Cost-Effective Resource Allocation in C-RAN with Mobile Cloud

Kezhi Wang[1], Kun Yang[1], Xinhou Wang[2] and Chathura Sarathchandra Magurawalage[1]
[1]School of Computer Sciences and Electrical Engineering, University of Essex, CO3 4HG, Colchester, U.K.
E-mails: {kezhi.wang, kunyang, csarata}@essex.ac.uk
[2]Service Computing Technology and System Lab, Cluster and Grid Computing Lab
School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China.
E-mail: xwang@hust.edu.cn

***Abstract*—Taking full advantages of two cloud-based techniques, i.e., cloud radio access network (C-RAN) and mobile cloud computing (MCC), mobile operators will be able to provide the good service to the mobile user as well as increasing their revenue. This paper aims to minimize the mobile operator's cost while at the same time, meet the task time constraints of the mobile users. In particular, we assume that the mobile cloud first completes the tasks for the mobile user and then transmits the results back to the users through C-RAN. Joint cost-effective resource allocation is proposed between MCC and C-RAN and simulation results confirm that the proposed cost minimization and resource allocation solution outperforms non-optimal solutions.**

Intex Terms - C-RAN, Cost-Effective Resource Allocation, MCC.

I. Introduction

Nowadays, smart phones have been gaining more and more popularity. The most impressive part of the smart phones is that they can run attractive and interactive applications for human beings such as online gaming, facial recognition, multimedia applications, high definition video, etc. Those applications are always resource-hungry and will be inevitably increasing the computation burden of the mobile device. Thus, the mobile device always offload those task to base station via wireless network. In traditional cellular networks, each base station (BS) transmits data signal separately to the mobile user, where the transmission signal may be affected by the interference. Cooperative communication has been proposed to mitigate and combat the deleterious effects of interference, but the total energy cost of the cooperative communication may be still a little bit high [1], [2]. Recently, cloud radio access network (C-RAN) has been presented and soon received much attention from both academia and industry [3], [4]. C-RAN divides the traditional BS into several remote radio heads (RRHs), the baseband unit (BBU) pool, and the high-bandwidth fronthaul connecting RRH to the BBU pool.

Another cloud based technology, i.e., mobile cloud computing (MCC) [5], [6], which is inspired by integrating the popular cloud computing into mobile environment, can enable mobile user with increasing computing demands but limited computing resource offloading tasks to the powerful computation platforms in the cloud. Reference [7] studies how to maximize the system throughput with constraints on the response latency of each MCC user. Reference [8] has shown that computing resources and communication resources can be coupled for enhancing connected devices. Also, software defined network (SDN) has been proposed to offer scalable and flexible management with a logical centralized control model to MCC [9], [10].

Moreover, the price of the bandwidth and the computation resource is a critical component of the MCC, as it directly affects mobile operator's revenue and UE's budget [11]. Pursuing computational intensive or high bandwidth tasks by the UE increases the operating expense and capital expenditure of the mobile operators, which drastically reduce their profit and make them face a very hard situation [11]. Thus, how to save the whole system's cost is of huge importance and interest in the operators' eyes. Pricing has been widely studied in both mobile wireless networks [12]–[15] and conventional cloud computing [11], [16], [17]. For instance, [12] leverages pricing to enable efficient resource management in wireless ad hoc network while [14] uses pricing as a lever to enable technology adoption in communication market. In MCC, [17] designs a computationally efficient pricing to enable fair competition for cloud resource and [16] develops optimized fine-grained pricing to satisfy both cloud customers and cloud service providers with maximized utility.

In this paper, we aim to minimize the mobile operator's cost under the time constraints of the given task from the mobile users. We assume that the mobile cloud is responsible for the computational intensive task while the BBU is in charge of returning the execution results to the UE via RRHs. Joint cost-effective resource allocation is proposed between MCC and C-RAN. In particular, we model the mobile operator's cost in two sides, i.e., executing the task in mobile cloud plus the cost in transmitting the results back to UE through RRHs. We assume the mobile operator has to meet the users' task time requirement. We formulate the joint cost minimization into a non-convex optimization, which is NP-hard. By converting it to the equivalent weighted minimum mean square error (WMMSE) and using the iterative algorithm, we can successfully address the joint resource allocation between the mobile cloud and C-RAN. Simulation results confirm that the proposed cost minimization and resource allocation solution outperforms non-optimal solutions.

978-1-4799-6664-6/16/$31.00 ©2016 IEEE

The remainder of this paper is organized as follows. Section II introduces the system model. Section III presents the cost model and problem formulation, while Section IV introduces the joint cost minimization solution in mobile cloud and mobile network. Simulation results are shown in Section V, followed by conclusion in Section VI and the acknowledgement in VII.

II. System Model

In this section, the mathematical models for the MCC as well as the C-RAN are presented. The whole system architecture is given by Fig. 1.

