a fast barrier method proposed in [27] and [28] to substantially reduce the storage and computational complexity.

It is worth noting that the Hessian matrix of Eq. (21)is singular. We can replace the objective function with a convex function which has same monotonicity in the feasible region to guarantee the Hessian matrix is non-singular. It is also worth noting that $\sum_{n\in\mathcal{N}_s} r_{k,n} = R_k^{peak}$ always holds since otherwise we can decrease the assigned bandwidth without violating the peak rate requirement constraints. We reform (21) as follows:

$$\begin{aligned}
\underset{\bar{b},\bar{r},\xi}{\text{minimize}} \quad & \xi^2 \\
s.t.\ C_1: \quad & \sum_{k\in\mathcal{K}} p_{k,n} \le p_n^{max}, \quad \forall n \in \mathcal{N}_s, \\
C_2: \quad & \sum_{k\in\mathcal{K}} b_{k,n} \le b_n^{max} + \xi, \quad \forall n \in \mathcal{N}_s, \\
C_3: \quad & \sum_{n\in\mathcal{N}_s} r_{k,n} = R_k^{peak}, \quad \forall k \in \mathcal{K}, \\
C_4: \quad & \bar{b} \in \mathbf{R}_+^{KN_s},\ \bar{r} \in \mathbf{R}_+^{KN_s}.
\end{aligned} \tag{22}$$

where

$$p_{k,n} = \tfrac{b_{k,n}}{H_{k,n}} \cdot \left(2^{r_{k,n}/b_{k,n}} - 1\right). \tag{23}$$

Note that variables are replaced by $\bar{b}, \bar{r}, \xi$ in (22). ξ^2 and ξ have different monotonicity in the interval of $(-\infty, 0)$. We can conclude that the solution to (17) exists if and only if the optimal value of (22) is 0.

We collect all variables into one vector $\bar{x} \in \mathbf{R}^{2KN_s}$,

$\bar{x} = (r_{1,1}, b_{1,1}, r_{1,2}, b_{1,2}, \ldots, r_{K,N_s}, b_{K,N_s}, \xi)^T$. Define

$$b_n = b_n^{max} + \xi - \sum_{k\in\mathcal{K}} b_{k,n},$$
$$p_n = p_n^{max} - \sum_{k\in\mathcal{K}} p_{k,n}, \quad \forall n \in \mathcal{N}_s. \tag{24}$$

Then we can convert all inequality constraints into a logarithmic barrier function $\phi(\bar{x})$:

$$\phi\bar{x} = -\sum_{n\in\mathcal{N}_s}\sum_{k\in\mathcal{K}} \log b_{k,n} - \sum_{n\in\mathcal{N}_s}\sum_{k\in\mathcal{K}} \log r_{k,n} - \sum_{n\in\mathcal{N}_s} \log b_n - \sum_{n\in\mathcal{N}_s} \log p_n. \tag{25}$$

We reform C_3 in (22) as $\Lambda\bar{x} = \bar{\varrho}$, where $\bar{\varrho} = (R_1^{peak}, \ldots, R_k^{peak})^T$ and $\Lambda \in \mathbf{R}^{K\times 2KN_s}$ with

$$\Lambda_{k,i} = \begin{cases} 1 & i = 2(k-1)N_s + 2n - 1, \quad \forall k, n, \\ 0 & \text{otherwise.} \end{cases} \tag{26}$$

Eq. (22) can be converted into a sequence of minimization problems by introducing a logarithmic barrier function with a parameter t. The optimal solution to (22) can be approximated by solving the following unconstrained convex problem:

$$\underset{\bar{x}}{\text{minimize}} \ \ \psi_t(\bar{x}) = t\xi^2 + \phi(\bar{x})$$
$$s.t.\ \Lambda\bar{x} = \bar{\varrho}. \tag{27}$$

Generally, Newton method is preferred to solve the convex problem with equality constraints due to its quadratic convergence property [31]. For a given parameter t, Newton step $\triangle\bar{x}$ can be calculated by solving the following equation:

$$\begin{bmatrix} \nabla^2\psi_t(\bar{x}) & \Lambda^T \\ \Lambda & \mathbf{0} \end{bmatrix} \begin{bmatrix} \triangle\bar{x} \\ \bar{\nu} \end{bmatrix} = \begin{bmatrix} -\nabla\psi_t(\bar{x}) \\ \mathbf{0} \end{bmatrix}. \tag{28}$$

$\nabla^2\psi_t(\bar{x})$ is the Hessian matrix, $\nabla\psi_t(\bar{x})$ is the gradient of $\psi_t(\bar{x})$, and $\bar{\nu}$ is the associated dual vector. The symbol $\mathbf{0}$ denotes a vector or a matrix all of whose components are 0. The Hessian of $\psi_t(\bar{x})$ can be written as

$$\nabla^2\psi_t(\bar{x}) = \Pi + \sum_{n\in\mathcal{N}_s} \frac{\nabla p_n \nabla p_n^T}{p_n^2} + \sum_{n\in\mathcal{N}_s} \frac{\nabla b_n \nabla b_n^T}{b_n^2}, \tag{29}$$

where $\Pi = diag(\Pi_{1,1}, \ldots, \Pi_{K,N_s}, t)$ with

$$\Pi_{k,n} = \begin{bmatrix} \frac{1}{p_n}\frac{\partial^2 p_{k,n}}{\partial r_{k,n}^2} + \frac{1}{r_{k,n}^2} & \frac{1}{p_n}\frac{\partial^2 p_{k,n}}{\partial r_{k,n}\partial b_{k,n}} \\ \frac{1}{p_n}\frac{\partial^2 p_{k,n}}{\partial b_{k,n}\partial r_{k,n}} & \frac{1}{p_n}\frac{\partial^2 p_{k,n}}{\partial b_{k,n}^2} + \frac{1}{b_{k,n}^2} \end{bmatrix}. \tag{30}$$

