Robust joint user association and resource partitioning in heterogeneous cloud RANs with dual connectivity

Elaheh Vaezpour a,b,c, Mehdi Dehghan b, Homayoun Yousefi’zadeh c

a The Department of Communication Technology, Iran Telecommunication Research Center, Tehran, Iran
b The Department of Computer and Information Technology Engineering, Amirkabir University of Technology, Tehran, Iran
c The Center for Pervasive Communications and Computing, University of California Irvine, Irvine, CA, USA

A R T I C L E   I N F O

Keywords:
User association
Resource allocation
Heterogeneous cloud radio access networks
Dual connectivity
Robust optimization

A B S T R A C T

Dual connectivity (DC) allows users to connect to both a macrocell and a small cell simultaneously and is introduced to combat constrained fronthaul links while improving throughput and user mobility. In this paper, we investigate the problem of joint user association and resource allocation in heterogeneous cloud radio access networks (H-CRAN) with DC. The problem is formulated as a non-convex mixed integer non-linear program with constraints on non-ideal fronthaul links and quality of service (QoS) requirements. We propose a block coordinate descent (BCD)-based algorithm that successively solves associated subproblems and prove its convergence. We also utilize successive convex approximation (SCA) to deal with binary association variables. Moreover, we discuss a realistic case in which imperfect knowledge of channel state information causes the user’s rate to be uncertain at the time of decision making. Then, we extend our model to a robust optimization problem by introducing protection functions for constraints. We introduce parameters to address the trade-off between robustness and performance and derive probability bounds of constraint violation. Our numerical results validate the effectiveness of both our nominal and robust algorithms in uncertain environments, confirm the convergence of our proposed algorithms, and examine the cost of robustness.

1. Introduction

1.1. Motivation

Despite many advantages of heterogeneous networks (HetNets), high densities of small cells may incur severe interference and therefore limit the spectral efficiency gains as well as the operational deployment. Meanwhile, cloud computing technology has emerged as a key technology for providing centralized on-demand resource processing. By leveraging this technology and integrating it with HetNets, inter-tier interference can be alleviated and cooperative processing gains can be achieved. To that end, the H-CRAN architecture has been recently proposed [1].

In H-CRAN, small cell base stations are simplified by connecting to a cloud through constrained fronthaul links and performing physical layer computations in virtual entities. More specifically, baseband processing units (BBUs) are centralized in a pool referred to as BBU pool, while radio functionality is distributed among remote radio heads (RRHs). By utilizing cloud computing capabilities, resource sharing gains are achieved and network extensions and upgrades will be much easier. Moreover, the overall optimization of resource allocation is facilitated by centralized processing. In order to provide backward compatibility and seamless coverage, macrocell base stations (MBSs) are equipped with their own BBUs and are interfaced to the BBU pool via backhaul links. This also alleviates the burden on the fronthaul network.

If users only associate with RRHs, H-CRAN is simplified to C-RAN [1]. The control and user planes are decoupled in H-CRAN. All control data and low bit-rate messages are forwarded to users by MBSs, while high-data packet traffic is transferred by RRHs. The delivery of control and broadcast signaling is carried out in MBSs in order to alleviate capacity and time delay constraints in fronthaul links between RRHs and the BBU pool and to provide seamless coverage along with user QoS guarantees. RRHs are preferred to provide applications with high speed data rates [1,2]. Recently, connecting to more than one transmission and reception end points has received a considerable attention since it leads to a more efficient deployment of wireless networks. One such architecture is the coordinated multipoint (CoMP) transmission and reception [3]. However, there is a need for high bandwidth and low delay fronthaul links in order to offer tight node synchronization in CoMP. Dual connectivity [4] is recently proposed in 3GPP Release 12 [4] considering relaxed fronthaul and less stringent synchronization requirements. Different from CoMP where bandwidth is shared among transmissions from different access points, bandwidth resources are split...
between transmitters in dual connectivity. Users in DC utilize radio resources from two distinct schedulers. Limited attention has been paid to user association in heterogeneous networks with dual connectivity in the literature. It is also not straightforward to apply the results of single connectivity approaches to the dual connectivity paradigm. While users can be assigned to different RRHs/MBSs, it is not always efficient for users to connect to the strongest receiving power RRH/MBS [5]. User association highly affects network spectral efficiency. Another challenging issue is that fronthaul links impose restrictions on traffic load at each RRH [6]. It is also crucial to determine resources that have to be allocated to macrocells and RRHs in order to satisfy users’ rate requirements. Nonetheless, the problem of joint user association and resource allocation in H-CRAN with DC is not well explored in prior works.

It is well understood that assuming access to perfect channel state information (CSI) in the problem of joint user association and resource allocation is unrealistic. The reason is that on one hand, the exact knowledge of channel gains may not be available due to limited bandwidth availability, especially, in fast fading environments [7]. Moreover, estimation errors do not allow for having exact knowledge of CSI in moderate and slow fading environments. On the other hand, the delay imposed by the fronthaul network makes channel gains outdated at the time of user association decision making [8]. In the presence of uncertainty, user association and resource allocation become more challenging, since the assignment should be performed in a way that prevents unnecessary handovers and undesirable ping-pong effects. Nonetheless, majority of previous research works on this topic ignore this uncertainty. Some previous works have somewhat considered uncertainty of the channel gain matrix in resource allocation problems of wireless networks [9–16]. However, most such works have made limiting assumptions about channel uncertainty. Some assume that uncertain parameters follow a known probability distribution function [9–13] or a family of distribution functions satisfying certain assumptions [14,15]. Some others tackle channel uncertainty by assuming that the statistics of channel variables are known [16]. The problems associated with the limiting assumptions noted above are listed below. First, the information about the distribution or statistics of uncertain parameters may not be available. Furthermore, the mismatch between the assumed probability distribution function and the real one may result in not meeting an acceptable level of QoS required by users. Our previous work of [17] presented a robust formulation for the joint power and subchannel allocation problem in femtocell networks without making any assumption on probability distribution functions of channel gains. Uncertainty was studied in the form of ellipsoidal bounded sets. However, extending this model to the case of RANs introduces a non-convex problem whose solution has a prohibitively high computational complexity. Therefore, there is a need to propose a robust approach that not only is as flexible as the previous method but also has a low complexity solution.

1.2. Related work

User association schemes in traditional single-tier cellular networks are unsuitable for HetNets because they result in load imbalance among base stations in different tiers due to different power levels of BSs [18]. There exists a large body of research in order to enhance load balancing in HetNets with single connectivity. The authors of [19] propose a mathematical formulation of the problem of load balanced user association and resource allocation among users of each base station. In [20], the same problem is investigated for energy-constrained HetNets. The authors of [21] investigate joint user association and resource partitioning among different tiers. In [22], a user association problem is formulated in HetNets where the total bandwidth is divided between the macro and small cell tiers based on a resource partitioning parameter. A joint user association and resource partitioning method is proposed in [23] for a macrocell and a selected set of picocells using KKT conditions and primal–dual update optimization.