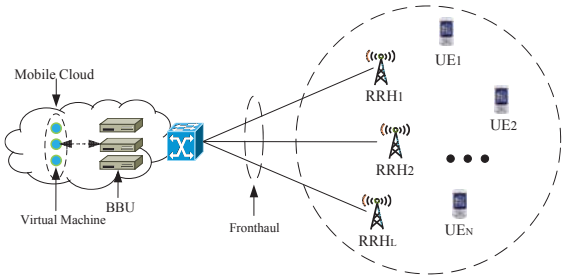

Fig. 1. A mobile cloud computing system with cloud radio access network assisted.

A. Task Model

We consider there are $\mathcal{N} = \{1, 2, ..., N\}$ UEs, each with one antenna, deployed in the C-RAN, as shown in Fig. 1. Consider there are $\mathcal{L} = \{1, 2, ..., L\}$ RRHs, each of which has $K \geq 1$ antennas, connecting to the BBU pool through high-speed fiber fronthaul link. We consider that the mobile cloud is co-located with the BBU and each mobile user has one task to be completed in the cloud. After the task execution completion, the mobile cloud will transmit the computation results back to the mobile user through C-RAN. Similar with [6] and [18], we assume that each of UE i has the computational intensive task U_i in the mobile cloud as follows:

$$U_i = (F_i, D_i),\ i = 1, 2, ..., N \tag{1}$$

where F_i is the total number of the CPU cycles to be accomplished for this computational task U_i to the i-th UE, while D_i describes the whole output date size transmitting to the i-th UE through C-RAN after task execution, including the task's output parameter and the calculation results, etc. D_i and F_i can be obtained by using the approaches provided in [19].

We assume that all the channel state information (CSI) are available in the BBU pool, which facilitate interference cancelation and signal cooperation. We do not consider the time and cost consumption in the process where the UE transmits the indication signal and configuration information to the mobile cloud to instruct the task to be executed. Also, we do not consider the time and cost in the fronthaul link, but we will consider the the fronthaul constraints by using the transmitting data rate.

B. Network Model

C-RAN wireless channel is responsible for returning the calculation results from the cloud to the mobile user. The received signal at the UE i under the complex baseband equivalent channel is given by

$$y_i = \sum_{j\in\mathcal{C}} \mathbf{h_{ij}}^H \mathbf{v_{ij}} x_i + \sum_{k\neq i}^{N} \sum_{j\in\mathcal{C}} \mathbf{h_{ij}}^H \mathbf{v_{kj}} x_k + \sigma_i, \quad i = 1, 2, ..., N \tag{2}$$

where x_i denotes the transmission data for the ith UE with $E\{|x_i|^2\} = 1$, $\mathcal{C} \subseteq \mathcal{L}$ is the set of serving RRHs to the UE i, $\mathbf{h_{ij}} \in \mathbb{C}^{K\times 1}$ denotes the channel vector from RRH j to UE i, while σ_i denotes the white Gaussian noise which is assumed to be distributed as $\mathcal{CN}(0, \sigma_i^2)$. Denote $\mathbf{v_{ij}} \in \mathbb{C}^{K\times 1}$ as the transmitting beamforming vector from RRH j to UE i. Therefore, the signal-to-interference-plus-noise ratio (SINR) can be expressed by

$$\text{SINR}_i = \frac{|\sum_{j\in\mathcal{C}} \mathbf{v_{ij}}^H \mathbf{h_{ij}}|^2}{\sum_{k\neq i}^{N} |\sum_{j\in\mathcal{C}} \mathbf{v_{kj}}^H \mathbf{h_{kj}}|^2 + \sigma^2},\ \ i = 1, 2, ..., N. \tag{3}$$

Then, the system capacity and the achievable rate for UE i can be given as

$$r_i = B_i \log\left(1 + \text{SINR}_i\right),\ \ i = 1, 2, ..., N \tag{4}$$

where B_i is the wireless channel bandwidth assigning to UE i and we assume is it fixed in this paper. The time in transmitting the execution results back to UE i can be written as

$$T_i^{Tr} = \frac{D_i}{r_i},\ i = 1, 2, ..., N. \tag{5}$$

Also, we assume that each RRH j has its own power constraint P_j as follows:

$$\sum_{i=1}^{N} |\mathbf{v_{ij}}|^2 \leq P_j, \quad j = 1, 2, ..., L. \tag{6}$$

C. Fronthaul Model

The fronthaul link is responsible of carrying the data from the BBU to the RRH and its consumption can be explained as the accumulated data rates of the users served by RRHs. References [20], [21] use $\boldsymbol{l}0$-norm to model the j-th fronthaul capability as

$$C_j = \sum_{i=1}^{N} ||\mathbf{v_{ij}}|^2|_0 \cdot r_i, \quad j = 1, 2, ..., L \tag{7}$$

where $||\mathbf{v_{ij}}|^2|_0$ denotes the $\boldsymbol{l}0$-norm of the beamforming vector $|\mathbf{v_{ij}}|^2$ and can be mathematically expressed as

$$||\mathbf{v_{ij}}|^2|_0 = \begin{cases} 0, \text{if } |\mathbf{v_{ij}}|^2 = 0 \\ 1, \text{otherwise} \end{cases}. \tag{8}$$

