Distributed Cloud Association and Beamforming in Downlink Multi-cloud Radio Access Networks

Alaa Alameer Ahmad*, Hayssam Dahrouj †, Anas Chaaban ‡, Aydin Sezgin*, Tareq Y. Al-Naffouri §
and Mohamed-Slim Alouini§
*Institute of Digital Communication Systems, Ruhr-Universität Bochum, Germany
†Department of Electrical and Computer Engineering, Effat University, Saudi Arabia
‡School of Engineering, The University of British Columbia, Kelowna, Canada
§King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia

***Abstract*—Conventional cloud-radio access networks assume the existence of a single processor, responsible for managing a plurality of devices. To cope with the current drastic increase in the number of data-hungry systems, several clouds would be practically needed, and so the joint provisioning of inter-cloud and intra-cloud interference becomes a fundamental challenge. This paper considers a multi-cloud radio access network model (MC-RAN) where each cloud is connected to a distinct set of base stations (BSs) via limited capacity fronthaul links. The paper investigates the problem of jointly assigning users to clouds and determining their beamforming vectors so as to maximize the network-wide sum-rate utility. The paper solves such a difficult non-convex combinatorial problem using a heuristic algorithm which uses fractional programming techniques to deal with the non-convexity of the continuous part of the problem, and l_0-norm approximation to account for the binary association part. A highlight of the proposed algorithm is its ability to be implemented in a distributed fashion across the multiple clouds. The simulations illustrate how close is the sum-rate performance of the proposed approach as compared to the sum-rate achieved by a centralized single processor, especially in dense networks.**

I. Introduction

C-RANs are a promising architecture for realizing the requirements of 5G and beyond wireless networks [1]. In C-RAN, a large set of geographically distributed base stations (BSs) are connected to a central processor (CP) at the cloud via high-speed digital links [2]. Such deployment enables large scale signal coordination between BSs, which increases the overall system performance as compared to conventional BSs cellular operation. The majority of works on C-RAN consider a single cloud scenario, in which a single CP in the cloud is responsible for coordinating the operation of the well-spread multi-cell networks, which often contain a large number of BSs and users (see [3], [4] and references therein). However, the plurality and wide spread of devices in next generation systems, would necessitate the deployment of multiple CPs, each responsible for managing a distinct set of BSs [5]–[7]. Each CP at the cloud coordinates the data processing and beamforming vectors of the set of BSs associated with it. The coordination between CPs, however, is on a message passing level, rather than on a signal-processing level. We refer to a C-RAN with multiple CPs as (MC-RAN) to distinguish it from the classical single CP C-RAN. In MC-RAN, the *inter-cloud* interference becomes significant due to the limited communication between distributed CPs; thus the need for managing both the inter-cloud and the intra-cloud interference. In this paper, we consider an MC-RAN, where each cloud coordinates the operation of a set of base-stations. The system performance becomes, therefore, a function of the user-to-cloud association strategy, as well as the beamforming vector of each user. The paper tackles such a problem, and devises an efficient algorithm that can be implemented in a distributed fashion across the multiple CPs.

Recently, MC-RAN systems have been studied in references [5]- [6]. In [5], the authors study MC-RAN problem in which each CP adopts a compression based transmission strategy. In the current paper, however, we focus on the data-sharing strategy since it is shown to achieve better performance in terms of sum-rate [8]. The authors in [7] consider the user-to-CP association problem in a multi-cloud setup and assume fixed beamforming and an infinite fronthaul capacity. The authors in [9] partially overcome this issue by assuming a discrete set of fixed resources associated with each cluster of BSs connected to a specific CP. The impact of finite fronthaul links is further considered in [6].

A part of the current paper solution relies on fractional programming (FP) approach, first proposed in [10] in the context of conventional cellular systems. Along the same direction, reference [11] proposes an iterative algorithm based on fractional programming techniques to deal with joint association and beamforming vectors design in multi-cell scenario. However, the setup in [11] focuses on an uplink scenario in which the transmit power constraints are decoupled between all users. Moreover, the setup in [11] does not adopt a cloud architecture and assumes that the users can be associated only with the BS in their cell. In the downlink MC-RAN, however, the beamforming vectors of the users are coupled with a single per-BS transmit power constraints. Further, due to the limited fronthaul capacity constraints in cloud-based architectures, the approaches proposed in [10], [11] are no-longer applicable. This paper, therefore, in part tactfully tweaks the FP approach to fit the current paper system model by accounting for its underlying optimization variables, objective and constraints.

The paper considers the downlink of an MC-RAN system, and focuses on jointly determining the user-to-cloud association and the users beamforming vectors by maximizing

978-1-7281-7440-2/20/$31.00 ©2020 IEEE

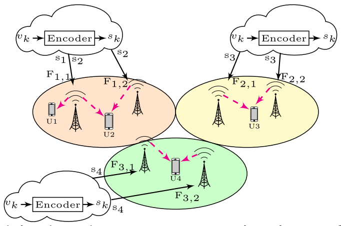

Fig. 1: A multi-cloud system constituting of three clouds and 2 base-stations per cloud.

the sum-rate (SR) subject to a per-BS power and per-BS fronthaul constraints. To tackle such a difficult mixed discrete-continuous non-convex optimization problem, we propose using a distributed iterative algorithm based on fractional programming framework and a l_0-norm heuristic approximation. A highlight of the proposed algorithm is its ability to determine the user-to-cloud association and beamforming vectors in a distributed fashion across the multiple clouds, which makes it amenable to practical implementation. Through extensive numerical simulations, we further show that the performance of our distributed approach is close to the performance which can be achieved by an ideal centralized single CP scenario.