The studies above propose single connectivity paradigm and overlook the constraints associated with fronthaul capacity and QoS. The works of [19–21] develop schemes to obtain upper bounds of the user association problem by relaxing integer association variables. In contrast, we consider dual connectivity paradigm, QoS requirement of users, and non-ideal fronthauls. Accordingly, we propose techniques to provide binary association variables.

There is limited research on user association in dual connectivity. In [24], the problem of user association to a pair of base stations is considered. A two-stage method based on stable matching theory is proposed to first pair a macrocell with a small cell and then associate users with each group. In [25], the problem of dual connectivity user association with the objective of throughput maximization is formulated as an integer program. Then, a sub-optimal solution is provided without considering backhaul requirements and resource partitioning. An opportunistic cell association algorithm for DC is presented in [26] under the assumption of having ideal backhaul connections. The authors of [27] consider the problem of user association in DC in order to minimize the radio resource consumption and maximize the backhaul capacity utilization. The authors in [28] address the problem user association and power allocation with the objective to minimize a linear cost model as a function of rates. They formulate the problem as signomial geometric programming. None of the above-mentioned works consider resource partitioning among tiers. Since determining the split ratio depends on the cell load conditions, there is a need to study joint user association and resource allocation for H-CRAN with DC.

1.3. Contributions and organisation

Unlike previous works focusing only on user association in the DC paradigm, this paper investigates joint user association and resource allocation in H-CRANs with DC which can improve both load balancing and network performance. A robust user association and resource allocation scheme is also developed to combat rate uncertainties and parameters controlling the trade-off between performance and robustness are introduced. To the best of our knowledge, this is the first work which studies the problem of joint user association and resource allocation in H-CRAN networks with dual connectivity and also considers channel gain uncertainty. The contributions of this paper are more specifically summarized below.

1. We consider H-CRANs in which users can connect to both macrocell and RRHs simultaneously. Considering a non-ideal fronthaul, QoS requirements, and multiple MBSs and RRHs, we propose a mathematical formulation for the problem of joint user association and resource allocation with the objective of maximizing network throughput while considering proportional fairness. First, we solve the problem for an idealistic case in which a perfect knowledge of users’ rates is assumed. The problem is formulated as an intractable mixed integer non-linear program (MINLP). We decompose the original problem to three sub-problems which are successively solved. Further, we develop techniques from successive convex approximation (SCA) scheme to provide binary association variables.

2. We propose a robust formulation of the problem in order to deal with rate uncertainties in real systems. In order to do so, we utilize the $\Gamma$-robustness approach [29] in which we make no assumptions about channel uncertainty statistics and offer QoS satisfaction guarantees regardless of the distribution of uncertain data. We can also achieve a trade-off between robustness and optimality that is reflected by the choice of $\Gamma$. Several decompositions are proposed to deal with complexity and solve the robust problem.

3. We prove the convergence of our proposed algorithms. We also present the probability of constraint violation and discuss the trade-off between performance and robustness. The robustness and performance of the proposed algorithms are evaluated through extensive numerical experiments and computational complexities of our algorithms are presented.
2. Joint User Association and Resource Allocation (JUARA)

2.1. System model

We consider the downlink of an H-CRAN with $|K_S| = N_S$ RRHs overlaying $|K_M| = N_M$ MBSs where $K_M$ and $K_S$ are the set of MBSs and RRHs, respectively. RRHs are connected to the BBU pool via fronthaul links, while MBSs are interfaced to the BBU pool with backhaul links. Various technologies such as optical networks or microwave transmissions have been proposed for the fronthaul network to transport baseband signals from RRHs to the cloud [30]. However, the capacities of these communication links are limited. Let $C_j$ be the maximum capacity of the fronthaul link that connects RRH $j$ to the BBU pool. The total system bandwidth is denoted by $W$. A typical example of our network model is depicted in Fig. 1. We denote the set of users in the system as $K_U$, where $|K_U| = N_U$. The channel gain between RRH $j$ and user $i$ is denoted by $h_{ij}$. Similarly, the channel gain between macrocell $m$ and user $i$ is denoted by $h_{im}^m$. We assume that the effect of path loss, fading, and shadowing are captured in channel gain parameters. Transmit powers of RRH $j$ and MBS $m$ are $P_j$ and $P_m^m$, respectively. The SINR on the link between RRH $j$ and user $i$ is then given by

\[ \Gamma_{ij} = \frac{P_j h_{ij}^i}{\sigma^2}, \]

where $\sigma^2$ is the noise power. The SINR between MBS $m$ and user $i$ is expressed as

\[ \Gamma_{im}^m = \frac{P_m^m h_{im}^m}{\sigma^2}. \]

Parameters and variables are summarized in Table 1.

2.2. Nominal problem formulation

We formulate the problem of joint user assignment and resource allocation such that the sum utility of users is maximized subject to QoS and fronthaul constraints. More specifically, we consider the inter-frequency scenario which is one of the deployment scenarios of dual connectivity as stated in 3GPP Release 12 [4]. In the inter-frequency scenario, bandwidth is split between MBS and RRH user connections.

We define the real variable $\eta$ to represent the ratio of bandwidth which is allocated to RRH tier. In other words, total bandwidth is split between MBS tier and RRH tier based on the value of $\eta$. Moreover, let $x_{ij}$ be the binary variable denoting whether user $i$ is associated with RRH $j$. We also denote the association of user $i$ with MBS $m$ by the binary variable $y_{im}$.

Since power levels of base stations are different in H-CRANs, maximum SINR and maximum RSSI user association schemes cause load imbalance among base stations. This may force MBSs to carry most of data traffic in H-CRANs [18]. Therefore, a utility of the load-weighted imbalance among base stations. This may force MBSs to carry most of data traffic in H-CRANs [18]. Therefore, utility function of the total effective rate of user $i$ denoted by $R_i$. Respectively, we adopt the logarithmic objective function of the effective rate to achieve load balancing among base stations and improved network performance. This objective function is also inline with the proportional fairness utility function [19]. Hence, we select the proportional fairness utility function as $U_i(R_i) = \log (R_i)$. In DC, users can be assigned to both an RRH and an MBS simultaneously. Hence, the total effective rate $R_i$ is the sum of effective rates, $R_i^j$ and $R_i^m$, that user $i$ can attain from connections to the RRH $j$ and MBS $m$, respectively. The effective downlink rate $r_i^j$ associated with RRH $j$ and user $i$ is expressed as $r_i^j = \sum_{k \in K_j} \frac{y_{ik}}{\eta \Gamma_{ik}}$, where $r_{ij} = W \log (1 + y_{ij})$. It is shown in [19] that equal resource allocation among users of one base station is the optimal strategy for a logarithmic utility. Therefore, if user $i$ connects with RRH $j$, i.e., $x_{ij} = 1$, a portion of the total bandwidth will be assigned to it. Similarly, the achieved effective data rate by user $i$ from MBS $m$ is $r_i^m = \frac{1-\eta}{\eta} \sum_{k \in K_m} y_{ik}$, where $\eta = W \log (1 + y_{im})$. Hence, the utility function is presented as $U_i(R_i) = \log \left( \sum_{k \in K_M} \frac{y_{ik}}{\eta \Gamma_{ik}} + \sum_{k \in K_S} \frac{y_{ik}}{\eta \Gamma_{ik}} \right)$.