Denote

$$\bar{f}_i = \begin{cases} \nabla\psi_t(\bar{x}) & i = 0, \\ \nabla p_i / p_i & i = 1, \ldots, N_s, \\ \nabla b_{i-N_s}/b_{i-N_s} & i = N_s + 1, \ldots, 2N_s, \end{cases} \tag{31}$$

the Hessian of $\psi_t(\bar{x})$ can be rewritten as

$$\nabla^2\psi_t(\bar{x}) = \Pi + \sum_{i=1}^{2N_s} \tilde{f}_i. \tag{32}$$

Fact 1: Given a non-singular matrix $\Phi \in \mathbf{R}^{2KN_s\times 2KN_s}$, vectors $f, b \in \mathbf{R}^{2KN_s\times 1}$, where f satisfies $1 + f^T\Phi^{-1}f \neq 0$. Then, if $\Phi x = b, (\Phi + ff^T)\tilde{x} = b$, it always holds $\tilde{x} = x - \frac{f^T x}{1+f^T g} g$, where $g = \Phi^{-1} f, g \in \mathbf{R}^{2KN_s\times 1}$.

By using the mathematical fact, we can employ an iterative algorithm to calculate Newton step efficiently.

Step 1: Solve the following equations:

$$\begin{bmatrix} \Pi & \Lambda^T \\ \Lambda & \mathbf{0} \end{bmatrix} \tilde{x}_i = \tilde{f}_i, \quad \forall i = 0, 1, \ldots, 2N_s, \tag{33}$$

where $\tilde{x}_i = \begin{bmatrix} \bar{x}_i \\ \bar{\nu}_i \end{bmatrix}$, $\tilde{f}_i = \begin{bmatrix} \bar{f}_i \\ \mathbf{0} \end{bmatrix}$. Consider the ith equation, since Π is a diagonal matrix, we have

$$\Pi_{k,n} \begin{bmatrix} \bar{x}_i[\vartheta_1(k,n)] \\ \bar{x}_i[\vartheta_2(k,n)] \end{bmatrix} = \begin{bmatrix} \bar{f}_i[\vartheta_1(k,n)] - \bar{\nu}_i[k] \\ \bar{f}_i[\vartheta_2(k,n)] \end{bmatrix}, \quad \forall k, n, \tag{34}$$

$$\sum_{n=1}^{N_s} \bar{x}_i[\vartheta_1(k,n)] = 0, \quad \forall k, \tag{35}$$

and

$$\bar{x}_i[2KN_s + 1] = \bar{f}_i[2KN_s + 1]/t, \tag{36}$$

where $\vartheta_1(k,n) = 2(k-1)N_s + 2n - 1$, $\vartheta_2(k,n) = \vartheta_1(k,n) + 1$, $\bar{f}[j]$ denotes the jth component of the vector $\bar{f}$. According to (34), we can obtain

$$\bar{x}_i[\vartheta_1(k,n)] = \Pi_{k,n}^{-1}[1,1]\Big(\bar{f}_i[\vartheta_1(k,n)] - \bar{\nu}_i[k]\Big) + \Pi_{k,n}^{-1}[1,2]\bar{f}_i[\vartheta_2(k,n)] \tag{37}$$

and

$$\bar{x}_i[\vartheta_2(k,n)] = \Pi_{k,n}^{-1}[2,1]\bar{f}_i[\vartheta_1(k,n)] + \Pi_{k,n}^{-1}[2,2]\bar{f}_i[\vartheta_2(k,n)]. \tag{38}$$

Substituting (37) into (35), we have

$$\bar{\nu}_i[k] = \frac{1}{\sum_{n=1}^{N_s}\Pi_{k,n}^{-1}[1,1]} \sum_{n=1}^{N_s}\Big(\Pi_{k,n}^{-1}[1,1]\bar{f}_i[\vartheta_1(k,n)] + \Pi_{k,n}^{-1}[1,2]\bar{f}_i[\vartheta_2(k,n)]\Big). \tag{39}$$

In brief, we calculate $\bar{\nu}_i$ by using (39) in the first place and then obtain $\bar{x}_i$ by using (36)-(38).

Step 2: Update $\tilde{x}_0, \tilde{x}_1, \ldots, \tilde{x}_{2N_s-1}$ by solving the following equations:

$$\left(\begin{bmatrix} \Pi & \Lambda^T \\ \Lambda & \mathbf{0} \end{bmatrix} + \tilde{f}_{2N_s}\tilde{f}_{2N_s}^T\right)\tilde{x}_i = \tilde{f}_i, \quad \forall i = 0, 1, \ldots, 2N_s - 1. \tag{40}$$

We can update $\tilde{x}_i$'s efficiently based on Fact 1.

⋮

Step $\mathbf{2N_s + 1}$: Update $\tilde{x}_0$ that satisfies the following equation:

$$\left(\begin{bmatrix} \Pi & \Lambda^T \\ \Lambda & \mathbf{0} \end{bmatrix} + \sum_{i=1}^{2N_s}\tilde{f}_i\tilde{f}_i^T\right)\tilde{x}_0 = \tilde{f}_0. \tag{41}$$

TABLE II
BARRIER METHOD

Algorithm 1

Initialization: Feasible point $\bar{x}$
while $KN_s/t > \epsilon_b$
  while True
    Compute $\triangle\bar{x}$ by using (28) and $\lambda^2 = \nabla\psi_t(\bar{x})\triangle\bar{x}$;
    Set $s = 1$;
    if $\lambda^2/2 \leq \epsilon_n$
      break;
    end if
    while $\psi_t(\bar{x} + s\triangle\bar{x}) > \psi_t(\bar{x}) - \alpha s\lambda^2$
      $s = \beta s$;
    end while
    Update $\bar{x} = \bar{x} + s\triangle\bar{x}$;
  end while
  $t = \mu t$;
end while
return $\bar{x}$

$\bar{x}_0$ is Newton step.