One can see that $\boldsymbol{l}0$-norm is the number of nonzero entries in the vector. Therefore, the j-th fronthaul constraint can be modeled as the maximum data rates $C_{j,max}$ as

$$C_j \leq C_{j,max}, \quad j = 1, 2, ..., L. \tag{9}$$

D. Computation Model and QoS Requirement

The mobile cloud is responsible of completing the task for the mobile user. Assume the time spent to complete the task U_i is as follows:

$$T_i^C = \frac{F_i}{f_i}, \quad i = 1, 2, ..., N \tag{10}$$

where f_i is the assigned computation capability of the i-th virtual machine serving UE i in the mobile cloud. Also, we assume that the constraint of the sum of all the virtual machines is given by

$$\sum_{i=1}^{N} f_i \leq f_{max} \tag{11}$$

where f_{max} is the maximum computation capacity of the whole mobile cloud and can be allocated to different virtual machine per demand.

The QoS can be given as the whole time for completing the task and returning the results back to the mobile user. We define the total time spent in executing and transmitting the task results to UE i as

$$T_i = T_i^{Tr} + T_i^C, i = 1, 2, ..., N. \tag{12}$$

We assume that the task has to be accomplished in time constraints $T_{i,max}$ in order to satisfy the mobile user's requirement, thus the time (QoS) constraint can be given as

$$T_i \leq T_{i,max}, i = 1, 2, ..., N. \tag{13}$$

III. Cost and Problem Formulation

In this section, we introduce the model for the expenditure cost of the mobile operators and provide the cost-effective resource allocation problem formulation. Our design aims to minimize the expenditure cost while satisfying the time constraints. In other words, mobile operator can maximize their profits by meeting the requirements of all the users.

We define the cost model for the i-th virtute machine as follows

$$Cost_i^C = \kappa^C \cdot f_i, i = 1, 2, ..., N \tag{14}$$

where $\kappa^C \geq 0$ is the mobile operator's cost for each computation unit (e.g., CPU cycle) per second in the mobile cloud. We take the popular cloud service provider Amazon EC2 [22] as an example to show how to calculate κ^C. If we look at i2.xlarge On-demand Instance on Amazon EC2, which contains 14 ECU (each ECU providing the equivalent CPU capacity of a 1.0-1.2 GHz 2007 Opteron or 2007 Xeon processor), and the hourly price is \$0.853. Thus we can roughly calculate the price of each CPU cycle per second κ^C as $\$1.7 * 10^{-14}$. Although Amazon EC2 is the conventional cloud computing provider which is different from the MCC, its pricing still can reflect the cost of computation capacity of MCC as they have similar hardware resources (e.g., CPU, memory, storage).

Also, the cost for transmitting the calculation results from the mobile cloud to each UE i via the C-RAN can be given as

$$Cost_i^{Tr} = \kappa^{Tr} \cdot r_i, i = 1, 2, ..., N. \tag{15}$$

where $\kappa_i^{Tr} \geq 0$ is the price for the data rate (bit/s) in wireless network. We take one of the largest mobile services provider i.e., China Mobile [23] as an example to show how to calculate κ^{Tr}. If we look at one of the goody-bags of China Mobile, which contains 11 GigaBytes resource and the monthly price is \$45.6. Thus we can roughly calculate the price for data rate κ^{Tr} as $\$1.6 * 10^{-15}$.

Therefore, the whole cost in executing i-th task can be written as

$$Cost_i = Cost_i^C + Cost_i^{Tr}, i = 1, 2, ..., N \tag{16}$$

Then the joint cost minimization optimization problem can be given as

$$\begin{aligned} \mathcal{P}1: \ & \underset{f_i, r_i, \mathbf{v_{ij}}, \mathcal{C}}{\text{minimize}} \quad \sum_{i=1}^{N} Cost_i \\ & \text{subject to}: (6), (9), (11), (13). \end{aligned} \tag{17}$$

IV. Cost-effective resource allocation algorithm

In this section, we aim to provide the joint cost-effective resource allocation in C-RAN with mobile cloud. The objective is to minimize the total cost for the mobile operators (maximize their profits) by allocating the resource jointly between mobile cloud and mobile network.