II. System Model and Problem Formulation

A. System Model

Consider the downlink of a MC-RAN, consisting of C CPs each coordinating a group of B_c BSs, each equipped with L antennas, over a network comprising K single-antenna users. Fig. 1 illustrates an exemplary MC-RAN with $C = 3$ CPs, $B_c = 2$ where each BS is connected to one (and only one) CP via a digital fronthaul link with finite capacity.

Let $\mathcal{C} = \{1, \cdots, C\}$ be the set of CPs, $\mathcal{B} = \{1, \cdots, B\}$ be the set of BSs in the network and $\mathcal{K} = \{1, \cdots, K\}$ be the set of users. We assume that each user $k \in \mathcal{K}$ can be assigned to one and only one CP $c \in \mathcal{C}$. Furthermore, we assume that every CP $c \in \mathcal{C}$ is connected to a disjoint cluster of BSs denoted as $\mathcal{B}_c = \{1, \cdots, B_c\}$, where $\cup_{c\in\mathcal{C}}\mathcal{B}_c = \mathcal{B}$, $\mathcal{B}_c \cap \mathcal{B}_{c'} = \varnothing \quad \forall c \neq c'$ and B_c is the total number of BSs in the cluster connected to CP c.

Let $\mathbf{h}^b_{c,k} \in \mathbb{C}^L$ be the channel vector from the b-th BS of the c-th cloud to the k-th user. Let $\mathbf{h}_{c,k} \in \mathbb{C}^{B_c L\times 1}$ be the channel vector from the c-th cloud to the k-th user, i.e., $\mathbf{h}_{c,k} \triangleq [(\mathbf{h}^1_{c,k})^T, \cdots, (\mathbf{h}^{B_c}_{c,k})^T]^T$. Define $\mathbf{w}_{c,k} \in \mathbb{C}^{B_c L\times 1} \triangleq [(\mathbf{w}^1_{c,k})^T, \cdots, (\mathbf{w}^{B_c}_{c,k})^T]^T$ as the transmit beamforming vector used in transmission from CP c's BS's to user k.

We focus on the downlink with the data-sharing strategy. Under such strategy, each CP performs joint encoding and precoding of messages requested by the users associated with it. The precoded message of user k is then shared with a subset of BSs $\mathcal{B}_k \subset \mathcal{B}_c$ and sent over the finite capacity fronthaul links between CP c and each BS in $\mathcal{B}_k$. Note that since each user may be connected with only a subset of BSs $\mathcal{B}_k$, we expect the vector $\mathbf{w}_{c,k}$ to be sparse, i.e., $\mathbf{w}^b_{c,k} = \mathbf{0} \in \mathbb{C}^L$ is a vector with zero valued elements if BS b does not transmit to user k. Let $z_{c,k}$ denote the binary variable of associating cloud c with user k, and 0 otherwise. Let s_k denote the symbol of the encoded message at CP c requested by user k. Let $(\cdot)^\dagger$ denote the hermitian transpose operator.

We can write the signal-to-interference plus noise-ratio (SINR) of user k, when potentially associated with cloud c as

$$\mathrm{SINR}_{c,k} = \frac{\left|\mathbf{h}^\dagger_{c,k}\mathbf{w}_{c,k}\right|^2 z_{c,k}}{\sigma^2 + \sum_{(c',k')\neq(c,k)} \left|\mathbf{h}^\dagger_{c',k}\mathbf{w}_{c',k'}\right|^2 z_{c',k'}} \tag{1}$$

where σ^2 is the Gaussian noise variance, $(c', k') \neq (c, k)$ means that either $c' \neq c$ (inter-cloud interference), or $c' = c$ and $k' \neq k$ (intra-cloud interference). The achievable rate of user k is defined as

$$R_{c,k} = B \log_2(1 + \mathrm{SINR}_{c,k}), \tag{2}$$

where B is the system transmit bandwidth.