Now, we can present the mathematical formulation for the nominal problem of joint user association and resource allocation in H-CRAN with DC as follows.

\[
\begin{align*}
\max_{X,Y,\eta, \xi} & \quad \sum_{i \in K_U} U_i(R_i^d(X,Y,\eta)) \\
\text{S.T.} & \quad \sum_{j \in K_S} x_{ij} \leq 1, \quad \forall i \in K_U \quad (1a) \\
& \quad \sum_{m \in K_M} y_{im} \leq 1, \quad \forall i \in K_U \quad (1b) \\
& \quad \sum_{m \in K_M} y_{im} + \sum_{j \in K_S} x_{ij} \geq 1, \quad \forall i \in K_U \quad (1c) \\
& \quad \sum_{i \in K_U} \frac{\eta x_{ij}}{\Gamma_{ij}} \leq C_j, \quad \forall j \in K_S \quad (1d) \\
& \quad \text{if } x_{ij} = 1 \quad \text{then } \quad \sum_{i \in K_U} \frac{\eta x_{ij}}{\Gamma_{ij}} \geq R_i^j, \quad \forall i \in K_U, \forall j \in K_S \quad (1e) \\
& \quad \text{if } y_{im} = 1 \quad \text{then } \quad \sum_{i \in K_U} \frac{\eta y_{im}}{\Gamma_{im}^m} \geq R_i^m, \quad \forall i \in K_U, \forall m \in K_M \quad (1f) \\
& \quad x_{ij} \in \{0,1\}, \quad \forall i \in K_U, \forall j \in K_S \quad (1g) \\
& \quad y_{im} \in \{0,1\}, \quad \forall i \in K_U, \forall m \in K_M \quad (1h) \\
& \quad 0 \leq \eta \leq 1 \quad (1i)
\end{align*}
\]

where $X = \{x_{ij}, i \in K_U, j \in K_S\}$ and $Y = \{y_{im}, i \in K_U, m \in K_M\}$. In the above formulation, the objective is to maximize the overall network utility function. The constraints of the optimization problem are specified as (1b)–(1j). Constraints (1b) and (1c) require that each user is assigned to at most one RRH and one MBS respectively. In DC, users can connect to both an MBS and an RRH simultaneously and utilize radio resources provided by two distinct schedulers. Therefore, constraint (1d) is imposed to reflect this property and guarantee that each user is associated with at least one base station. Constraint (1e) guarantees that the maximum capacity of fronthaul links are not exceeded. Let $R_i^j$ and $R_i^m$ reflect the required rate of user $i$ in RRH and MBS, respectively. Then,
Table 1

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_P$</td>
<td>Set of RRHs</td>
</tr>
<tr>
<td>$N_P$</td>
<td>Number of RRHs</td>
</tr>
<tr>
<td>$K_M$</td>
<td>Set of MBSs</td>
</tr>
<tr>
<td>$N_M$</td>
<td>Number of MBSs</td>
</tr>
<tr>
<td>$K_U$</td>
<td>Set of users</td>
</tr>
<tr>
<td>$N_U$</td>
<td>Number of users</td>
</tr>
<tr>
<td>$C_I$</td>
<td>The maximum capacity of the fronthaul link that connects RRH $j$ to the BBU pool</td>
</tr>
<tr>
<td>$W$</td>
<td>Total bandwidth</td>
</tr>
<tr>
<td>$b_j$</td>
<td>The channel gain between RRH $j$ and user $i$</td>
</tr>
<tr>
<td>$b_m'$</td>
<td>The channel gain between MBS $m$ and user $i$</td>
</tr>
<tr>
<td>$P_j$</td>
<td>Transmit power of RRH $j$</td>
</tr>
<tr>
<td>$P_m'$</td>
<td>Transmit power of MBS $m$</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>The noise power</td>
</tr>
<tr>
<td>$\gamma_i$</td>
<td>The SINR on the link between RRH $j$ and user $i$</td>
</tr>
<tr>
<td>$\gamma_m$</td>
<td>The SINR on the link between MBS $m$ and user $i$</td>
</tr>
<tr>
<td>$R_f$</td>
<td>The total effective rate of user $i$</td>
</tr>
<tr>
<td>$\eta_i$</td>
<td>The number of users in RRH $j$ with a deviated value of $r_{ij}$</td>
</tr>
</tbody>
</table>

(1f) and (1g) are conditional constraints satisfying minimum required rates of users if users associate with the corresponding base station. In H-CRANs, $R_m^c$ can be set to a small value to ensure delivering of control packets and $R_f^c$ can be set to a larger value to guarantee a high data rate. Constraints (1h), (1i), and (1j) reflect binary user associations and thresholds of the split ratio factor $\eta$.

Proposition 1. The optimization problem (1) is equivalent to the following problem.

$$\text{max}_{X, Y, \eta} P(X, Y, \eta) \triangleq \sum_{i \in K_U} U_i(R_i^c(X, Y, \eta))$$ (2a)

S.T. (1b)–(1e), (1h)–(1j)

$$\left( M_1 - (M_1 - 1)x_{ij} \right) \gamma_i r_{ij} \geq R_f^c \sum_{i \in K_U} x_{ij}, \forall j \in K_S, \forall i \in K_U$$ (2b)

$$\left( M_2 - (M_2 - 1)y_{im} \right) (1 - \eta) r_{im}^c \geq R_f^c \sum_{i \in K_U} y_{im}, \forall m \in K_M, \forall i \in K_U$$ (2c)

where

$$U_i(R_i^c) = \sum_{m \in K_M} y_{im} \log \left( \frac{1 - \eta r_{im}^c}{\sum_{j \in K_U} y_{im} x_{ij}} + \frac{\eta r_{ij}}{\eta r_{im}^c} \right)$$ (2d)

and $M_1$ and $M_2$ are sufficiently large numbers.

Appendix A contains the proof of Proposition 1.

Noting that problem (2) is a MINLP, an exhaustive search of all possible user associations, i.e., $X$ and $Y$, combined with finding the optimal split ratio $\eta$ for each of these associations may solve the problem. Such approach incurs exponential complexity and is hence computationally intractable. In the following subsection, we propose a low complexity iterative approach to solve this problem.