A. Achievable Rates

From (3) and (4), one can get the achievable rate for i-th UE as

$$\begin{aligned} & r_i = B_i \log \left(1 + \frac{|\sum_{j \in \mathcal{C}} \mathbf{h_{ij}}^H \mathbf{v_{ij}}|^2}{\sum_{k=1,\ k \neq i}^{N} |\sum_{j \in \mathcal{C}} \mathbf{h_{ij}}^H \mathbf{v_{kj}}|^2 + \sigma^2} \right), \\ & i = 1, 2, ..., N. \end{aligned} \tag{18}$$

If one ignores the interference term $\sum_{k=1,\ k \neq i}^{N} |\sum_{j \in \mathcal{C}} \mathbf{h_{ij}}^H \mathbf{v_{kj}}|^2$ and use (6), as well as apply Cauchy-Schwarz inequality, one may get

$$r_i \leq R_{i,max}, i = 1, 2, ..., N, \tag{19}$$

where

$$\begin{aligned} R_{i,max} &= B_i \log \left(1 + \frac{\sum_{j \in \mathcal{C}} |\mathbf{h_{ij}}^H|^2 \sum_{j \in \mathcal{C}} |\mathbf{v_{ij}}|^2}{\sigma^2} \right) \\ &\leq B_i \log \left(1 + \frac{\sum_{j \in \mathcal{C}} |\mathbf{h_{ij}}^H|^2 P_j}{\sigma^2} \right). \end{aligned} \tag{20}$$

By using $T_{i,max} > 0$, $f_i > 0$, (12) and (13), one can get

$$T_{i,max} \geq T_i^{Tr} + T_i^C = \frac{D_i}{r_i} + \frac{F_i}{f_i}. \tag{21}$$

Then, one can have the minimum computational capacity for i-th cloud as

$$f_i \geq f_{i,min}, \quad i = 1, 2, ..., N \tag{22}$$

where

$$f_{i,min} = \frac{F_i}{T_{i,max} - \frac{D_i}{R_{i,max}}}, \quad i = 1, 2, ..., N. \tag{23}$$

Also, one can get the maximum computational capacity allocated to i-th cloud as

$$f_{i,max} = f_{max} - \sum_{k=1, k\neq i}^{N} f_{k,min}, \quad i = 1, 2, ..., N. \tag{24}$$

Therefore, the minimum achievable rate can be given by

$$r_i \geq R_{i,min} \tag{25}$$

where

$$R_{i,min} = \frac{D_i}{T_{i,max} - \frac{F_i}{f_{i,max}}}. \tag{26}$$

As the arbitrary phase rotation of the beamforming vectors $\mathbf{v_{ij}}$ does not affect (25), it can be rewritten as a second-order cone (SOC) constraint as [24]

$$\begin{aligned} &\sqrt{1 - \frac{1}{2^{\frac{R_{i,min}}{B_i}}}} \sqrt{\sum_{k=1}^{N} |\sum_{j\in\mathcal{C}} \mathbf{h_{ij}}^H \mathbf{v_{kj}}|^2 + \sigma^2} \\ &\leq \text{Re}\left(|\sum_{j\in\mathcal{C}} \mathbf{h_{ij}}^H \mathbf{v_{ij}}|^2 \right), \quad i = 1, 2, ..., N. \end{aligned} \tag{27}$$

Moreover, to achieve the minimal value, the equality will hold for the time constraints of $\mathcal{P}1$ as

$$T_{i,max} = T_i^{Tr} + T_i^C = \frac{D_i}{r_i} + \frac{F_i}{f_i}. \tag{28}$$

Then, one can use the data rate r_i to replace the computational capacity in the objective function of $\mathcal{P}1$ by

$$f_i = \frac{F_i}{T_{i,max} - \frac{D_i}{r_i}}. \tag{29}$$

Also, according to [25], the non-convex $\boldsymbol{l}0$-norm can be approximated by a convex reweighed $\boldsymbol{l}1$-norm as $|\mathbf{V}|_0 = \sum_{k=1}^{N} \rho_k |v_k|$, where v_k is the k-th element of the vector $\mathbf{V}$ and ρ_k is the corresponding weight. Following reference [21] and using (7) and (9), the fronthaul constraints can be written as

$$C_j = \sum_{i=1}^{N} \rho_{ij} \left|\mathbf{v_{ij}}\right|^2 \cdot r_i \leq C_{j,max}, \quad j = 1, 2, ..., L \tag{30}$$

where

$$\rho_{ij} = \frac{1}{\left|\mathbf{v_{ij}}\right|^2 + \epsilon} \quad j = 1, 2, ..., L. \tag{31}$$

and ϵ is a small positive factor to ensure stability. Then by using (27), (29) and (30), $\mathcal{P}1$ can be transformed to

$$\begin{aligned} &\underset{r_i, \mathbf{v_{ij}}, \mathcal{C}}{\text{minimize}} \quad \phi(r_i) \\ &\text{subject to} : (6), (27), (30) \end{aligned} \tag{32}$$

where $\phi(r_i) = \sum_{i=1}^{N} \kappa^C \cdot \varphi_i(r_i) + \kappa^{Tr} \cdot r_i$ and $\varphi_i(r_i) = \frac{F_i \cdot r_i}{T_{i,max} \cdot r_i - D_i}$.