B. Problem Formulation

We focus on SR maximization by adjusting the joint user-to-cloud association and beamforming vectors variables. The optimization problem considered in the paper can then be written as

$$\max_{\mathbf{z},\mathbf{w},\mathbf{R}} \quad \sum\nolimits_{(c,k)\in\mathcal{C}\times\mathcal{K}} R_{c,k} \tag{P0}$$

$$\text{s.t.} \quad (2) \tag{3a}$$

$$\sum_{k\in\mathcal{K}} \left\|\mathbf{w}^b_{c,k}\right\|_2^2 z_{c,k} \leq P^{\max}_{c,b}, \tag{3b}$$

$$\sum_{k\in\mathcal{K}} R_{c,k}\mathbb{1}\left\{\left\|\mathbf{w}^b_{c,k}\right\|_2^2\right\} z_{c,k} \leq F_{c,b}, \tag{3c}$$

$$z_{c,k} \in \{0, 1\} \quad \forall c \in \mathcal{C}, \forall k \in \mathcal{K} \tag{3d}$$

where $\mathbf{z} = \{z_{c,k}|\,(c, k) \in \mathcal{C} \times \mathcal{K}\}$ groups all discrete association variables, $\mathbf{R} = \{R_{c,k}|\,(c, k) \in \mathcal{C} \times \mathcal{K}\}$ and $\mathbf{w} = \{\mathbf{w}_{c,k}|(c, k) \in \mathcal{C} \times \mathcal{K}\}$. $P^{\max}_{c,b}$ and $F_{c,b}$ are the maximum transmit power and the fronthaul capacity of BS b in cloud c, respectively. Constraint (3b) represents the maximum transmit power available to BS b, $z_{c,k}$ ensures that this constraint is only active when user k is associated with cloud c. Constraint (3c) represents the fronthaul capacity constraint for BS b, where $\mathbb{1}\{\|\mathbf{w}^b_{c,k}\|_2^2\} = 1$ if $\|\mathbf{w}^b_{c,k}\|_2^2 > 0$, and 0 otherwise. The above optimization is over the binary variables $\mathbf{z}$, and the continuous beamforming vectors $\mathbf{w}$ and rates $\mathbf{R}$ variables. The problem is challenging to solve due to the non-convexity of the objective function and constraint (3c), besides the discrete nature of variables $\mathbf{z}$. The paper addresses the difficulties imposed by the considered optimization problem, and proposes an efficient heuristic algorithm. A highlight of the proposed algorithm is that it can be implemented in a distributed fashion across the multiple clouds in the network.

III. Joint Cloud-Association and Beamforming

In this section, we propose an algorithm that uses fractional programming to deal with the non-convexity of problem (P0), and l_0-norm approximation to account for the binary association part. In [11], the authors use fractional programming and a matching approach to solve the problem of user association and power control in an uplink multi-cell network. However, this approach can not be applied in our case for the following reasons. First, the user-to-CP association in our case is a one-to-many matching because each CP can associate many users with its cluster of BSs, as opposed to the scenario in [11], where each user can be assigned to only one BS in the uplink, where each cell is served by only one BS. Moreover, the downlink transmit beamforming vectors are coupled through the power constraint in (3b), as opposed to the uplink case in [11] where each transmit beamforming vector has its own transmit power budget. Furthermore, the fronthaul capacity constraints makes our problem particularly challenging. Our paper, therefore, develops a different approach that starts by reformulating the fronthaul constraints in such a way that enables applying the FP technique.

A. Fronthaul Constraint Reformulation

To make problem (P0) more tractable, we start by relaxing the complicated fronthaul constraint (3c) which consists of three components: 1) the rate R_k which is a non-convex function of the beamforming vector, 2) the indicator function $\mathbb{1}\{\|\mathbf{w}_{c,k}^b\|_2^2\}$ which represents the association between user k and BS b and 3) the binary cloud-user association variable. We note that the indicator function in our problem is equivalent to the l_0-norm, i.e., $\mathbb{1}\{\|\mathbf{w}_{c,k}^b\|_2^2\} = \left\|\|\mathbf{w}_{c,k}^b\|_2^2\right\|_0$, because $\|\mathbf{w}_{c,k}^b\|_2^2 \in \mathbb{R}_+$. l_0-norm is also a non-convex function of the beamforming vector; however, it is amenable for approximation with a convex function as we see in section III. Now, we can write the fronthaul capacity constraint (3c) as

$$\sum_{k\in\mathcal{K}} \hat{R}_{c,k}\left\|\|\mathbf{w}_{c,k}^b\|_2^2\right\|_0 z_{c,k} \leq F_{c,b}, \tag{4}$$

where $\hat{R}_{c,k}$ is the fixed rate of user k when assigned to a cloud c. While $\hat{R}_{c,k}$ is fixed at this step, it is subsequently updated in an outer loop, i.e., whenever other variables get updated, as shown in the description of Algorithm 1 later in this section. Hence, by choosing the users which contribute *efficiently* to the sum-rate, each BS can render the constraint in (4) satisfied, by serving the best group of users for which the aggregate sum-rate does not exceed the fronthaul capacity limit $F_{c,b}$. Although the problem is still difficult to solve, the fronthaul constraint reformulation allows finding the beamforming vectors, whenever the user-to-cloud association is decided.