The computational complexity of the fast barrier method is $O(KN_s^3)$. Moreover, the storage complexity is bounded by $O(KN_s)$ since only the diagonal elements of the Hessian, $2N_s$ vectors with $2K$ nonzero elements and the gradient need recording. Generally, $K \gg N_s$ generally holds in practical systems. Thus the fast barrier method can reduce the complexity and be widely applied to large-scale wireless system. The outline of the barrier method is summarized in Table II. ϵ_b and ϵ_n are the tolerances of the barrier method and the Newton method, respectively. α and β are two constants utilized in backtracking line search with $\alpha \in (0, 0.5)$ and $\beta \in (0, 1)$. The step size of the backtracking line search is s with $s > 0$. t and μ are parameters associated with a trade-off between outer iterations and inner iterations.

Define a boolean function $\mathcal{F}(N_s)$ for each $\mathcal{N}_s$: $\mathcal{F}(N_s)$ is *True* if the RRHs in $\mathcal{N}_s$ can satisfy the peak rate requirements; otherwise, $\mathcal{F}(N_s)$ is *False*. Denote ξ^* as the optimal indicator parameter for problem (22), we have

$$\mathcal{F}(N_s) = \begin{cases} Ture & \xi^* = 0, \\ False & \xi^* > 0, \end{cases} \quad \forall N_s \subseteq \mathcal{N}. \tag{42}$$

B. Local Search Algorithm for RRH Selection Problem

We generalize the methods in [29] and [30] to address the formulated RRH selection problem. First, we need to obtain the minimum transmission power of the RRHs in $\mathcal{N}_s$ to satisfy network constraints under general scenario. The power and bandwidth allocation problem is formulated as follows:

$$\begin{aligned} \underset{\bar{b},\bar{p}}{\text{minimize}} \quad & \sum_{n\in\mathcal{N}_s} \frac{1}{\eta_n} \sum_{k\in\mathcal{K}} p_{k,n} \\ s.t.\ C_1: \quad & \sum_{n\in\mathcal{N}_s} r_{k,n} \geq R_k^{avg}, \quad \forall k \in \mathcal{K}, \\ C_2: \quad & \sum_{k\in\mathcal{K}} p_{k,n} \leq p_n^{max}, \quad \forall n \in \mathcal{N}_s, \\ C_3: \quad & \sum_{k\in\mathcal{K}} b_{k,n} \leq b_n^{max}, \quad \forall n \in \mathcal{N}_s, \\ C_4: \quad & p_{k,n} \geq \Delta_{k,n} b_{k,n}, \quad \forall n \in \mathcal{N}_s,\ k \in \mathcal{K}, \\ C_5: \quad & \bar{b} \in \mathbf{R}_+^{KN_s},\ \bar{p} \in \mathbf{R}_+^{KN_s}. \end{aligned} \tag{43}$$

Eq. (43) has the same form of (22) so it can be solved by the fast barrier method in the same way. Define $P_r(\mathcal{N}_s)$ as the optimal value of (43), which also denotes the minimum transmission power of the RRHs in $\mathcal{N}_s$ to meet the average rate requirements. Specifically, if $\mathcal{N}_s$ is infeasible to (43), we set $P_r(\mathcal{N}_s) = +\infty$.

Define

$$P(\mathcal{N}_s) = P_t(\mathcal{N}_s) + P_r(\mathcal{N}_s). \tag{44}$$

as the minimum objective value of (20) with the given $\mathcal{N}_s$, where

$$P_t(\mathcal{N}_s) = \sum_{n\in\mathcal{N}_s} P_n. \tag{45}$$

Based on the optimal solution to (43), we design an effective local search algorithm to address the RRH selection problem. Starting with a feasible solution, e.g., $\mathcal{N}_s = \mathcal{N}$, we introduce the following three local improvement operations:

add(n): We select RRH n which is not in current solution $\mathcal{N}_s$, i.e. $n \notin \mathcal{N}_s$. If $P(\mathcal{N}_s \cup \{n\}) < P(\mathcal{N}_s)$, add RRH n to $\mathcal{N}_s$, i.e. $\mathcal{N}_s \leftarrow \mathcal{N}_s \cup \{n\}$.

open$(n, \mathcal{N}')$: We select RRH $n \notin \mathcal{N}_s$ and close a subset of RRHs $\mathcal{N}'$, i.e. $\mathcal{N}_s \leftarrow \mathcal{N}_s \cup \{n\} \backslash \mathcal{N}'$. Note that the possibilities for the set $\mathcal{N}'$ are exponential scale. To make the procedure terminated in polynomial time, "open" operation proposed in [29] and [30] does not compute the exact cost of the new solution but only an estimated cost that is always overestimated, that is, all the rate requirements served by $\mathcal{N}'$ is only reassigned to n. However, such a simplified operation cannot be applied to our formulated problem due to the spectral efficiency requirements. We propose a modified "open" operation based on the following theorem:

Theorem 1: If there exists a set of RRHs $\mathcal{N}' \subseteq \mathcal{N}_s$ that satisfies

$$P(\mathcal{N}_s \cup \{n\} \backslash \mathcal{N}') < P(\mathcal{N}_s \cup \{n\}), \tag{46}$$

there must exist an RRH $n' \in \mathcal{N}'$ that satisfies

$$P(\mathcal{N}_s \cup \{n\} \backslash \{n'\}) < P(\mathcal{N}_s \cup \{n\}). \tag{47}$$

Proof: The proof is presented in Appendix B. ■

According to Theorem 1, we do not need to search all possible combinations of set $\mathcal{N}'$ because if there exists a subset of RRHs that can decrease the total power consumption, there must exist an RRH that can also lower power consumption. So the key idea of our proposed "open" operation is as follows: Try to find RRH $n' \in \mathcal{N}_s$ that satisfies (47) without violating rate and spectral efficiency requirements, if no such an RRH exists, the operation cannot improve current solution; otherwise, close RRH n' and repeatedly search the remaining RRHs in $\mathcal{N}_s$ until no RRH satisfies (47). Obviously, our proposed "open" operation would terminate in polynomial time. It is worthy to highlight that we do not need to find a specific $\mathcal{N}'$ to improve the solution since we only need to confirm the existence of such a set of RRHs. Once we find an RRH that satisfies (47), implying the existence of $\mathcal{N}'$ is confirmed, the local improvement operation can help find a better solution to the original RRH selection problem. If no RRH satisfies (47), we can conclude that there does not exist a set $\mathcal{N}'$ that satisfies (46), indicating the operation cannot find any a better solution.