B. WMMSE-based Solution

By using (20), one can get the upper bound of the objective function of (32) as

$$\phi(r_i) \leq \kappa^C \cdot \varphi_i(r_i) + B_i \frac{\kappa^{Tr}}{\sigma^2} \sum_{j\in\mathcal{C}} |\mathbf{h_{ij}}^H|^2 \sum_{j\in\mathcal{C}} |\mathbf{v_{ij}}|^2. \tag{33}$$

As minimizing the objective function of (32) can be relaxed to minimizing its upper bound, (32) can be transformed to

$$\begin{aligned} &\underset{r_i, \mathbf{v_{ij}}, \mathcal{C}}{\text{minimize}} \quad \sum_{i=1}^{N} \kappa^C \cdot \varphi_i(r_i) + B_i \frac{\kappa^{Tr}}{\sigma^2} \sum_{j\in\mathcal{C}} |\mathbf{h_{ij}}^H|^2 \sum_{j\in\mathcal{C}} |\mathbf{v_{ij}}|^2 \\ &\text{subject to} : (6), (27), (30). \end{aligned} \tag{34}$$

Inspired by the solutions provided in [26], [21], [27] and [28], one can use the WMMSE-based method to solve (34). Assume that the receive beamforming vector in mobile user i as $\mathbf{u_i} \subseteq \mathbb{C}^{1\times 1}$, as there is only one antenna in the UE. Also, denote $\mathbf{v_i} = [\mathbf{v_{i1}}, \mathbf{v_{i2}}, ..., \mathbf{v_{iC}}]^H$, $\mathbf{h_i} = [\mathbf{h_{i1}}, \mathbf{h_{i2}}, ..., \mathbf{h_{iC}}]^H$ for notation simplification. Thus, the corresponding mean square error (MSE) at UE i can be given as

$$\begin{aligned} e_i &= E\left[(\mathbf{u_i} y_i - x_i)(\mathbf{u_i} y_i - x_i)^H\right] \\ &= \sum_{i=1}^{N} \mathbf{u_i}^H (\mathbf{h_i}^H \mathbf{v_i} \mathbf{v_i}^H \mathbf{h_i} + \sigma_i^2) \mathbf{u_i} - 2\,\text{Re}\left[\mathbf{u_i}^H \mathbf{h_i}^H \mathbf{v_i}\right] + 1. \end{aligned} \tag{35}$$

Thus, one can rewrite (34) as

$$\begin{aligned} \mathcal{P}2: \ &\underset{\omega_i, \mathbf{v_{ij}}, \mathbf{u_i}, \mathcal{C}}{\text{minimize}} \quad \sum_{i=1}^{N} \kappa^C (\omega_i e_i + \psi_i(\gamma_i(\omega_i)) - \omega_i \gamma_i(\omega_i)) + \\ &B_i \frac{\kappa^{Tr}}{\sigma^2} \sum_{j\in\mathcal{C}} |\mathbf{h_{ij}}^H|^2 \sum_{j\in\mathcal{C}} |\mathbf{v_{ij}}|^2 \\ &\text{subject to} : (6), (27), (30) \end{aligned} \tag{36}$$

where $\psi_i(e_i) = \varphi_i(-B_i \log(e_i))$ and $\gamma_i(\cdot)$ is the inverse mapping of the gradient map $\frac{\partial \psi_i(e_i)}{\partial e_i}$. One can see that $\psi_i(e_i)$ is a strictly concave function in $\mathcal{P}2$, as $\varphi_i(r_i)$ is the decreasing utility function of the data rate r_i. One can see that $\mathcal{P}2$ is convex with respect to each of the individual variables ω_i, $\mathbf{v_{ij}}$ and $\mathbf{u_i}$. Therefore, one can use the block coordinate descent method to solve it [26], [21], [27], [28]. The process to solve $\mathcal{P}2$ is as follows:

Step 1: By fixing all the transmit beamforming vector $\mathbf{v_i}$, the optimal receive beamforming vector can be obtained by the well-known minimum mean square error (MMSE) receiver as

$$\mathbf{u_i} = \left(\mathbf{h_i}^H \mathbf{v_i}\right) \cdot \left(\sum_{k=1}^{N} \mathbf{h_i}^H \mathbf{v_k} \mathbf{v_k}^H \mathbf{h_i} + \sigma_i^2\right)^{-1}. \tag{37}$$

Step 2: By fixing all the transmit beamforming vector $\mathbf{v_i}$ and the MMSE receiver $\mathbf{u_i}$, the corresponding optimal MSE weight ω_i can be given by

$$\omega_i = \frac{\partial \psi_i(e_i)}{\partial e_i} = \frac{B_i D_i F_i \ln(2)}{e_i (B_i T_{i,max} \ln(e_i) + D_i \ln(2))^2}. \tag{38}$$

Step 3: By fixing all the optimal MSE weight ω_i and MMSE receiver $\mathbf{u_i}$, the optimal transmit beamforming vector $\mathbf{v_i}$ can be calculated by solving the following SOCP problem as

$$\begin{aligned} &\underset{r_i,\mathbf{v_{ij}},\mathcal{C}}{\text{minimize}} \quad \tau \\ &\text{subject to}: (6), (27), (30) \end{aligned} \tag{39}$$

where $\tau = \sum_{i=1}^{N} \kappa^C \omega_i e_i + B_i \frac{\kappa^{Tr}}{\sigma^2} \sum_{j\in\mathcal{C}} |\mathbf{h_{ij}}^H|^2 \sum_{j\in\mathcal{C}} |\mathbf{v_{ij}}|^2$. Therefore, one can use the algorithm 1 to address $\mathcal{P}2$, where ε is a small constant to guarantee convergence.