B. Fractional Programming Approach

After reformulating the fronthaul constraints, we introduce a variable $\boldsymbol{\gamma} = [\gamma_{1,1}, \ldots, \gamma_{C,K}]^T$. The problem (P0) can then be reformulated as

$$\begin{aligned} &\underset{\{\mathbf{z},\mathbf{w},\boldsymbol{\gamma}\}}{\text{maximize}} \quad \sum_{(c,k)} \log_2(1+\gamma_{c,k}) && \text{(P1)} \\ &\text{s.t.} \quad (3b), (3d) \text{ and } (4) \\ &\gamma_{c,k} = \text{SINR}_{c,k} && \text{(5a)} \end{aligned}$$

The problem (P1) is still not convex; however, it is now amenable for applying FP framework. As we show next, the form in (P1) allows to move the non-convex fraction outside the logarithm. After introducing auxiliary variables, the original optimization problem enables deriving simpler algorithms. Towards this end, we define the *partial Lagrangian* function as

$$\begin{aligned} L(\mathbf{z},\mathbf{w},\boldsymbol{\gamma},\boldsymbol{\lambda}) &= \sum_{(c,k)} \left(\log_2(1+\gamma_{c,k}) - \lambda_{c,k}\gamma_{c,k}\right) \\ &+ \sum_{(c,k)} \lambda_{c,k} \frac{\left|\mathbf{h}_{c,k}^{\dagger}\mathbf{w}_{c,k}\right|^2 z_{c,k}}{\sigma^2 + \sum_{(c',k')\neq(c,k)} \left|\mathbf{h}_{c',k}^{\dagger}\mathbf{w}_{c',k'}\right|^2 z_{c',k'}} \end{aligned} \tag{6}$$

where $\boldsymbol{\lambda} = [\lambda_{1,1}, \ldots, \lambda_{C,K}]^T$ are the Lagrange multipliers related to constraint (5a). By setting the partial derivatives $\frac{\partial L}{\partial \gamma_{c,k}}$ to zero, we get:

$$\gamma_{c,k} = \frac{1}{\lambda_{c,k}} - 1. \tag{7}$$

Then, we express $\lambda_{c,k}$ as a function of $\mathbf{z}$ and $\mathbf{w}$ by using (5a) and (7) to obtain

$$\lambda_{c,k} = \frac{\left(\sigma^2 + \sum_{(c',k')\neq(c,k)} \left|\mathbf{h}_{c',k}^{\dagger}\mathbf{w}_{c',k'}\right|^2 z_{c',k'}\right)}{\sigma^2 + \sum_{(c',k')} \left|\mathbf{h}_{c',k}^{\dagger}\mathbf{w}_{c',k'}\right|^2 z_{c',k'}} \tag{8}$$

By substituting (8) in (6), we obtain the following function:

$$\begin{aligned} f_r(\mathbf{z},\mathbf{w},\boldsymbol{\gamma}) &= \sum_{(c,k)} \left(\log_2(1+\gamma_{c,k}) - \gamma_{c,k}\right) \\ &+ \sum_{(c,k)} \frac{(1+\gamma_{c,k})\left|\mathbf{h}_{c,k}^{\dagger}\mathbf{w}_{c,k}\right|^2 z_{c,k}}{\sigma^2 + \sum_{(c',k')} \left|\mathbf{h}_{c',k}^{\dagger}\mathbf{w}_{c',k'}\right|^2 z_{c',k'}} \end{aligned} \tag{9}$$

This can be used to introduce an equivalent formulation to the optimization problem in (P1) as follows

$$\begin{aligned} &\underset{\{\mathbf{z},\mathbf{w},\boldsymbol{\gamma}\}}{\text{maximize}} \quad f_r(\mathbf{z},\mathbf{w},\boldsymbol{\gamma}) && \text{(P2)} \\ &\text{s.t.} \quad (3b), (3d) \text{ and } (4) \end{aligned}$$

Note that when $\mathbf{z}$ and $\mathbf{w}$ are held fixed, we can find the optimal variables $\boldsymbol{\gamma}^* = \left\{\gamma_{c,k}^* \,|\, (c,k) \in \mathcal{C}\times\mathcal{K}\right\}$ in closed form. To show this, we set the partial derivatives $\frac{\partial f_r}{\partial \gamma_{c,k}}$ to zero and solve the resulting equation for $\gamma_{c,k}$. This yields the optimal $\gamma_{c,k}^*$ as

$$\gamma_{c,k}^* \triangleq \frac{\left|\mathbf{h}_{c,k}^{\dagger}\mathbf{w}_{c,k}\right|^2 z_{c,k}}{\sigma^2 + \sum_{(c',k')\neq(c,k)} \left|\mathbf{h}_{c',k}^{\dagger}\mathbf{w}_{c',k'}\right|^2 z_{c',k'}} \tag{11}$$

which is exactly the SINR of user k if assigned to cloud c.