2.3. An iterative approach for solving JUARA

We utilize the block coordinate descent (BCD) method [31] toward solving (2). In doing so, we decompose the decision variables of the problem (2) into three blocks, $X$, $Y$, and $\eta$. At each iteration, problem (2) is sequentially optimized over each block of variables while the remaining ones are held fixed at their last value. This procedure is performed iteratively until convergence. The latter results in identifying the answer set

$$X^{(t)} \rightarrow Y^{(t)} \rightarrow \eta^{(t)} \rightarrow \cdots \rightarrow X^{(t)} \rightarrow Y^{(t)} \rightarrow \eta^{(t)}$$

(3)

to the problem set below.

$$X^{(t)} = \arg \max_{X \in \mathcal{A}_1} P_1(X) \triangleq P(X, Y^{(t-1)}, \eta^{(t-1)})$$ (4a)

$$Y^{(t)} = \arg \max_{Y \in \mathcal{A}_2} P_2(Y) \triangleq P(X^{(t)}, Y, \eta^{(t-1)})$$ (4b)

$$\eta^{(t)} = \arg \max_{\eta \in \mathcal{A}_3} P_3(\eta) \triangleq P(X^{(t)}, Y^{(t)}, \eta)$$ (4c)

where $\alpha$ is the joint constraints (1b)–(1e), (1h)–(1j), (2b), (2c), and the sets $\mathcal{A}_1$, $\mathcal{A}_2$, and $\mathcal{A}_3$ only include constraints on a corresponding variable. The optimization algorithm includes sequentially solving three subproblems to find the optimal RRH-user association matrix $X$, MBS-user association matrix $Y$, and resource allocation factor $\eta$ at each iteration. Notice that each of the subproblems has a smaller number of variables and constraints and hence may be solved with a much lower complexity than the original problem. In the sequel, we present solution methods to solve each of the subproblems.

2.3.1. RRH-user association given $Y$ and $\eta$

Consider the RRH-user association problem defined as follows.

$$\text{max}_{X} P_1(X) \triangleq \sum_{i \in K_U} \left( \sum_{j \in K_S} x_{ij} \log \left( \frac{\eta r_{ij}}{\sum_{j' \in K_S} y_{ij'}} + \frac{1 - \eta r_{ij}}{1 - \eta r_{ij}^c} \right) \right)$$ (5a)

S.T. $\alpha = \{ (1b), (1d), (1e), (1h), (2b) \}$

Note that the term $\sum_{i \in K_M} y_{im} (1 - \eta r_{im}^c)$ is not a function of $X$ and has no impact on the iterative approach of finding the optimal solution. Therefore, we omit this term from the objective function.

Note that Subproblem (5) is a MINLP program. To reduce complexity, we relax integer constraints on variables $x_{ij}$ allowing them to assume any value over the interval $[0, 1]$. Applying the condition $x_{ij}^* = x_{ij}$ forces $x_{ij}$ to take either value 0 or 1 and hence stay binary. This condition is equivalent to the set of conditions \( \{ x_{ij} \geq x_{ij}^*, x_{ij} \leq x_{ij}^* \} \). It is trivial to see that since $x_{ij} \in [0, 1]$, $x_{ij} \geq x_{ij}^*$ always holds. Therefore, we can only impose the constraint $x_{ij} \leq x_{ij}^*$ to imply such condition. However, this
constraint is non-convex. Following [32], we use Lagrangian relaxation with respect to the constraint set Ω = \{x_{ij} ≤ x_{ij}, ∀i ∈ K_U, ∀j ∈ K_S\} and introduce a Lagrange multiplier \( \mu \) to add a new constraint to the objective function as follows.

\[
\max \quad P_1(X) = P(X) + \mu \sum_{i \in K_U} \sum_{j \in K_S} (x_{ij}^2 - x_{ij})
\]

Subject to:

\[
a'_1 = \{(1b), (1d), (1e), (2a), \quad x_{ij} \in [0, 1], \quad ∀i \in K_U, ∀j \in K_S\}
\]

(6a)

The Lagrange multiplier \( \mu \) prohibits violations of the corresponding constraint. In other words, \( \mu \) controls the degree of satisfaction of the constraint. It is shown in [32] that this relaxation is exact with the right choice of \( \mu \) since the objective function \( P(X) \) is not concave due to non-concavity of the term \( x_{ij}^2 - x_{ij} \). To overcome such difficulty, we use SCA method in which variables \( x_{ij} \) are updated successively until converging to the maximum of an approximated version of the objective function \( P(X) \). This approximation can be a tight lower bound of \( P(X) \). Let the separable non-concave part of \( P(X) \) be \( g(X) = x_{ij}^2 - x_{ij} \). Since \( g(X) \) is a convex quadratic function, its first order approximation at a given point \( X^{(a)} \), also representing its global lower bound, can be written as follows.

\[
x_{ij}^2 - x_{ij} ≥ (x_{ij}^{(a)})^2 - x_{ij}^{(a)} + (2x_{ij}^{(a)} - 1)(x_{ij} - x_{ij}^{(a)}) = g_{ij}^{(a)}(X)
\]

Hence, the approximated problem of (6) at a given point \( X^{(a)} \) is expressed as follows.

\[
\max \quad P_1^{(a)}(X) = P_1(X) + \mu \sum_{i \in K_U} \sum_{j \in K_S} g_{ij}^{(a)}(X)
\]

Subject to:

\[
a'_2 = \{(1b), (1d), (1e), (2b), \quad x_{ij} \in [0, 1], \quad ∀i \in K_U, ∀j \in K_S\}
\]

(6b)

2.3.2. MBS-User association given \( X \) and \( \eta \)

Now, we consider the subproblem associated with finding the optimum value of variable \( Y \).

\[
\max \quad P_2(Y) ≜ \left( \sum_{i \in K_U} \sum_{m \in K_M} y_{im} \log \left( \frac{1 - \eta^{(t-1)} x_{im}^{(t)}}{1 - \eta^{(t-1)} y_{im}} \right) \right.

+ \left. \sum_{m \in K_M} \sum_{j \in K_S} x_{ij} y_{im} \log \left( \frac{\sum_{i \in K_U} x_{ij}^{(t)} + \sum_{i \in K_U} y_{im}}{\eta^{(t-1)} y_{im}} \right) \right)
\]

Subject to:

\[
a'_3 = \{(1c), (1d), (2c), (11)\}
\]

(9a)

Similarly to the solution method of (5), we relax integer constraints on the elements of vector \( Y \). Then, the constraint set \( \{y_{im} ≤ y_{im}^{(a)}, ∀i \in K_U, ∀m \in K_M\} \) imposing integrality conditions is added to the objective function with the Lagrangian variable \( \lambda > 0 \) introducing problem below.

\[
\max \quad P_2^{(a)}(Y) ≜ P_2(Y) + \lambda \sum_{i \in K_U} \left( y_{im}^{(a)} - y_{im} \right)
\]

Subject to:

\[
a'_4 = \{(1c), (1d), (2c), \quad y_{im} \in [0, 1], ∀i \in K_U, ∀m \in K_M\}
\]

(10a)

With the right choice of \( \lambda \), the optimal solutions of (10) and (9) are equivalent. We now use the SCA method to successively solve the convex optimization problem (11) until convergence to a limit point. This limit point is a KKT point of problem (10).

\[
\max \quad \tilde{P}_2^{(a)}(Y) ≜ \tilde{P}_2(Y) + \lambda \sum_{i \in K_U} \left( y_{im}^{(a)}^2 - y_{im}^{(a)} + (2y_{im}^{(a)} - 1)y_{im} - y_{im}^{(a)} \right)
\]