close$(n, \mathcal{N}')$: In this operation, an RRH $n \in \mathcal{N}_s$ is closed and a subset of RRHs $\mathcal{N}'$ that is disjoint from $\mathcal{N}_s$ is opened, i.e. $\mathcal{N}_s \leftarrow \mathcal{N}_s \cup \mathcal{N}' \backslash \{n\}$. The key idea of our proposed "close" operation is as follows: Try to find RRH $n' \notin \mathcal{N}_s$ that can decrease the total power consumption while satisfying the peak rate requirements. If no such an RRH exists, the operation cannot improve the current solution; otherwise, close RRH n' and repeatedly search the remaining RRHs in $\mathcal{N} \backslash \mathcal{N}_s$ until no operation can decrease the total power consumption. We have the following theorem to help confirm the existence of $\mathcal{N}'$.

Theorem 2: If there exists a set of RRHs $\mathcal{N}' \subseteq \mathcal{N}_s$ that satisfies

$$P(\mathcal{N}_s \cup \mathcal{N}' \backslash \{n\}) < P(\mathcal{N}_s \backslash \{n\}), \tag{48}$$

there must exist an RRH $n' \in \mathcal{N}'$ that satisfies

$$P(\mathcal{N}_s \cup \{n'\} \backslash \{n\}) < P(\mathcal{N}_s \backslash \{n\}). \tag{49}$$

Proof: The proof is presented in Appendix C. ∎

It is also worth set $\mathcal{N}'$ which can satisfy (48) and a subset $\mathcal{N}_1'$ of $\mathcal{N}'$ that $\mathcal{N}_s \cup \mathcal{N}_1' \backslash \{n\}$ is not feasible due to the peak rate requirement constraints. So we ignore the peak rate requirement constraints and find an $\mathcal{N}'$ at first in the "close" operation because if the peak rate requirement constraints is still violated by opening RRHs $\mathcal{N}'$ and closing RRH n, we can conclude that $\mathcal{N}_s \cup \mathcal{N}_1' \backslash \{n\}$ is not feasible.

Based on Theorem 1 and Theorem 2, we can conclude that $\mathcal{N}_s$ is locally optimal solution if none of the three operations can decrease the total power consumption and the algorithm stops at this point. It is worth noticing that "multi" operation, which is generalized the "open", "close" operations, is introduced in [29] and [30]. It is straightforward to generalize our proposed "open", "close" operations to the "multi" operation. Due to its high computational complexity, we do not employ the operation in this paper. Our proposed local search algorithm is summarized in Table III.

TABLE III
LOCAL SEARCH ALGORITHM FOR THE RRH SELECTION PROBLEM

Algorithm 2

Initialization: Feasible solution $\mathcal{N}_s$
repeat
  "add" operation
  for $n \in \mathcal{N} \backslash \mathcal{N}_s$
    Calculate $P(\mathcal{N}_s \cup \{n\})$;
    if $P(\mathcal{N}_s \cup \{n\}) < P(\mathcal{N}_s)$
      $\mathcal{N}_s \leftarrow \mathcal{N}_s \cup \{n\}$;
    end if
  end for
  "open" operation
  for $n \in \mathcal{N} \backslash \mathcal{N}_s$
    $\mathcal{N}_s^{temp} = \mathcal{N}_s \cup \{n\}$;
    repeat
      for $n' \in \mathcal{N}_s^{temp}$, $\mathcal{F}(\mathcal{N}_s^{temp} \backslash \{n'\}) == True$
        if $P(\mathcal{N}_s^{temp} \backslash \{n'\}) < P(\mathcal{N}_s^{temp})$
          $\mathcal{N}_s^{temp} \leftarrow \mathcal{N}_s^{temp} \backslash \{n'\}$;
        end if
      end for
    until $\mathcal{N}_s^{temp}$ does not change
    if $P(\mathcal{N}_s^{temp}) < P(\mathcal{N}_s)$
      $\mathcal{N}_s = \mathcal{N}_s^{temp}$;
    end if
  end for
  "close" operation
  for $n \in \mathcal{N}_s$
    $\mathcal{N}_s^{temp} = \mathcal{N}_s \backslash \{n\}$;
    repeat
      for $n' \in \mathcal{N} \backslash \mathcal{N}_s^{temp}$
        if $P(\mathcal{N}_s^{temp} \cup \{n'\}) < P(\mathcal{N}_s^{temp})$
          $\mathcal{N}_s^{temp} \leftarrow \mathcal{N}_s^{temp} \cup \{n'\}$;
        end if
      end for
    until $\mathcal{N}_s^{temp}$ does not change
    if $P(\mathcal{N}_s^{temp}) < P(\mathcal{N}_s)$ and $\mathcal{F}(\mathcal{N}_s^{temp}) == True$
      $\mathcal{N}_s = \mathcal{N}_s^{temp}$;
    end if
  end for
until No operation can decrease the total power consumption
return $\mathcal{N}_s$