Proposition 1. *Any feasible vector* $\mathbf{w}$ *for problem (P2) is also feasible for (P1). Moreover, the variables* $\boldsymbol{\gamma}^*$ *given in* (11) *are optimal for both problems (P1) and (P2).*

Proof. Note that the feasible set of beamforming vectors is the same for problems (P1) and (P2), which is characterized through constraints (3b) and (3d). Furthermore, we note that $f_r(\mathbf{z},\mathbf{w},\boldsymbol{\gamma}) = L(\mathbf{z},\mathbf{w},\boldsymbol{\gamma},\boldsymbol{\lambda}^*)$, where $L(\mathbf{z},\mathbf{w},\boldsymbol{\gamma},\boldsymbol{\lambda})$ is the Lagrangian function of problem (P1), λ^* is the optimal Lagrange multipliers defined as $\boldsymbol{\lambda} = \{\lambda_{c,k} | (c,k) \in \mathcal{C} \times \mathcal{K}\}$, and $\lambda_{c,k}$ is given in equation (8). This completes the proof. $\square$

C. Problem Reformulation

We now use the quadratic transform in a multidimensional and complex form, first introduced in [10, Theorem 2], which when tailored to our problem yields (12), where $\mathbf{u} = \{u_{c,k} | (c,k) \in \mathcal{C} \times \mathcal{K}\}$ is the set of complex valued auxiliary variables. When all other variables $\mathbf{z}, \mathbf{w}$ and $\boldsymbol{\gamma}$ are fixed, the optimal auxiliary variables are given by

$$u^*_{c,k} = \frac{\sqrt{(1+\gamma_{c,k})}\mathbf{w}^\dagger_{c,k}\mathbf{h}_{c,k} z_{c,k}}{\sum_{(c',k')} \left|\mathbf{h}^\dagger_{c',k}\mathbf{w}_{c',k'}\right|^2 z_{c',k'} + \sigma^2}. \tag{14}$$

We can verify this by computing the partial derivative of function $f_q(\mathbf{z}, \mathbf{w}, \boldsymbol{\gamma}, \mathbf{u})$ with respect to variable $\mathbf{u}$, set it to zero, and solve for $\mathbf{u}$ (the details of the proof are relatively straightfard, and are omitted due to space limit). Problem (P2) can now be reformulated as

$$\begin{aligned} \max_{\mathbf{z},\mathbf{w},\boldsymbol{\gamma},\mathbf{u}} \quad & f_q(\mathbf{z}, \mathbf{w}, \boldsymbol{\gamma}, \mathbf{u}) \\ \text{s.t.} \quad & (3\text{b}), (3\text{d}), (4) \end{aligned} \tag{15}$$

Note that the function $f_q(\mathbf{z}, \mathbf{w}, \boldsymbol{\gamma}, \mathbf{u})$ is concave in $\mathbf{w}$ when $\mathbf{u}$ is fixed, and is linear in $\mathbf{z}$ when all other variables are fixed. Furthermore, the function $f_q(\mathbf{z}, \mathbf{w}, \gamma, \mathbf{u})$ is in a form that allows us to optimize the association variables and their associated beamforming vectors independently. We, thus, propose an iterative coordinate-assent algorithm to find a solution of problem (15) in the next subsection.

D. Overall Algorithm

As shown in the previous section, the structure of the function $f_q(\mathbf{z}, \mathbf{w}, \boldsymbol{\gamma}, \mathbf{u})$ allows us to optimize it iteratively with respect to the optimization variables. One potential implementation is to update the variables independently; however, this may lead to premature association, i.e., a user associated to a CP in an iteration remains unchanged in future iterations. The rationale behind such pre-mature convergence is that the beamforming vectors of other CP's would be set to zero in this iteration and would remain zero for all future iterations. To avoid this, we propose to use *auxiliary* beamforming vectors in the utility function for optimizing the association decisions. Then, we optimize the actual beamforming vectors that meet constraints (3b) and (4). By doing so, each cloud assigns its available resources to the users assigned to it by solving problem (15) in a distributed fashion per cloud, while all other variables are fixed. The *auxiliary* beamforming vectors are computed by setting the partial derivatives of $\partial f_q / \partial \boldsymbol{w}_{c,k}$ to zero and solving for $\boldsymbol{w}_{c,k}$ in (12), which yields

$$\boldsymbol{\tau}^b_{c,k} = \frac{\sqrt{(1+\gamma_{c,k})}u_{c,k}\mathbf{h}^b_{c,k}}{\sum_{(c',k')} ||u_{c',k'}\mathbf{h}^b_{c,k'}||^2_2 + \mu^*_{c,b}} \tag{16}$$

where $\boldsymbol{\tau}^b_{c,k}$ represents the optimal auxiliary beamforming vector, and $\mu^*_{c,b} \geq 0$ is a coefficient which satisfies the following:

$$\mu^*_{c,b} = \min\left\{\mu_{c,b} \geq 0 : \sum\nolimits_{k\in\mathcal{K}} \left\|\boldsymbol{\tau}^b_{c,k}\right\|^2_2 \leq P^{\max}_{c,b}\right\}. \tag{17}$$

Let $\boldsymbol{\tau}_{c,k} = \left\{\boldsymbol{\tau}^b_{c,k} | (c,k) \in \mathcal{C} \times \mathcal{K}, b \in \mathcal{B}_c\right\}$ and $\boldsymbol{\tau} = \{\boldsymbol{\tau}_{c,k} | (c,k) \in \mathcal{C} \times \mathcal{K}\}$. Now, define the benefit of assigning user k to a cloud c in equation (13). Based on (13), the maximization of f_q with respect to the binary variables $\mathbf{z}$ boils down to the following simple problem:

$$z_{c,k} = \begin{cases} 1, & \text{if} \quad c = \underset{d}{\text{argmax}}\ \xi_{d,k} \\ 0, & \text{otherwise.} \end{cases} \tag{18}$$