Algorithm 1 Iterative algorithm for $\mathcal{P}2$

Initialize: $n = 1$, $\rho_{ij}^{(0)}$, $\mathbf{v_{ij}}^{(0)}$, $r_i^{(0)}$, $i = 1, 2, ..., N$, $j = 1, 2, ..., L$;
Repeat:
1: Obtain the receive beamforming vector $\mathbf{u_i}^{(n)}$ according to (37) by fixing $\mathbf{v_{ij}}^{(n-1)}$;
2: Obtain the MSE weight ω_i according to (38) by fixing $\mathbf{v_{ij}}^{(n-1)}$ and $\mathbf{u_i}^{(n)}$;
3: Obtain the transmit beamforming vector $\mathbf{v_{ij}}^{(n)}$ according to SOCP (39) by fixing $\omega_i^{(n)}$, $\mathbf{u_i}^{(n)}$;
4: Update $r_i^{(n+1)} = r_i^{(n)}$ according to (4);
5: Update $\rho_{ij}^{(n+1)} = \rho_{ij}^{(n)}$ according to (31);
6: Update $n = n + 1$;
Until $|\tau^{(n+1)} - \tau^{(n)}| < \varepsilon$
Return: RRH cluster $\mathcal{C}$, $\mathbf{v_{ij}}$, r_i, for $i = 1, 2, ..., N$, $j = 1, 2, ..., L$.

C. Computation Capacity Constraints

By using (29) and r_i from Algorithm 1, f_i for $i = 1, 2, ..., N$ can be obtained. Next, we have to check if those f_i meet the overall capacity constraint by using

$$\sum_{i=1}^{N} f_i \leq f_{max}. \tag{40}$$

If the sum of computational capacity of all the victual machines is greater than the total computation of the mobile data center, we propose to normalize f_i according to

$$f'_{i,max} = \frac{f_i}{\sum_{i=1}^{N} f_i} f_{max}, \quad i = 1, 2, ..., N. \tag{41}$$

Then, we set the minimal data rate for the i-th user as

$$r_i \geq R'_{i,min}, \quad i = 1, 2, ..., N \tag{42}$$

where

$$R'_{i,min} = \frac{D_i}{T_{i,max} - \frac{F_i}{f'_{i,max}}}, \quad i = 1, 2, ..., N. \tag{43}$$

We assume f_{max} is enough to make sure $R'_{i,min} \geq 0$ after normalization. Therefore, the optimization problem becomes

$$\begin{aligned} \mathcal{P}3: \quad &\underset{r_i,\mathbf{v_{ij}},\mathcal{C}}{\text{minimize}} \quad \sum_{i=1}^{N} \kappa^C \cdot \varphi_i(r_i) + \\ &B_i \frac{\kappa^{Tr}}{\sigma^2} \sum_{j\in\mathcal{C}} |\mathbf{h_{ij}}^H|^2 \sum_{j\in\mathcal{C}} |\mathbf{v_{ij}}|^2 \\ &\text{subject to}: (6), (30), (42). \end{aligned} \tag{44}$$

Then, the overall cost-effective resource allocation can be given by Algorithm 2.

Algorithm 2 Overall algorithm for joint optimization problem

1: Obtain the data rate r_i, for $i = 1, 2, ..., N$, by using **Algorithm 1**;
2: Obtain the computation capacity f_i according to (29);
3: Check the total cloud computation capacity by using (40);
4: **if** feasible **then**
  stop and go to **return**
5: **else** set $f'_{i,max} = \frac{f_i}{\sum_{i=1}^{N} f_i} f_{max}, \quad i = 1, 2, ..., N.$
6: Obtain the data rate r_i by solving $\mathcal{P}3$, for $i = 1, 2, ..., N$;
7: Obtain the computation capacity f_i according to (29);
8: **end if**
9: **Return**: $\mathbf{v_{ij}}$, f_i, r_i, for $i = 1, 2, ..., N$, $j = 1, 2, ..., L$.