IV. NUMERICAL RESULTS

We perform a series of experiments to evaluate the performance of our proposed algorithms. Simulation parameters of the considered system, such as path-loss model, maximum transmission power, system bandwidth, spectral efficiency requirement, etc., are based on the specifications proposed in [32]. We adopt the transport network power consumption model proposed in [11] and [17]. All results are averaged over 200 Monte Carlo simulations. The service area is 2×2 km^2. The average spatial traffic is distributed uniformly in the service area, as well as the site of each RRH. The area is discretized into K TDAs by using the method proposed in [21] and [22]. The total rate requirement of the region is 1Gbps and the required spectral efficiency of each TDA is 0.1bit/Hz. We set $R_k^{peak} = 3R_k^{avg}$ in all considered scenarios. The maximum transmission power is 1W and the available bandwidth is 100MHz for each RRH. The drain efficiency of radio frequency power amplifier η_n is 25%. The power consumption of the OLT is 20W. The consumed powers of each transport link in active mode and sleep mode are 3.85W and 0.75W, respectively. Thus P_n is 3.1W for each RRH and P_{fixed} can be calculated as $20 + 0.75N$ (in W). The path loss (in dB) between an RRH and a TDA is calculated as $140.7 + 36.7\log_{10}(D)$, where D (in km) is the distance between the RRH and the center of TDA for simplicity. The standard deviation of lognormal shadowing is 10dB. The noise PSD is −184dBm/Hz. The parameters of the barrier method are set as follows: $t^{(0)} = 0.1, \epsilon_b = \epsilon_n = 10^{-2}, \mu = 10, \alpha = 0.01, \beta = 0.1$.

We compare our proposed RRH selection scheme with the following ones: No RRH selection and greedy-based RRH selection proposed in [17]. For the former, all RRHs are always active (i.e. the initialization of Algorithm 2). We only minimize the total transmission power by using the barrier method. For the latter, all RRHs are active at first, then at each iteration, we switch off the RRH that can yield the largest power saving [17]. The greedy-based RRH selection scheme can also obtain a locally optimal solution to our formulated problem.

First, we investigate the convergence of our proposed local search algorithm as shown in Fig. 2 and Fig. 3. Fig. 2 illustrates the power consumption during each iteration with

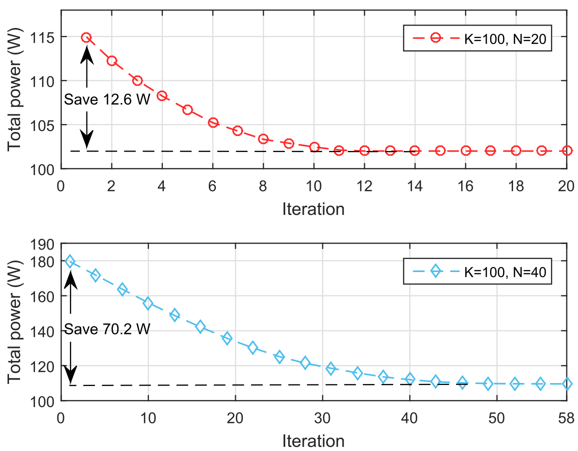

Fig. 2. Power consumption during each iteration with different number of RRHs. $R_D^{avg} = 1$ Gbps, $S_k^{min} = 0.1$ bps/Hz, $\forall k$.

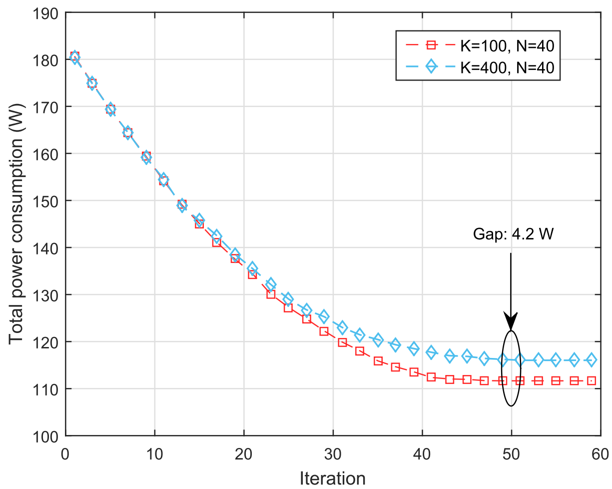

Fig. 3. Power consumption during each iteration with different number of TDAs. $R_D^{avg} = 1$ Gbps, $S_k^{min} = 0.1$ bps/Hz, $\forall k$.

different number of RRHs in different cases: $K = 100$, $N = 20$ and $K = 100, N = 40$. We can observe from Fig. 2 that our proposed algorithm converges rapidly. It requires about 12 iterations in the case that $K = 100$, $N = 20$ and 46 iterations in the case that $K = 100, N = 40$. Obviously, more RRHs usually require more iterations for convergence. Moreover, more RRHs consume more powers to serve the same number of TDAs because of the extra power consumption of the transport links in sleep mode. As seen from Fig. 2, our proposal can reduce the power consumption of the C-RAN as compared to no RRH selection scheme (the start points in Fig. 2). About 10% and 40% powers can be saved by our proposed algorithm for the cases that $K = 100, N = 20$ and $K = 100, N = 40$, respectively.

Fig. 3 shows the power consumption during each iteration with different number of TDAs. It can be observed from Fig. 3 that the number of iterations is almost the same for different number of TDAs. The total power consumption sightly increases as the increasing of TDAs. The reason is that the RRHs need to consume more powers to meet the spectral efficiency requirements of the new coming TDAs

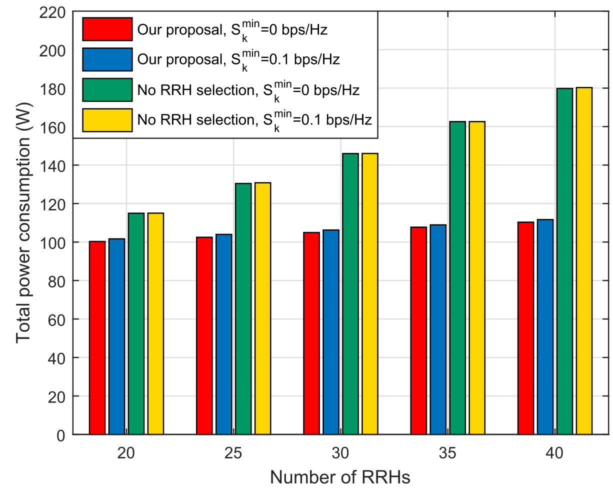

Fig. 4. Total power consumption as a function of the number of RRHs with different spectral efficiency requirements. $R_D^{avg} = 1$ Gbps, $K = 100$.