After association each user to one cloud, each cloud assigns its resources to its scheduled users by solving problem (15) with respect to beamforming vector $\mathbf{w}$, while all other variables are fixed. The feasible beamforming vectors, therefore, are solved using the following problem:

$$\begin{aligned} \max_{\mathbf{w}} \quad & f_q(\mathbf{z}, \mathbf{w}, \boldsymbol{\gamma}, \mathbf{u}) \\ \text{s.t.} \quad & (3\text{b}), (4) \end{aligned} \tag{19}$$

Problem (19) is still not convex due to the combinatorial nature of the variables $\left\|\|\mathbf{w}^b_{c,k}\|^2_2\right\|_0$. To overcome such challenge, we propose to approximate the non-convex l_0-norm with a weighted l_1-norm, and then reformulate the fronthaul constraint. Such approximation results in a convex constraint as a function of the beamforming vectors. Approximating an l_0-norm with a weighted sum of l_1 norm is widely used in the literature to minimize a non-convex l_0-norm objective function, e.g., see [3] and references therein. To enable the use of such approximation in the context of our paper, we write the function $\left\|\|\mathbf{w}^b_{c,k}\|^2_2\right\|_0$ as a reweighed l_1-norm as follows:

$$\left\|\|\mathbf{w}^b_{c,k}\|^2_2\right\|_0 = \beta^b_k \left\|\mathbf{w}^b_{c,k}\right\|^2_2 \tag{20}$$

Here, β^b_k is a constant weight associated with BS b and user k, and is defined in this work as:

$$\beta^b_k = \frac{1}{\delta + \left\|\mathbf{w}^b_{c,k}\right\|^2_2} \tag{21}$$

where $\delta > 0$ is a regularization constant. Based on this approximation, reformulate constraint[1] (4) as:

$$\sum_{k\in\mathcal{K}_c} \beta^b_k \hat{R}_k \left\|\mathbf{w}^b_{c,k}\right\|^2_2 \leq F_{c,b}, \tag{22}$$

where $\mathcal{K}_c = \{k \in \mathcal{K} | z_{c,k} = 1\}$. The particular choice of weights as given in (21) is motivated by the fact that BSs with small transmit power allocated to user k get higher weights β^b_k. Such BSs eventually drop out of the cluster of BSs sharing the message of user k, whereas only those BSs with reasonable transmit power to user k participate in serving user k. It is also worth mentioning that due to constraint (22), some users may not be scheduled for transmission in the current time slot. This happens when

$$z_{c,k} = 1 \quad \text{and} \quad \left\|\mathbf{w}^b_{c,k}\right\|^2_2 = \mathbf{0} \quad \forall b \in \mathcal{B}_c. \tag{23}$$

[1]Note that constraint (22) is an approximation of the constraint (4). The regularization parameter is chosen to make the approximation error arbitrary small. In our simulations, we choose $\delta = 10^{-12}$ and the beamforming vector $\mathbf{w}^b_{c,k} = \mathbf{0}$ in iteration t if $\left\|\mathbf{w}^b_{c,k}\right\|^2_2 \leq \delta$. This results in negligible error on the achievable data rate of user k.

$$f_q(\mathbf{z},\mathbf{w},\boldsymbol{\gamma},\mathbf{u}) = \sum_{(c,k)} \Big(\log_2(1+\gamma_{c,k}) - \gamma_{c,k} - |u_{c,k}|^2\sigma^2 + 2\sqrt{(1+\gamma_{c,k})}\Re\Big\{u^{\dagger}_{c,k}\mathbf{w}^{\dagger}_{c,k}\mathbf{h}_{c,k}\Big\}z_{c,k} \\ - u^{\dagger}_{c,k}\big(\textstyle\sum_{(c',k')}\mathbf{h}^{\dagger}_{c',k}\mathbf{w}_{c',k'}\mathbf{w}^{\dagger}_{c',k'}\mathbf{h}_{c',k}z_{c',k'}\big)u_{c,k}\Big) \quad (12)$$

$$\xi_{c,k} = \log_2(1+\gamma_{c,k}) - \gamma_{c,k} - |u_{c,k}|^2\sigma^2 + 2\sqrt{(1+\gamma_{c,k})}\Re\Big\{u^{\dagger}_{c,k}\mathbf{h}^{\dagger}_{c,k}\boldsymbol{\tau}_{c,k}\Big\} - \\ \textstyle\sum_{(c',k')} u^{\dagger}_{c',k'}\mathbf{h}^{\dagger}_{c,k'}\boldsymbol{\tau}_{c,k}\boldsymbol{\tau}^{\dagger}_{c,k}\mathbf{h}_{c,k'}u_{c',k'} \quad (13)$$

The optimal beamforming vectors can be determined by solving the following problem:

$$\begin{aligned} \max_{\mathbf{w}} \quad & f_q(\mathbf{z},\mathbf{w},\boldsymbol{\gamma},\mathbf{u}) \\ \text{s.t.} \quad & (3b), (22) \end{aligned} \quad (24)$$