V. Simulation Results

In this section, simulation results are provided to show the effectiveness of the cost-effective resource allocation. Matlab with CVX tool [29] has been used in the simulation. The simulation environment is shown as Fig. 2, in which we

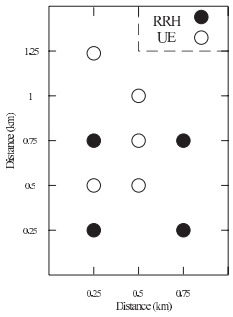

Fig. 2. C-RAN network with $L = 4$ RRHs and $N = 5$ UEs.

consider the C-RAN network with $L = 4$ RRHs, each equipped with $K = 2$ antennas. Also, we assume that there are $N = 5$ mobile users, each of which has only one antenna. We assume there is the mobile cloud co-located with the BBUs, and each mobile virtual machine has the same software stack as its corresponding mobile users and can execute the task for the mobile user. Moreover, we assume that the maximum transmit power for each RRH is 1 W, while the maximum total computation capacity is set to 100G CPU cycles per second. Similar with [30], we model the path and penetration loss as

$$p(d) = 127 + 25\log10(d) \tag{45}$$

where d (km) is the propagation distance. Also, we model the small scale fading as independent circularly symmetric Gaussian process distributed as $\mathcal{CN}(0, 1)$, whereas the noise power spectral density is assumed to be -100 dBm/Hz. We assume that the cost for data rate (bit/s) is $\kappa^{Tr} = 1.6 * 10^{-15}$\$ while the cost for the computational capacity is $\kappa^c = 1.7 * 10^{-14}$\$. Also, we assume the wireless channel bandwidth as 100 MHz and the fronthaul capacity constraint as 100 Mbps. In Fig. 3, we compare the proposed joint cost minimization optimization with the separate cost minimization solutions. For the separate energy minimization, we set two time constraints as $T_i^{Tr} \leq T_{i,max}^{Tr}$ and $T_i^C \leq T_{i,max}^C$, where $T_{i,max}^{Tr} + T_{i,max}^C = T_{i,max}$. $T_{i,max} = 2$s and $D_i = 1$ Mbit are set in Fig. 3.

One can see that with the increase of the total computational resource of the task F_i (CPU cycles per second), the cost ($) rise correspondingly. One can also see that the joint energy minimization achieves the best performance, followed by the second best solution when setting $T_{i,max}^{Tr} = T_{i,max}/2$ in Fig. 3. The performance of $T_{i,max}^{Tr} = T_{i,max} * 3/4$ can be shown as the worst solution among all these solutions.

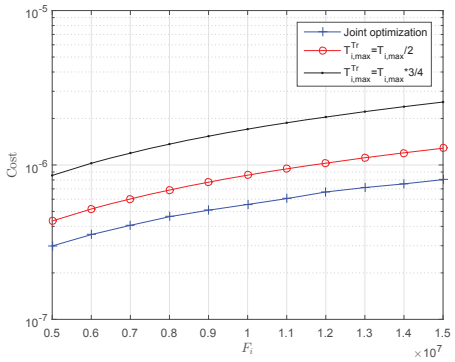

Fig. 3. Total cost vs. total number of CPU cycles with $T_{i,max} = 2$s.

VI. Conclusion

This paper has proposed a new architecture which allows the mobile operator having mobile cloud co-located with their C-RAN BBU. A joint cost-effective resource allocation algorithm in C-RAN with mobile cloud has been proposed. This algorithm can minimize the mobile operator's cost while at the same time, meet the task time constraints of the mobile users.

VII. Acknowledgement

This work was supported by UK EPSRC NIRVANA project (EP/L026031/1), EU Horizon 2020 iCIRRUS project (GA-644526) and EU FP7 Project CROWN (GA-2013-610524).

References

[1] K. Wang, Y. Chen, M.-S. Alouini, and F. Xu, "BER and optimal power allocation for amplify-and-forward relaying using pilot-aided maximum likelihood estimation," *IEEE Transactions on Communications*, vol. 62, no. 10, pp. 3462–3475, Oct 2014.

[2] K. Wang, Y. Chen, and M. Di Renzo, "Outage probability of dual-hop selective af with randomly distributed and fixed interferers," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 10, pp. 4603–4616, Oct 2015.

[3] X. Rao and V. Lau, "Distributed fronthaul compression and joint signal recovery in cloud-ran," *IEEE Transactions on Signal Processing*, vol. 63, no. 4, pp. 1056–1065, Feb 2015.

[4] C. M. R. Institute., "C-RAN white paper: The road towards green Ran. [online]," *(Jun. 2014)*, Available: http://labs.chinamobile. com/cran.

[5] S. Kosta, A. Aucinas, P. Hui, R. Mortier, and X. Zhang, "Thinkair: Dynamic resource allocation and parallel execution in the cloud for mobile code offloading," in *2012 IEEE Proceedings INFOCOM*, March 2012, pp. 945–953.

[6] K. Kumar and Y.-H. Lu, "Cloud computing for mobile users: Can offloading computation save energy?" *Computer*, vol. 43, no. 4, pp. 51–56, April 2010.