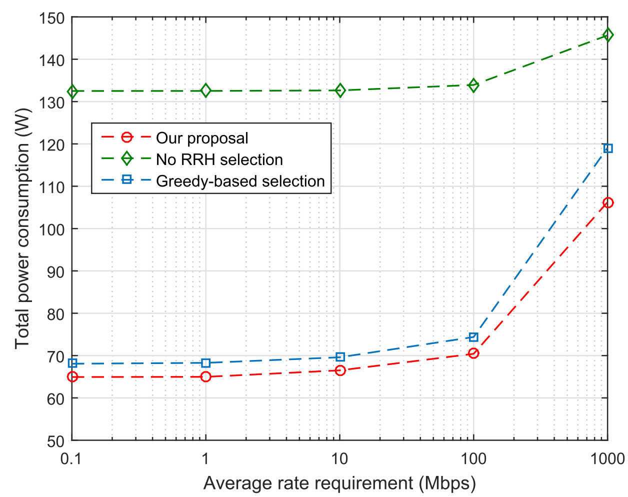

Fig. 5. Total power consumption as a function of average rate requirement of the C-RAN. $K = 100$, $N = 30$, $S_k^{min} = 0.1$ bps/Hz, $\forall k$.

or more RRHs are activated to serve these TDAs. Based on the numerical results shown in Fig. 2 and Fig. 3, we can conclude conservatively that our proposed local search algorithm converges stably and rapidly.

Then we investigate the saved powers as the function of the number of RRHs in different cases: $S_k^{min} = 0.1$ bps/Hz and $S_k^{min} = 0$ bps/Hz, which is illustrated in Fig. 4, where the number of TDAs is 100. Again, we can see our proposed RRH selection scheme can save power as compared to no RRH selection scheme in all scenarios. We also notice that the power consumption for different spectral efficiency requirements is almost the same, indicating that the spectral efficiency is not a significant factor in the power consumption of the C-RAN.

Finally, we give the power consumption as the function of average rate requirement of the C-RAN as shown in Fig. 5. The numbers of TDAs and RRHs are 100 and 30, respectively. The spectral efficiency requirement of each TDA is 0.1bps/Hz. R_D^{avg} varies from 100Kbps to 1Gbps. We can observe from

Fig. 5 that our proposal can achieve at least 45% and 7% power saving when $R_D^{avg} \leq 100$ Mbps as compared to no RRH selection scheme and greedy-based selection scheme, respectively. When $R_D^{avg} = 1$Gbps, our proposal can also save about 30% total power consumption of the C-RAN compared with no RRH selection scheme. Meanwhile, the gap between our proposed scheme and greedy-based selection scheme becomes larger with the increase of R_D^{avg}. The gap is larger than 12.5% when $R_D^{avg} = 1$ Gbps. Our proposal can reduce the power consumption of the C-RAN efficiently.

V. Conclusions

In this paper, we studied the RRH selection problem in the C-RAN, where our goal is to select a subset of RRHs with the minimum total power consumption of the C-RAN while satisfying the average rate requirement, the peak rate demand and the spectral efficiency requirement. The formulated problem is NP-hard. We first design an efficient algorithm to minimize the transmission power while satisfying the rate requirement of the service area with a given subset of RRHs. Then a general transformation to yield an equivalent convex problem is developed and solved by a fast barrier method with low complexity. Based on answering the feasibility problem for a given set of active RRHs to serve users with rate requirements, we propose a local search algorithm to tackle the RRH selection problem, where three modified local improvement operations are proposed to find out locally optimal solutions quickly. Numerical results verify the effectiveness and efficiency of our proposal. Moreover, the proposed algorithms shed insights on how to design energy-efficient C-RANs in practical network scenarios.

Appendix A
Proof of Theorem 1

If all RRHs in $\mathcal{N}'$ can not satisfy (47), that is:

$$P(\mathcal{N}_s \cup \{n\} \backslash \{n'\}) \geq P(\mathcal{N}_s \cup \{n\}), \quad n' \in \mathcal{N}', \tag{50}$$

based on the definition of $P(\mathcal{N}_s)$, we can obtain

$$-P_{n'} + P_r(\mathcal{N}_s \cup \{n\} \backslash \{n'\}) \geq P_r(\mathcal{N}_s \cup \{n\}), \quad n' \in \mathcal{N}'. \tag{51}$$

Inequality (51) is equivalent to

$$\sum_{n' \in \mathcal{N}'} [-P_{n'} + P_r(\mathcal{N}_s \cup \{n\} \backslash \{n'\})] \geq \sum_{n' \in \mathcal{N}'} P_r(\mathcal{N}_s \cup \{n\}). \tag{52}$$

Since $\sum_{n' \in \mathcal{N}'} P_{n'} = P_t(\mathcal{N}')$, we can obtain

$$-P_t(\mathcal{N}') + \sum_{n' \in \mathcal{N}'} [P_r(\mathcal{N}_s \cup \{n\} \backslash \{n'\}) - P_r(\mathcal{N}_s \cup \{n\})] \geq 0. \tag{53}$$

Notice that $P_r(\mathcal{N}_s \cup \{n\}) - P_r(\mathcal{N}_s \cup \{n\} \backslash \{n'\})$ denotes the transmission power that can be saved by n', indicating that

$$\sum_{n' \in \mathcal{N}'} [P_r(\mathcal{N}_s \cup \{n\}) - P_r(\mathcal{N}_s \cup \{n\} \backslash \{n'\})] \geq P_r(\mathcal{N}_s \cup \{n\}) - P_r(\mathcal{N}_s \cup \{n\} \backslash \mathcal{N}'). \tag{54}$$