E. Distributed Implementation

Problem (24) is, indeed, a convex optimization problem and can be solved efficiently. This is the case since, the objective is a concave function of the beamforming vector, and constraints (3b) and (22) are quadratic in $\mathbf{w}$. Interestingly, the optimization problem (24) can be implemented distributively such that each cloud designs the beamforming vectors of users associated to it. To this end, the clouds need to exchange interference information only, i.e., cloud c needs to receive the following term $\sum_{(c',k')\neq(c,k)} \left|\mathbf{h}^{\dagger}_{c',k}\mathbf{w}_{c',k'}\right|^2$ from all other clouds $c' \neq c$ to solve problem (24) locally at each cloud c. Using a similar argument, a proper exchange of interference terms between all clouds allows solving for the binary variables $\mathbf{z}$ using (13) and (18) locally at each cloud. In Algorithm 1, we list an iterative coordinate-accent procedure for joint association and beamforming design problem. Note that this algorithm can be implemented either in a distributed or in a centralized manner.

Algorithm 1 Distributed Algorithm for joint association and beamforming design

1: Initialize: $\mathbf{z}$, $\mathbf{w}$ and $\boldsymbol{\gamma}$ to feasible values.
2: **Repeat:** until convergence
3: Update $\mathbf{u}$ by (14)
4: Update $\boldsymbol{\gamma}$ by (11)
5: Update $\mathbf{z}$ by using (13) and (18)
6: Update $\mathbf{w}$ by solving problem (24)
7: Update the rates $\left\{\hat{R}_k|\forall k \in \mathcal{K}\right\}$, using the current $\mathbf{w}$ and equation (2)
8: **End**

IV. Simulations

In this section, we present numerical simulations that illustrate the performance of proposed joint user-to-cloud association and beamforming algorithm. We consider a multi-cloud network deployed in a square area defined by $[-1000 \;\; 1000] \times [-1000 \;\; 1000]\ \mathrm{m}^2$. Both users and BSs are assumed to be uniformly and independently distributed in this area. The density of BSs is set to 39 BS/km^2, which is consistent with the requirements of 5G ultra-dense networks [12]. Furthermore, we assume that each BS has a maximum load of 8 users. The density of users in the network is given in the simulations as needed. The channel model consists of three components: 1) path-loss as $\mathrm{PL}_{b,k} = 148.1 + 37.6\log_{10}(d_{b,k})\mathrm{dB}$, where $d_{b,k}$ is the distance between BS b and user k in km; 2) log normal shadowing with 8dB standard deviation and 3) Rayleigh channel fading with zero mean and unit variance. The channel bandwidth is set to $B = 10$ MHz, the noise power spectrum density is set to -169 dBm/Hz and the maximum transmit PSD is set to -40 dBm/Hz for each BS. Each BS is assumed to be equipped with 2 transmit antennas.

To assess the sum-rate performance as a function of the fronthaul capacity limit, we first consider a four-cloud scenario, where each cloud controls the BSs in its sector. If a user is connected to a cloud, it can be served only by BSs in the area controlled by this cloud. We also evaluate a two-cloud scenario, wherein clouds 1 and 2 are merged to form a single cloud, and clouds 3 and 4 are also merged. Similarly, in single cloud scenario, all clouds are merged into a single cloud which controls the entire set of all BSs. For each multi-cloud scenario, we compare a distributed implementation of the joint association and beamforming vector design algorithm with a centralized implementation. In the distributed version, the clouds cooperate on a scheduling level with minimal information exchange, while beamforming vectors design and fronthaul bandwidth allocation decisions are made individually at each cloud. In the centralized implementation, we assume the existence of a virtual processor which performs joint resource allocation across the clouds, yet each cloud manages its distinct set of BSs and users. First, the user density is set to 45 users/km^2. We further compare our proposed algorithm with the state-of-the-art approach described in [7], which focuses on scheduling users to clouds while assuming fixed beamforming vectors design at the clouds. The sum-rate as a function of the fronthaul capacity is shown in Fig. 2 for different schemes. The figure shows that centralized solutions in different cloud setups outperform the distributed implementation. However, the local beamforming vectors design in the distributed algorithm is less demanding in terms of computational cost. Further, the communication overhead due to information exchange is significantly reduced, since the clouds need only to exchange minimal amount of information in the distributed case. Fig. 2, particularly, shows that our pro-

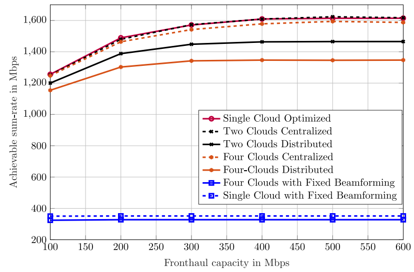

Fig. 2: Sum-rate as a function of fronthaul capacity. Users density is set to 45 users/km^2.