Subject to:

\[
a'_5 = \{(1c), (1d), (2c), \quad y_{im} \in [0, 1], ∀i \in K_U, ∀m \in K_M\}
\]

(11a)

2.3.3. Resource allocation given \( X \) and \( Y \)

It now remains to find \( \eta \) with a fixed user association matrices \( X^{(t)} \) and \( Y^{(t)} \). The problem is expressed as follows.

\[
\max \quad \tilde{P}_2(\eta) ≜ \tilde{P}_2(\eta) + \sum_{i \in K_U} \sum_{m \in K_M} \tilde{y}_{im} \log(1 - \eta) + \sum_{j \in K_S} \tilde{x}_{ij} \log(\eta)

+ \sum_{m \in K_M} \sum_{j \in K_S} \tilde{x}_{ij} \log \left( \frac{\sum_{i \in K_U} \tilde{x}_{ij} + \sum_{i \in K_U} \tilde{y}_{im}}{(1 - \eta) \eta_{mj}} \right)
\]

Subject to:

\[
a'_6 = \{(1e), (1j), (2b), (2c)\}
\]

(12)

Proposition 3. Subproblem (12) is a convex program and has a unique optimal solution.

Appendix C contains the proof of Proposition 3.

Subproblem (12) can be efficiently solved relying on interior point methods.

Algorithm 1 combines steps discussed in this section to an integrated solution to the nominal JUARA problem for DC. This procedure is performed at a slow time-scale to lower the communication overhead on the fronthaul network and the computational cost. Therefore, users are assigned to RRHs and MBS based on the estimated channel gains.

We investigate the effect of variations in the channel gains in Section 3 and propose a robust approach to deal with the mentioned uncertainty.

Algorithm 1 Nominal JUARA for DC

Initialize

Set iteration number \( t = 0 \) and maximum number of iterations \( r \).

Initialize variables \( x_{ij}^{(0)}, y_{im}^{(0)}, \) and \( \eta^{(0)} \) for each user \( i \), MBS \( m \), and RRH \( j \).

repeat

1. Set \( t = t + 1 \)

2. Keeping \( Y^{(t-1)} \) and \( \eta^{(t-1)} \) fixed, find optimal RRH-user association matrix \( X^{(t)} \) by applying the SCA method to iteratively solve problem (8).

3. Keeping \( X^{(t)} \) and \( \eta^{(t-1)} \) fixed, find optimal MBS-user association matrix \( Y^{(t)} \) by applying the SCA method to iteratively solve problem (11).

4. Keeping \( X^{(t)} \) and \( Y^{(t)} \) fixed, find optimal resource allocation factor \( \eta^{(t)} \) by solving (12).

until (All the variables converge to a stationary point) or \( (t = r) \)

2.4. Convergence analysis

In this section, we provide the following proposition to demonstrate that our proposed algorithm converges to a stationary point of (2).

Proposition 4. Algorithm 1 converges to a stationary point of (2).

Proof. Algorithm 1 follows the block successive upper-bound minimization (BSUM) framework [34]. It has been proven in [34] that a BSUM algorithm converges to a stationary point if a locally tight upper-bound of the objective function satisfying [34, Assumptions 2] is solved in each subproblem. All of the subproblems in our algorithm satisfy the conditions in [34, Assumptions 2] and therefore [34, Theorem 2] implies that every limit point generated by Algorithm 1 is a stationary point of the original problem (2). This concludes the proof.

3. Robust JUARA

In this section, we propose a robust framework for solving the previous problem in presence of channel gain uncertainties. Then, we derive the probability of constraint violation.
3.1. Problem formulation

Channel gain values and consequently $r_{ij}$ values are affected by uncertainty at the time of optimization decision making. Since improper handling of such uncertainty may result in violating the capacity constraint Eq. (1e), we aim at proposing a robust formulation to deal with this uncertainty. We model the values of $r_{ij}$ by a symmetric and bounded random variable assuming value in the interval $\left[r_{ij} - \delta r_{ij}, r_{ij} + \delta r_{ij}\right]$ where $\delta r_{ij}$ denotes the nominal value and $\delta r_{ij}$ is the highest deviation. We apply $\Gamma$-robustness approach presented in [29] and assume that the value of $r_{ij}$ for at most $\Gamma_j \in \{0, \ldots, N_j\}$ users of RRH $j$ varies from its nominal value. Rather than protecting the problem against a conservative case in which channel gains of all users change, we consider a case in which the value of $r_{ij}$ for at most $\Gamma_j$ users of RRH $j$ is allowed to deviate. The goal of the robust optimization framework is to compute optimal solutions that are feasible if the rates of at most $\Gamma_j$ users of RRH $j$ deviate from their nominal values. Hence, the robust counterpart of the JUARA problem is formulated as follows.

$$\max_{X, Y, \eta} P(X, Y, \eta)$$

S.T. : $\left\{ (1b)-(1e), (1h)-(1j), (2b), (2c) \right\}$

$$\sum_{i \in K_U} \eta_{X_i} f_{ij} + \psi_J(X, \eta) \leq C_j, \forall j \in K_S$$

where

$$\psi_J(X, \eta) = \max_{\{S_U \subseteq \{1, \ldots, N_U\} \subseteq K_U \mid \Gamma_j \}} \left\{ \sum_{i \in S_U} \eta_{X_i} f_{ij} \right\}$$

By varying the parameter $\Gamma_j$, we can adjust the robustness of the model against the conservatism level of the solution. If $\Gamma_j = 0$, then $\psi_J(X, \eta) = 0$ and no robustness is taken into account. If $\Gamma_j = |K_U|$, then $\psi_J(X, \eta) = \sum_{i \in K_U} \eta_{X_i} f_{ij}$ and the rate of all users is changed imposing a high level of conservatism.

We note that problem (13) is also a MINLP optimization problem and computationally intractable. Moreover, the set selection in the inner maximization adds to the non-convexity of the coupling constraint (13b). Fixing each block of variables $X$ or $\eta$ does not make the corresponding subproblem convex. In the sequel, we will develop a solution method for the proposed robust JUARA optimization problem. We apply an approach similar to the one provided in Section 2.3. However, for the robust problem, we have one more subproblem to solve in the iterative algorithm which is the protection function evaluation (PFE) subproblem. The PFE subproblem formulated below evaluates the inner maximization in 4th constraint (13b) for fixed values of $X^{(t-1)}$ and $\eta^{(t-1)}$.

$$\psi_J(X^{(t-1)}, \eta^{(t-1)}) = \max_{\{S_U \subseteq \{1, \ldots, N_U\} \subseteq K_U \mid \Gamma_j \}} \left\{ \sum_{i \in S_U} w_i f_{ij} \right\}$$

S.T. : $\left\{ \sum_{i \in K_U} w_i \leq \Gamma_j, 0 \leq w_i \leq 1, \forall i \in K_U \right\}$

where $W_j = \{w_i, i \in K_U\}$. Since the optimal value of (15) results in having exactly $\Gamma_j$ variables from all $w_i$ variables to be 1, the problem is equivalent to the selection of subset $S_U \subseteq K_U$ with objective function $\sum_{i \in S_U} w_i f_{ij}$, where $|S_U| = \Gamma_j$. In fact, the optimal value of variable $w_i$ shows whether user $i$ in RRH $j$ is selected in subset $S_U$.