Inequality (54) always holds because the transmission power saved by RRH $n' \in \mathcal{N}'$ when $\mathcal{N}'$ is active is always not more than the saved transmission power by exclusively activating RRH n'. Then we have

$$\begin{aligned} -P_t(\mathcal{N}') &+ P_r(\mathcal{N}_s \cup \{n\} \backslash \mathcal{N}') - P_r(\mathcal{N}_s \cup \{n\}) \\ &\geq -P_t(\mathcal{N}') + \sum_{n' \in \mathcal{N}'} [P_r(\mathcal{N}_s \cup \{n\} \backslash \{n'\}) - P_r(\mathcal{N}_s \cup \{n\})] \\ &\geq 0. \end{aligned} \tag{55}$$

With simple mathematical operations, we have

$$\begin{aligned} P(\mathcal{N}_s \cup \{n\} \backslash \mathcal{N}') &= P_t(\mathcal{N}_s \cup \{n\}) - P_t(\mathcal{N}') \\ &\quad + P_r(\mathcal{N}_s \cup \{n\} \backslash \mathcal{N}') \\ &\geq P_t(\mathcal{N}_s \cup \{n\}) + P_r(\mathcal{N}_s \cup \{n\}) \\ &= P(\mathcal{N}_s \cup \{n\}). \end{aligned} \tag{56}$$

Eq. (56) violates condition (46). Therefore, we can conclude that there must exist an RRH $n' \in \mathcal{N}'$ that satisfies (47).

Appendix B
Proof of Theorem 2

The proof is similar to that of Theorem 1. If there exists no RRH that satisfies (49), that is:

$$P(\mathcal{N}_s \cup \{n'\} \backslash \{n\}) \geq P(\mathcal{N}_s \backslash \{n\}), \tag{57}$$

based on the definition of $P(\mathcal{N}_s)$, we have

$$P_{n'} + P_r(\mathcal{N}_s \cup \{n'\} \backslash \{n\}) \geq P_r(\mathcal{N}_s \backslash \{n\}). \tag{58}$$

Then we can obtain an equivalent form of (58) as follows

$$P_t(\mathcal{N}') + \sum_{n' \in \mathcal{N}'} [P_r(\mathcal{N}_s \cup \{n'\} \backslash \{n\}) - P_r(\mathcal{N}_s \backslash \{n\})] \geq 0. \tag{59}$$

In (59), $P_r(\mathcal{N}_s \backslash \{n\}) - P_r(\mathcal{N}_s \cup \{n'\} \backslash \{n\})$ is the transmission power that can be saved by RRH n'. As the same reason introduced in Appendix A, we have

$$\begin{aligned} &\sum_{n' \in \mathcal{N}'} [P_r(\mathcal{N}_s \cup \{n'\} \backslash \{n\}) - P_r(\mathcal{N}_s \backslash \{n\})] \\ &\quad \leq P_r(\mathcal{N}_s \cup \mathcal{N}' \backslash \{n\}) - P_r(\mathcal{N}_s \backslash \{n\}). \end{aligned} \tag{60}$$

According to inequalities (59) and (60), we can obtain

$$P_t(\mathcal{N}') + P_r(\mathcal{N}_s \cup \mathcal{N}' \backslash \{n\}) - P_r(\mathcal{N}_s \backslash \{n\}) \geq 0, \tag{61}$$

and

$$\begin{aligned} P(\mathcal{N}_s \cup \mathcal{N}' \backslash \{n\}) &= P_t(\mathcal{N}_s \backslash \{n\}) + P_t(\mathcal{N}') \\ &\quad + P_r(\mathcal{N}_s \cup \mathcal{N}' \backslash \{n\}) \\ &\geq P_t(\mathcal{N}_s \backslash \{n\}) + P_r(\mathcal{N}_s \backslash \{n\}) \\ &= P(\mathcal{N}_s \backslash \{n\}). \end{aligned} \tag{62}$$

Since inequality (62) violates condition (48), there must exist an RRH n' that satisfies (49).

REFERENCES

[1] J. G. Andrews *et al.*, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[2] S. He, Y. Huang, S. Jin, and L. Yang, "Coordinated beamforming for energy efficient transmission in multicell multiuser systems," *IEEE Trans. Commun.*, vol. 61, no. 12, pp. 4961–4971, Dec. 2013.

[3] S. Wang, M. Ge, and W. Zhao, "Energy-efficient resource allocation for OFDM-based cognitive radio networks," *IEEE Trans. Commun.*, vol. 61, no. 8, pp. 3181–3191, Aug. 2013.

[4] X. Wang, F. Zheng, P. Zhu, C. Meng, and X. You, "Energy-efficient resource allocation for OFDMA relay systems with imperfect CSIT," *Sci. China Inf. Sci.*, vol. 58, no. 8, pp. 1–13, Aug. 2015.

[5] A. Ghosh *et al.*, "Heterogeneous cellular networks: From theory to practice," *IEEE Commun. Mag.*, vol. 50, no. 6, pp. 54–64, Jun. 2012.

[6] R. Q. Hu and Y. Qian, "An energy efficient and spectrum efficient wireless heterogeneous network framework for 5G systems," *IEEE Commun. Mag.*, vol. 52, no. 5, pp. 94–101, May 2014.

[7] *C-RAN: The Road Towards Green RAN*, China Mobile Res. Inst., Beijing, China, Oct. 2011.

[8] V. Suryaprakash, P. Rost, and G. Fettweis, "Are heterogeneous cloud-based radio access networks cost effective?" *IEEE J. Sel. Areas Commun.*, vol. 33, no. 10, pp. 2239–2251, Oct. 2015.

[9] R. Irmer *et al.*, "Coordinated multipoint: Concepts, performance, and field trial results," *IEEE Commun. Mag.*, vol. 49, no. 2, pp. 102–111, Feb. 2011.

[10] K. I. Pedersen, Y. Wang, S. Strzyz, and F. Frederiksen, "Enhanced inter-cell interference coordination in co-channel multi-layer LTE-advanced networks," *IEEE Wireless Commun.*, vol. 20, no. 3, pp. 120–127, Jun. 2013.