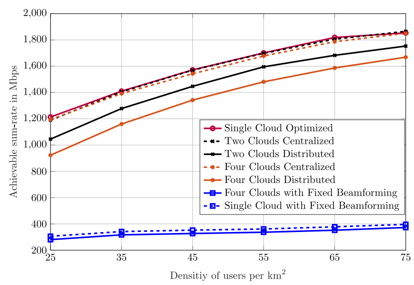

Fig. 3: Sum-rate as a function of users-density. The fronthaul capacity is set to 300 Mbps.

posed joint user-to-CP association and beamforming vectors design algorithm significantly improves the SR performance as compared to the scheme proposed in [7]. On the other hand, the distributed implementation of our algorithm is able to achieve a good performance compared to the solution where a single cloud controls the whole network, which highlights the ability of the proposed algorithm at improving the network throughout, besides its distributed implementation characteristics. Fig. 3 depicts the sum-rate performance of the studied schemes as we vary the density of users in the network from 25 users/km^2 to 75 users/km^2. It is evident from Fig. 3 that as the network becomes more dense, the performance gap between centralized implementation (full-cooperation) and the distributed implementation (limited-cooperation) of multi-cloud scheme becomes smaller, which highlights the important role our proposed algorithm presents, especially in the interference-limit regime at dense networks. Finally, Fig. 4 plots the network sum-rate versus the number of iterations. The figure particularly illustrates the fast convergence of the proposed algorithm, under both the centralized and distributed implementation.

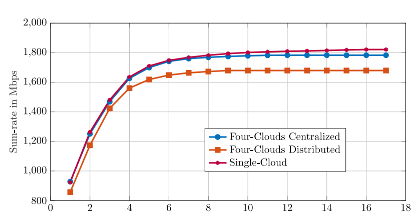

Fig. 4: Convergence of the proposed algorithms.

V. Conclusion

Managing wireless systems with multiple CPs is a promising avenue to achieve the high metrics of next generation ultra-dense networks. This paper investigates the joint user-to-cloud association and beamforming problem in MC-RAN system. We propose an efficient iterative algorithm, which can be implemented in a distributed fashion across the multiple clouds. The numerical simulations show that our proposed distributed implementation achieves comparable performance to the centralized implementation. Compared to the state-of-the-art, our algorithm shows much better performance than scheduling-only scheme, and comparable performance to the idealized single cloud model.

Acknowledgments

Hayssam Dahrouj would like to thank Effat University in Jeddah, Saudi Arabia, for funding the research reported in this paper through the Research and Consultancy Institute.

References

[1] L. Zhang, Y. Liang, and D. Niyato, "6G Visions: Mobile Ultra-Broadband, Super Internet-of-Things, and Artificial Intelligence," *China Communications*, vol. 16, no. 8, pp. 1–14, Aug. 2019.
[2] *Cloud Radio Access Networks: Principles, Technologies, and Applications*. Cambridge University Press, 2017.
[3] B. Dai and W. Yu, "Sparse Beamforming and User-Centric Clustering for Downlink Cloud Radio Access Network," *IEEE Access*, vol. 2, pp. 1326–1339, 2014.
[4] J. Tang, W. P. Tay, T. Q. S. Quek, and B. Liang, "System cost minimization in cloud RAN with limited fronthaul capacity," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 3371–3384, May 2017.
[5] S. Park, O. Simeone, O. Sahin, and S. Shamai, "Inter-cluster design of precoding and fronthaul compression for cloud radio access networks," *IEEE Wireless Commun. Lett.*, vol. 3, no. 4, pp. 369–372, Aug. 2014.
[6] O. Dhifallah, H. Dahrouj, T. Y. Al-Naffouri, and M. S. Alouini, "Distributed robust power minimization for the downlink of multi-cloud radio access networks," in *2016 IEEE Global Communications Conference (GLOBECOM)*, Dec. 2016, pp. 1–6.
[7] H. Dahrouj, T. Y. Al-Naffouri, and M. S. Alouini, "Distributed cloud association in downlink multicloud radio access networks," in *2015 49th Annual Conference on Information Sciences and Systems (CISS)*, Mar. 2015, pp. 1–3.
[8] L. Liu and W. Yu, "Cross-Layer Design for Downlink Multihop Cloud Radio Access Networks With Network Coding," *IEEE Trans. Signal Process.*, vol. 65, no. 7, pp. 1728–1740, Apr. 2017.
[9] A. Douik, H. Dahrouj, T. Y. Al-Naffouri, and M. S. Alouini, "Distributed hybrid scheduling in multi-cloud networks using conflict graphs," *IEEE Trans. Commun.*, vol. 66, no. 1, pp. 209–224, Jan. 2018.
[10] K. Shen and W. Yu, "Fractional Programming for Communication Systems Part I: Power Control and Beamforming," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2616–2630, May 2018.
[11] ——, "Fractional Programming for Communication Systems Part II: Uplink Scheduling via Matching," *IEEE Trans. Signal Process.*, vol. 66, no. 10, pp. 2631–2644, May 2018.
[12] C. Pan, H. Ren, M. Elkashlan, A. Nallanathan, and L. Hanzo, "The Non-Coherent Ultra-Dense C-RAN Is Capable of Outperforming Its Coherent Counterpart at a Limited Fronthaul Capacity," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 11, pp. 2549–2560, Nov. 2018.