[7] Y. Cai, F. Yu, and S. Bu, "Cloud radio access networks (C-RAN) in mobile cloud computing systems," in *2014 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, April 2014, pp. 369–374.

[8] C. S. Magurawalage, K. Yang, and K. Wang, "Aqua computing: Coupling computing and communications," *arXiv:1510.07250*, pp. 1–19, October 2015.

[9] M. Dong, H. Li, K. Ota, and J. Xiao, "Rule caching in sdn-enabled mobile access networks," *IEEE Network*, vol. 29, no. 4, pp. 40–45, July 2015.

[10] J. Ding, R. Yu, Y. Zhang, S. Gjessing, and D. Tsang, "Service provider competition and cooperation in cloud-based software defined wireless networks," *IEEE Communications Magazine*, vol. 53, no. 11, pp. 134–140, November 2015.

[11] H. Xu and B. Li, "Dynamic cloud pricing for revenue maximization," *Cloud Computing, IEEE Transactions on*, vol. 1, no. 2, pp. 158–171, July 2013.

[12] Y. Xue, B. Li, and K. Nahrstedt, "Optimal resource allocation in wireless ad hoc networks: a price-based approach," *Mobile Computing, IEEE Transactions on*, vol. 5, no. 4, pp. 347–364, April 2006.

[13] P. Hande, M. Chiang, R. Calderbank, and S. Rangan, "Network pricing and rate allocation with content provider participation," in *INFOCOM 2009, IEEE*, April 2009, pp. 990–998.

[14] S. Sen, Y. Jin, R. Guerin, and K. Hosanagar, "Modeling the dynamics of network technology adoption and the role of converters," *Networking, IEEE/ACM Transactions on*, vol. 18, no. 6, pp. 1793–1805, Dec 2010.

[15] S. Ha, S. Sen, C. Joe-Wong, Y. Im, and M. Chiang, "Tube: Time-dependent pricing for mobile data," in *Proceedings of the ACM SIGCOMM 2012 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*, ser. SIGCOMM '12. New York, NY, USA: ACM, 2012, pp. 247–258. [Online]. Available: http://doi.acm.org/10.1145/2342356.2342402

[16] H. Jin, X. Wang, S. Wu, S. Di, and X. Shi, "Towards optimized fine-grained pricing of iaas cloud platform," *Cloud Computing, IEEE Transactions on*, vol. PP, no. 99, pp. 1–1, 2014.

[17] Q. Wang, K. Ren, and X. Meng, "When cloud meets ebay: Towards effective pricing for cloud computing," in *INFOCOM, 2012 Proceedings IEEE*, March 2012, pp. 936–944.

[18] X. Chen, "Decentralized computation offloading game for mobile cloud computing," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 4, pp. 974–983, April 2015.

[19] L. Yang, J. Cao, S. Tang, T. Li, and A. Chan, "A framework for partitioning and execution of data stream applications in mobile cloud computing," in *2012 IEEE 5th International Conference on Cloud Computing (CLOUD)*, June 2012, pp. 794–802.

[20] V. N. Ha and L. B. Le, "Joint coordinated beamforming and admission control for fronthaul constrained cloud-RANs," in *2014 IEEE Global Communications Conference (GLOBECOM)*, Dec 2014, pp. 4054–4059.

[21] 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.

[22] Amazon Elastic Compute Cloud (EC2). [Online]. Available: http://aws.amazon.com/ec2/, 2015.

[23] China Mobile. [Online]. Available: www.chinamobileltd.com/, 2015.

[24] A. Wiesel, Y. Eldar, and S. Shamai, "Linear precoding via conic optimization for fixed mimo receivers," *IEEE Transactions on Signal Processing*, vol. 54, no. 1, pp. 161–176, Jan 2006.

[25] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted l1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5-6, pp. 877–905, 2008.

[26] J. Tang, W. P. Tay, and T. Quek, "Cross-layer resource allocation with elastic service scaling in cloud radio access network," *IEEE Transactions on Wireless Communications*, vol. 14, no. 9, pp. 5068–5081, Sept 2015.

[27] Q. Shi, M. Razaviyayn, Z.-Q. Luo, and C. He, "An iteratively weighted mmse approach to distributed sum-utility maximization for a mimo interfering broadcast channel," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4331–4340, Sept 2011.

[28] S. Christensen, R. Agarwal, E. Carvalho, and J. Cioffi, "Weighted sum-rate maximization using weighted mmse for mimo-bc beamforming design," *IEEE Transactions on Wireless Communications*, vol. 7, no. 12, pp. 4792–4799, December 2008.

[29] M. Grant and S. Boyd., "CVX: Matlab software for disciplined convex programming, version 3.0," *(June 2015)*, Available: http://cvxr.com/cvx.

[30] Y. Shi, J. Zhang, and K. Letaief, "Group sparse beamforming for green cloud radio access networks," in *2013 IEEE Global Communications Conference (GLOBECOM),*, Dec 2013, pp. 4662–4667.