[11] A. R. Dhaini, P.-H. Ho, G. Shen, and B. Shihada, "Energy efficiency in TDMA-based next-generation passive optical access networks," *Netw., IEEE/ACM Trans. Netw.*, vol. 22, no. 3, pp. 850–863, Jun. 2014.

[12] S. Tombaz, P. Monti, K. Wang, A. Vastberg, M. Forzati, and J. Zander, "Impact of backhauling power consumption on the deployment of heterogeneous mobile networks," in *Proc. IEEE GLOBECOM*, Dec. 2011, pp. 1–5.

[13] E. Oh, K. Son, and B. Krishnamachari, "Dynamic base station switching-on/off strategies for green cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2126–2136, May 2013.

[14] J. Zheng, Y. Cai, X. Chen, R. Li, and H. Zhang, "Optimal base station sleeping in green cellular networks: A distributed cooperative framework based on game theory," *IEEE Trans. Wireless Commun.*, vol. 14, no. 8, pp. 4391–4406, Aug. 2015.

[15] J. Gong, J. S. Thompson, S. Zhou, and Z. Niu, "Base station sleeping and resource allocation in renewable energy powered cellular networks," *IEEE Trans. Commun.*, vol. 62, no. 11, pp. 3801–3813, Nov. 2014.

[16] Y. Yang, L. Chen, W. Dong, and W. Wang, "Active base station set optimization for minimal energy consumption in green cellular networks," *IEEE Trans. Veh. Technol.*, vol. 64, no. 11, pp. 5340–5349, Nov. 2015.

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

[18] S. Luo, R. Zhang, and T. J. Lim, "Downlink and uplink energy minimization through user association and beamforming in C-RAN," *IEEE Trans. Wireless Commun.*, vol. 14, no. 1, pp. 494–508, Jan. 2015.

[19] D. Lee, S. Zhou, X. Zhong, Z. Niu, X. Zhou, and H. Zhang, "Spatial modeling of the traffic density in cellular networks," *IEEE Wireless Commun.*, vol. 21, no. 1, pp. 80–88, Feb. 2014.

[20] K. Aardal, "Capacitated facility location: Separation algorithms and computational experience," *Math. Program.*, vol. 81, no. 2, pp. 149–175, Apr. 1998.

[21] J. G. Carlsson, E. Carlsson, and R. Devulapalli, "Balancing workloads of service vehicles over a geographic territory," in *Proc. IEEE/RSJ IROS*, Nov. 2013, pp. 209–216.

[22] C. Ran, S. Wang, and C. Wang, "Cellular networks planning: A workload balancing perspective," *Comput. Netw.*, vol. 84, pp. 64–75, Jun. 2015.

[23] A. Abdelnasser, E. Hossain, and D. I. Kim, "Clustering and resource allocation for dense femtocells in a two-tier cellular OFDMA network," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1628–1641, Mar. 2014.

[24] S. Wang, W. Zhao, and C. Wang, "Budgeted cell planning for cellular networks with small cells," *IEEE Trans. Veh. Technol.*, vol. 64, no. 10, pp. 4797–4806, Oct. 2015.

[25] W. Zhao, S. Wang, C. Wang, and X. Wu, "Cell planning for heterogeneous networks: An approximation algorithm," in *Proc. IEEE INFOCOM*, Apr./May 2014, pp. 1087–1095.

[26] W. Zhao, S. Wang, C. Wang, and X. Wu, "Approximation algorithms for cell planning in heterogeneous networks," *IEEE Trans. Veh. Technol.*, to be published, doi: 10.1109/TVT.2016.2552487.

[27] S. Wang, Z.-H. Zhou, M. Ge, and C. Wang, "Resource allocation for heterogeneous cognitive radio networks with imperfect spectrum sensing," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 3, pp. 464–475, Mar. 2013.

[28] S. Wang, M. Ge, and C. Wang, "Efficient resource allocation for cognitive radio networks with cooperative relays," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 11, pp. 2432–2441, Nov. 2013.

[29] J. Zhang, B. Chen, and Y. Ye, "A multiexchange local search algorithm for the capacitated facility location problem," *Math. Oper. Res.*, vol. 30, no. 2, pp. 389–403, May 2005.

[30] M. Bansal, N. Garg, and N. Gupta, "A 5-approximation for capacitated facility location," in *Proc. 20th Annu. Eur. Symp. ESA*, Sep. 2012, pp. 133–144.

[31] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge Univ. Press, 2004.

[32] *Evolved Universal Terrestrial Radio Access (E-UTRA); Further Advancements for E-UTRA Physical Layer Aspects*, Document TR 36.814, 3GPP, Mar. 2010.

Wentao Zhao (S'13) received the M.S. degree from Nanjing University, China, in 2013, where he is currently pursuing the Ph.D. degree with the School of Electronic Science and Engineering. His research focuses on resource allocation in wireless networks, and cellular network planning, and optimization.

Shaowei Wang (S'06–M'07–SM'13) received the B.S., M.S., and Ph.D. degrees from Wuhan University, China, in 1997, 2003, and 2006, respectively, all in electronic engineering. From 1997 to 2001, he was with China Telecom as a Research and Development Scientist. Since 2006, he has been with the School of Electronic Science and Engineering, Nanjing University, China. From 2012 to 2013, he was a Visiting Scholar/Professor with Stanford University, Stanford, CA, USA, and The University of British Columbia, Vancouver, BC, Canada. He has authored over 80 papers in leading journals and conference proceedings in his areas of interest. He organized the Special Issue on Enhancing Spectral Efficiency for LTEAdvanced and Beyond Cellular Networks for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and the Feature Topic on Energy-Efficient Cognitive Radio Networks for the *IEEE Communications Magazine*. He serves on the Editorial Board of *IEEE Communications Magazine*, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and *Wireless Networks* (Springer). He serves/served on the technical or the executive committees of reputable conferences including the IEEE INFOCOM, the IEEE ICC, the IEEE GLOBECOM, and the IEEE WCNC.