The MBS-user association subproblem is the same as (9). The robust versions of RRH-user association and resource allocation subproblems are similar to their nominal counterparts, i.e., Subproblem (5) and (12) except for the fronthaul capacity constraint presented as follows.

$$\sum_{i \in K_U} \eta_{X_i} f_{ij} + w_i f_{ij} \leq C_j, \forall j \in K_S$$

The robust algorithm is detailed in Algorithm 2. The robust algorithm also follows the BSUM framework and its convergence is guaranteed. The proof is similar to that of Proposition 4.  

3.2. Probability bounds of constraint violation

Reducing the values of parameter $\Gamma_j$ results in constraint violation for some instances of uncertainty. In fact, the values of parameter $\Gamma_j$ control the trade-off between constraint violation and performance. In this section, we present probability bounds of violating capacity constraint (1e) under uncertainty model. The only assumption made in our system model is that the errors are independently and symmetrically distributed random variables. Let $X^*$, $Y^*$ and $\eta^*$ be the solutions of the model (13) under uncertainty. Further, assume random variables $r_{ij}$ are independently and symmetrically distributed. Then, the probability that the $\Gamma$th constraint of (1e) is violated satisfies the following bound [29].

$$Pr \left( \sum_{i \in K_U} \eta_{X_i} f_{ij} + \psi_J(X, \eta) > C_j \right) \leq 1 - \phi \left( \frac{\Gamma_j - 1}{\sqrt{2N_U}} \right), \forall j \in K_S$$

where $\phi$ is the cumulative Gaussian probability distribution function.

4. Performance evaluation

4.1. Time complexity

The nominal JUARA corresponding to problem (2) is a mixed integer non-linear program. The nominal iterative algorithm (i.e., Algorithm 1) solves three subproblems in each iteration. Since a convex program is iteratively solved in the RRH-user association subproblem, the total complexity of this subproblem is in the order of $O(T_1(N_U, N_S))$ where $T_1$ is the total number of iterations needed for the SCA method to converge. The time complexity of MBS-user association subproblem is in the order of $O(T_2(N_M, N_S))$ where $T_2$ represents the number of iterations needed for convergence of the SCA method. The resource allocation subproblem is a convex program with one variable and $O(N_U, N_S)$ constraints. Therefore, its time complexity is in the order of $O(N_U, N_S)$. We assume that Algorithm 1 iterates for $T_3$ iterations with $T_3 \geq \gamma$. The total time complexity of the nominal iterative algorithm is hence in the order of $O(T_1(N_U, N_S)^3)$ since $N_M \leq N_S$. In the case of robust JUARA, we also have to solve the PFE subproblem for every RRH at each iteration. The PFE subproblem is a linear program with a complexity order of $O(N_U^3)$. Since it has to be performed for each RRH, the complexity would be in the order of $O(N_U N_M^3)$. Hence, the total complexity of the robust algorithm is also in the order of $O(T_1(N_U, N_S)^3)$ where $T_1$ is the number of iterations required for Algorithm 2 to converge.
Table 2
A listing of simulations’ parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>RRH power</td>
<td>0.01 W</td>
</tr>
<tr>
<td>MBS power</td>
<td>1 W</td>
</tr>
<tr>
<td>Number of RRHs</td>
<td>10</td>
</tr>
<tr>
<td>Number of MBSs</td>
<td>3</td>
</tr>
<tr>
<td>Number of users</td>
<td>50</td>
</tr>
<tr>
<td>Fronthaul link maximum capacity</td>
<td>30 Mbps</td>
</tr>
<tr>
<td>$R$</td>
<td>100 kbps</td>
</tr>
<tr>
<td>$R^m$</td>
<td>50 kbps</td>
</tr>
<tr>
<td>AWGN power spectral density</td>
<td>$-174$ dBm/Hz</td>
</tr>
<tr>
<td>$W$</td>
<td>50 MHz</td>
</tr>
</tbody>
</table>

4.2. Experimental results

In this subsection, we present our numerical results noting that all of the simulations are conducted using MATLAB. We simulate an H-CRAN as follows. The locations of MBSs are fixed in an area of $1000 \times 1000$ m$^2$, while RRHs and users are uniformly distributed in this area. The channel gains between RRHs, MBS, and users are estimated considering Rayleigh fading, path loss, and shadowing. Rayleigh fading is represented by an exponentially distributed random variable with a unit mean. The path loss model between an RRH and a user is estimated using the following equation.

$$P_L(dB) = 37 + 30 \log_{10} d_{ij}$$

(18)

where $d_{ij}$ is the distance between the RRH $i$ and user $j$. This path loss model is chosen according to 3GPP technical report [35]. Shadowing is modeled as a log-normal random variable. The shadow fading standard deviation for RRH and user is set to 10 dB. The path loss between MBS and UE is modeled as $P_L(dB) = 34 + 40 \log_{10} d_{mj}$. The shadow fading deviation between MBS and user is set to 12 dB. We assume that $R^s_i = R^s$ and $R^m_i = R^m$ for all $i \in K_U$. In our robust approach, we assume that $\tilde{\tau}$ is equal to 10% of its nominal value. Unless otherwise mentioned, the remaining simulation parameters are summarized in Table 2. Detailed information of each experiment is given in the corresponding subsection. We report the results averaged over 50 Monte Carlo network realizations in the experiments. We compare nominal and robust JUARA algorithms against dual connectivity user association (UA) proposed in [25] and the single connectivity user association method proposed in [19]. We adopt equal and optimal resource allocation among different tiers for these methods. We also consider QoS and fronthaul capacity constraints in both schemes.

4.2.1. Convergence behavior of the proposed algorithms

First, we evaluate the convergence behavior of our robust and nominal JUARA algorithms for a random network configuration. Fig. 2a shows the iteration numbers versus utility for the inner loops of the nominal and robust algorithms. The convergence behavior of the outer loop of these algorithms is presented in Fig. 2b. For the robust algorithm, it is assumed that $\Gamma_j$ is equal for all RRHs with values $\Gamma_j = \Gamma = 2$. We see that both nominal and robust JUARA algorithms offer rapid convergence characteristics in each loop. Fig. 2b compares our solution with the optimal solution identified by exhaustive search for a small network with 2 MBSs, 2 RRHs and 6 users. As can be seen from Fig. 2b, our algorithm has a utility very close to that of the optimal solution with a much lower time complexity.

4.2.2. Performance comparison

To demonstrate the advantage of our proposed algorithms, we compare them against the single connectivity paradigm [19] (SCUA) and the user association algorithm proposed in [25] for the dual connectivity scheme (DCUA). Network utility versus the number of RRHs is reported in Fig. 3 for 25 and 50 number of users. As can be seen from Fig. 3, the utility is the highest in our nominal approach. This is due to the fact that users can utilize radio resources of more than one base station in the dual connectivity mode. We observe that the dual connectivity scheme [25] has the lowest utility, since it fails to provide fairness among users and this has a great impact on the logarithmic utility.

In addition and as evident in Fig. 3, network utility generally increases as the number of RRHs increases. It can also be seen that the increase in the number of RRHs has a more pronounced impact on dual connectivity compared to single connectivity. It is observed that the proposed method is more flexible in selecting the best base stations to connect. We note that the percentage of dual connected users increases from 30 to 100 with the increase in the number of RRHs. Therefore, we conclude that the higher density of RRHs leads to higher utility values in the dual connectivity paradigm. However, network utility has a slower increase when the RRH density is high. This is due to the limited resources and high interference among RRHs.
4.2.3. The effect of fronthaul and QoS constraints

In this experiment, we illustrate the effect of fronthaul capacity on the performance of algorithms. We vary the fronthaul capacity from 20 Mbps to 80 Mbps and present values of the network utility and the split ratio $\eta$ in Figs. 4a and 4b respectively. In the robust algorithm, we assume that $\Gamma_j = 2$ for all RRHs. As seen by Fig. 4a, the network utility decreases with the decrease in fronthaul capacity. The reason is that user’s data transmitted by the RRH is now constrained by the fronthaul capacity. The feasible region of the problem becomes smaller. We observe that the network utility has a slower increase when the fronthaul capacity is high. This is justified in the sense that when the fronthaul capacity is very high, the corresponding constraint becomes non-binding. Moreover, when the fronthaul capacity is high, the effect of uncertainty decreases and the difference between the utilities of nominal and robust algorithms reduces. Next, we conduct an experiment to evaluate the effect of minimum data rate requirement on the network utility. We assume that $R_m = R_s = R_i$. It is observed in Fig. 5 that the network utility decreases as the value of $R_i$ increases. The reason is that a larger value of $R_i$ results in a smaller feasible region of the optimization problem. In other words, when the values of $R_i$ are large, the remaining resources are insufficient to further improve the network utility after providing all users with their minimum data rate requirement. When $R_i$ is set to a very high value, the system becomes infeasible meaning that one or more constraints of set $\phi$ cannot be simultaneously satisfied while $R_i$ is guaranteed for all RRH users. Infeasible solutions are given zero utility in this experiment.

4.2.4. Robustness cost

In this section, we provide a performance evaluation comparison of the robust JUARA against nominal JUARA. In Fig. 6, we provide a comparison of the network utility as a function of $\Gamma$. We note that the value of $\Gamma$ represents the number of users in each RRH whose $r_{ij}$s are changed. The results are presented for realistic and unrealistic environments. In unrealistic environments, perfect CSI knowledge is available and fronthaul links are delay-free. While this type of environment is unrealistic, we present the results to offer insight about the robustness cost where uncertainty does not exist. As seen by Fig. 6, network utility in the nominal scheme is higher albeit not much than that of robust scheme in unrealistic environments. To the contrary and when uncertainties are present in realistic environments, the utility of the nominal scheme is much lower than that of the robust scheme and in some cases drops to zero. It must be mentioned that the optimal solution of the nominal scheme is not feasible for all values of uncertainty in realistic environments. The fronthaul constraint is met at equality and variations in the aggregated rate of each RRH violate the constraint. Infeasible nominal solutions are given a zero utility in Fig. 6. On the other hand, the robust approach guarantees that all the constraints are satisfied offering acceptable performance in each case. The robust approach has the same utility in both realistic and unrealistic environments.

4.2.5. The effect of network load

In this experiment, we vary the number of users and evaluate its effect on network utility. Each point in Fig. 7 is the average of 50 simulation runs with random topology instances. We include error bars which indicate 95% confidence that the actual average is within the range of depicted interval. As can be seen from Fig. 7, the network utility increases first and gradually reaches saturation with the increase of the number of users being due to limited available resources. When the number of users with a predefined rate requirement is very large, the system becomes infeasible. When the system is infeasible, we require an appropriate admission control policy. An admission control policy is to remove some users until the system becomes feasible. Here, we only assume feasible systems and leave the admission control problem to future work. Comparing the results show that our proposed scheme can yield more utility compared to other methods. While still having
better performance, the utility gain of our proposed method slightly decreases as the network load increases. This is mainly due to the increased interference and shrunken feasible region.

5. Conclusion

This paper focused on the fronthaul and QoS-aware joint user association and resource allocation problems, both nominal and robust, in H-CRANs with dual connectivity. The nominal problem aimed at maximizing proportional fairness utility and was formulated as an MINLP optimization problem. We proposed an iterative BCD-based algorithm successively solving three sub-problems. We also utilized the successive convex approximation method to keep the association variables binary. The convergence of the proposed algorithm was proved. Moreover, we noted that the existing methods have not systematically and explicitly factored rate uncertainties into the corresponding problem formulation. Such uncertainties may increase the likelihood of violating fronthaul rate constraint. Subsequently and absent from any prior work, we also presented a robust formulation of this problem considering the actual effect posed by rate uncertainties on the joint user association and resource allocation. The robust optimization framework utilized in this paper modeled rate errors as uncertain yet bounded perturbations. We then proposed another iterative BCD-based algorithm to offer a computationally efficient solution to the problem. Moreover, we derived the probability of the fronthaul constraint violation. Our proposed algorithms were widely analyzed through experiments and the tradeoff between optimization performance and robustness was validated. In this paper, the power of base stations were assumed to be static. Future works could include power control methods for dual connections. It also remains to consider the admission control problem. Moreover, investigating other dual connectivity deployment scenarios could be an interesting future research direction.

Appendix A. Proof of Proposition 1

We first prove the equivalency of the constraints (2b) and (1f). Constraint (1f) implies that if user $i$ is associated with RRH $j$ ($x_{ij} = 1$), it should be provided with a minimum data rate. Otherwise, the constraint should be non-binding. Constraint (2b) ensures such condition. When $x_{ij} = 1$, constraint (2b) becomes

$$
\eta_{r_{ij}} \geq R_f \sum_{j \in K_j} x_{ij}, \forall j \in K_S, \forall i \in K_U
$$

(A.1)

When $x_{ij} = 0$, the constraint is non-binding due to the positivity and large value of $M_1$,

$$
M_1 \eta_{r_{ij}} \geq R_f \sum_{j \in K_j} x_{ij}, \forall j \in K_S, \forall i \in K_U
$$

(A.2)

Similarly, constraints (2c) and (1g) are equivalent. We now prove the equivalency of the objective functions $U_i(R'_c)$ and $U'_i(R'_c)$. Considering constraint (1b) and the fact that variables $x_{ij}$ and $y_{im}$ are binary, we can write the objective function $U_i(R'_c)$ for all possible cases as follows.

1. The case of user $i$ connecting to both RRH and MBS: $\sum_{j \in K_j} x_{ij} = 1$ and $\sum_{m \in K_M} y_{im} = 1$.

$$
U_i(R'_c) = \sum_{m \in K_M} \sum_{j \in K_j} y_{im} x_{ij} \log \left( \frac{(1 - \eta) y_{im}^p}{\sum_{j \in K_j} y_{im} + \eta y_{ij}} \right)
$$

(A.3)

2. The case of user $i$ only connecting to an RRH: $\sum_{j \in K_j} x_{ij} = 1$ and $\sum_{m \in K_M} y_{im} = 0$.

$$
U_i(R'_c) = \sum_{j \in K_j} y_{ij} \log \left( \frac{\eta y_{ij}}{\sum_{j \in K_j} x_{ij}} \right)
$$

(A.4)

3. The case of user $i$ only connecting to an MBS: $\sum_{j \in K_j} x_{ij} = 0$ and $\sum_{m \in K_M} y_{im} = 1$.

$$
U_i(R'_c) = \sum_{m \in K_M} y_{im} \log \left( \frac{(1 - \eta) y_{im}^p}{\sum_{j \in K_j} x_{ij}} \right)
$$

(A.5)

4. The case of $\sum_{j \in K_j} x_{ij} = 0$ and $\sum_{m \in K_M} y_{im} = 0$ does not happen due to constraint (1d).

Therefore, we have

$$
U_i(R'_c) = \sum_{m \in K_M} \sum_{j \in K_j} y_{im} x_{ij} \log \left( \frac{(1 - \eta) y_{im}^p}{\sum_{j \in K_j} y_{im} + \eta y_{ij}} \right) + (1 - \sum_{m \in K_M} y_{im}) \sum_{j \in K_j} x_{ij} \log \left( \frac{\eta y_{ij}}{\sum_{j \in K_j} x_{ij}} \right) + (1 - \sum_{j \in K_j} x_{ij}) \sum_{m \in K_M} y_{im} \log \left( \frac{(1 - \eta) y_{im}^p}{\sum_{j \in K_j} x_{ij}} \right)
$$

$$
= \sum_{j \in K_j} \sum_{m \in K_M} x_{ij} y_{im} \log \left( \frac{(1 - \eta) y_{im}^p}{\sum_{j \in K_j} y_{im} + \eta y_{ij}} \right) + \sum_{j \in K_j} x_{ij} \log \left( \frac{\eta y_{ij}}{\sum_{j \in K_j} x_{ij}} \right) + \sum_{m \in K_M} y_{im} \log \left( \frac{(1 - \eta) y_{im}^p}{\sum_{j \in K_j} x_{ij}} \right)
$$

$$
= U_i(R'_c)
$$

This concludes the proof.

Appendix B. Proof of Proposition 2

First, note that the feasible set of Subproblem (8), i.e., $a_{ij}$ is convex. To prove the concavity of $\hat{P}_{1_{in}}(X)$, we just need to prove that $P_1(X)$ is concave since $g_{1_{in}}^p(X)$ is an affine function of $X$. Let $a_{ij} = \eta^{p-1} r_{ij}$ and $b_i = \sum_{j \in K_j} y_{im} \times \eta^{p-1} r_{ij}$, $P_1(X)$ can be written as

$$
P_1(X) = \sum_{j \in K_j} \sum_{k \in K_M} x_{ij} \log \left( \frac{a_{ij}}{\sum_{j \in K_j} x_{ij}} \right) + \sum_{j \in K_j} \sum_{k \in K_M} x_{ij} \log \left( \frac{b_k}{\sum_{j \in K_j} x_{ij} + b_k} \right)
$$

$$
= \sum_{j \in K_j} \sum_{k \in K_M} x_{ij} \log \left( \frac{\sum_{k \in K_M} x_{ij}}{a_{ij}} + b_k \sum_{j \in K_j} y_{im} \right)
$$

$$
= \sum_{j \in K_j} \sum_{k \in K_M} y_{im} \log \left( \frac{a_{ij} b_k}{\sum_{j \in K_j} x_{ij}} \right)
$$

$$
\sum_{m \in K_M} y_{im}^{p-1} \text{ is binary and takes a value of either 0 or 1. Therefore,}
$$

$$
\sum_{m \in K_M} y_{im}^{p-1} = 1 \Rightarrow \phi(X) = \sum_{i \in K_i} x_{ij} \log \left( \frac{a_{ij} b_k}{\sum_{j \in K_j} x_{ij} + 1} \right)
$$

$$
\sum_{m \in K_M} y_{im}^{p-1} = 0 \Rightarrow \phi(X) = \sum_{i \in K_i} x_{ij} \log \left( \frac{a_{ij}}{\sum_{j \in K_j} x_{ij}} \right)
$$

Fig. 7. A comparison of network utility as a function of the number of RRHs.
In both cases, the Hessian matrix is negative definite and hence \( f_\eta (X) \) is concave. \( P_3 (\eta) \) is the summation of a set of concave functions which is in turn concave. This concludes the proof.

Appendix C. Proof of Proposition 3

First, note that the feasible set of Subproblem (12) is convex. Now, we prove that the objective function \( P_3 (\eta) \) is strictly concave. Consider the following functions

\[
\begin{align*}
 f_1^{\prime} (\eta) &= y_{im}^{(i)} \log (1 - \eta) + \sum_{k \in K_u} x_{im}^{(i)} / \eta_{ij} \\
 f_1^{\prime \prime} (\eta) &= x_{im}^{(i)} / \eta_{ij}^2 + \sum_{k \in K_u} y_{im}^{(i)} / (1 - \eta) r_{im}^{(i)}
\end{align*}
\]

(C.1)

It is straightforward to see that \( f_1^{\prime} (\eta) \) and \( f_1^{\prime \prime} (\eta) \) are strictly concave functions of \( \eta \). Let \( Q(\eta) = \sum_{k \in K_u} x_{im}^{(i)} / \eta_{ij} + \sum_{k \in K_u} y_{im}^{(i)} / (1 - \eta) r_{im}^{(i)} \). The second derivative of \( Q(\eta) \) is

\[
\frac{d^2 Q}{d \eta^2} = \frac{2 \sum_{k \in K_u} x_{im}^{(i)} / \eta_{ij}^3}{\eta_{ij}^2} + \frac{2 \sum_{k \in K_u} y_{im}^{(i)} / (1 - \eta) r_{im}^{(i)}}{(1 - \eta) r_{im}^{(i)}}
\]

(C.2)

Since \( 0 \leq \eta \leq 1 \) and \( \sum_{k \in K_u} x_{im}^{(i)} / \eta_{ij} \geq 1, \frac{d^2 Q}{d \eta^2} > 0 \) and \( Q(\eta) \) is strictly convex. Based on the composition rules stated in [36], \( \log (Q(\eta)) \) is strictly concave. The objective function \( P_3 (\eta) \) is the summation of a set of strictly concave functions (i.e., \( f_1^{\prime}, f_1^{\prime \prime} \) and \( f_2^{\prime} \)) which in turn is strictly concave and has a unique global optimum. This concludes the proof.

References