

# Joint Energy Minimization and Resource Allocation in C-RAN with Mobile Cloud

Kezhi Wang, Kun Yang, *Senior Member, IEEE*, and Chathura Sarathchandra Magurawalage

## Abstract

Cloud radio access network (C-RAN) has emerged as a potential candidate of the next generation access network technology to address the increasing mobile traffic, while mobile cloud computing (MCC) offers a prospective solution to the resource-limited mobile user in executing computation intensive tasks. Taking full advantages of above two cloud-based techniques, C-RAN with MCC systems are presented in this paper to enhance both performance and energy efficiencies. In particular, this paper studies the joint energy minimization and resource allocation in C-RAN with MMC under the time constraints of the given tasks. We first give the computational model and network model with the energy and time cost. Then, we formulate the joint energy minimization into a non-convex optimization with fronthaul rate constraints. An equivalent convex feasibility problem is reformulated and the iterative algorithm based on weighted minimum mean square error (WMMSE) is given to deal with the non-convexity. Simulation results confirm that the proposed energy minimization and resource allocation solution can improve the system performance and save energy.

**Intex Terms** - C-RAN, Joint Energy Minimization, MCC, Resource Allocation.

## I. INTRODUCTION

Nowadays, the number of smart devices and the corresponding mobile traffic have grown rapidly, which poses an increasingly high burden on the existing cellular network. It is predicted that the mobile device traffic will increase one thousand times and the cost is expected to decrease one hundred times by 2020, with the help of new network and computation paradigm [1]. Recently, more and more computational resource intensive tasks, such as multimedia applications,

Kezhi Wang, Kun Yang and Chathura Sarathchandra Magurawalage are with the School of Computer Sciences and Electrical Engineering, University of Essex, CO3 4HG, Colchester, U.K. (e-mails: {kwangj, kunyang, csarata}@essex.ac.uk).

high definition video playing and gaming appear in our daily life, make the load of both the mobile phone and the network, in terms of energy and bandwidth, increase hugely. Also, those types of applications have the trend of attracting more attention from the smartphone users.

In traditional cellular networks, each base station (BS) transmits data signal separately to the UE, so that the energy cost in the BS will be very high, in order to overcome the path loss and the interference from other BSs. Coordinated Multi-Point (CoMP) technique has been proposed to mitigate interference by using cooperation techniques, such as joint transmission (JT) and coordinated beamforming (CBF), between different BSs. However, CoMP technique sometimes cannot achieve the best performance, due to traditional X2 interface limitation, i.e., low bandwidth, high latency and inaccurate synchronization.

It is very fortunate that recently, a new promising network infrastructure, i.e., cloud radio access network (C-RAN), has been presented and soon received a large amount of attention in both academia and industry [2]. C-RAN is a cloud computing based, centralized, clean and collaborative radio access network [3]. It divides the traditional BS into three parts, namely, several remote radio heads (RRHs), the baseband unit (BBU) pool, and the high-bandwidth, high-speed, low latency fiber transport (or fronthaul) link connecting RRH to the BBU cloud pool. In C-RAN, most of the intensive network computational tasks, such as baseband signal processing, precoding matrix calculation, channel state information estimation are moved to BBU pool in the cloud, which is composed of numerous software defined virtual machines with the feature of dynamically configurable, scalable, sharable, re-allocatable per demand. On the other hand, RRH only needs to up-convert the received baseband signal from the BBU cloud and transmit them in the RF frequency band. In this case, RRHs with limited functions, only including A/D, D/A conversion, amplification, frequency conversion, make them very easy to distribute, according to the network requirement. Thanks to the separation of BBU and RRH and the cooperation between different BBUs, significant performance gain can be achieved in terms of efficient interference cancellation and management as well as the increase of network capacity and decrease of the energy cost.

Another very impressive technique, i.e. mobile cloud computing (MCC) has also attracted a huge number of interest recently [4], [5]. MCC is inspired by integrating the popular cloud computing into mobile environment, which enables that mobile user with increasing computing demands but limited computing resource can offload tasks to the powerful platforms in the cloud.

The reference [5] has investigated if the offloading operation to the cloud can save energy and extend battery lifetimes for UEs. The reference [6] has provided a theoretical framework of energy optimal mobile cloud computing under stochastic wireless channel while the reference [7] has proposed a game theoretical approach for achieving efficient computation offloading for MCC. Although the cloud computing has demonstrated the potential ability to improve the performance, in not only the MCC, but also C-RAN, the research of integration between them is rarely less. Fortunately, [8], [9] have shown that the combination of MCC and C-RAN is of huge interest.

Also, pursuing computational intensive or high bandwidth tasks in the UE side increases the operating expense and capital expenditure of the mobile operators, which drastically reduce their profit and make them face a very hard situation. It has been shown that the energy overhead or the electricity cost are among the most important factors in the overall operational expenditure [10]. Thus, how to save the whole system's energy is of huge importance and interest in the operators' eyes.

To address the above-mentioned questions, in this paper, we propose a novel C-RAN structure with the mobile clone (virtual machine) co-located with the BBU in the cloud pool. The mobile clone is responsible for the computational intensive task while the BBU is in charge of returning the execution results to the UE via RRHs. We aim to jointly reduce the total energy cost under the time constraints of the given task in C-RAN and mobile cloud. In particular, we model the energy cost in executing the task in mobile cloud and the energy cost in transmitting the results back to UE through RRHs. We also model the time spent in the mobile cloud and in wireless transmission process. Then we formulate the joint energy minimization into a non-convex optimization, which is NP-hard. By converting it to the equivalent weighted mean square error (WMMSE) and using the iterative algorithm, we can successfully address the joint resource allocation between the mobile cloud and C-RAN and deal with beamforming vector design in RRHs.

The remainder of this paper is organized as follows. Section II introduces the system model including the mobile cloud computational model and the network model. Section III presents the optimization problem formulation as well as two separate energy minimization solutions in mobile cloud and C-RAN, while Section IV introduces the joint energy minimization in mobile cloud and mobile network. Simulation results are shown in Section V, followed by concluding

remarks in Section VI.

## II. SYSTEM MODEL

In this section, the mathematical models for the mobile cloud computation as well as the C-RAN are presented. First, we introduce the concept of the mobile clone in MCC and the whole system design, and then we describe the computation models, including the energy and time consumption model in the cloud and in the network. Finally, the QoS requirement is given through the time constraint of the given task in the last subsection.

### A. Mobile Clone and System Architecture

We have noticed that when the mobile users encounter the computational intensive or high energy required tasks, they sometimes do not want to offload those tasks into the mobile cloud, as transmitting those program data to the cloud still costs some energy [5]. In some cases, it is even better to execute those tasks locally if transmission overhead is too high. Therefore, it is better to have the mobile user's computational tasks and corresponding data in the mobile cloud first. We can give the name of those mobile cloud with the user task and data on board as *mobile clone*. Mobile clone can be implemented by the cloud-based virtual machine which holds the same software stack, such as operating system, middleware, applications, as the mobile user. Then, if the mobile user wants to execute some task, it only needs to send the indication signal and the corresponding user configuration information to the mobile clone (virtual machine), which will execute those task on mobile user's behalf. In this case, the mobile user only needs to cost a small amount of energy and time overhead. After the task execution completion, the mobile clone will transmit the computation result data back to the mobile user through C-RAN. Another advantage of having mobile clone is that each mobile clone can talk to each other in the cloud without through the wireless link. In this case, each mobile user's communication can be possibly transferred into the communication between the mobile clones, thereby saving a great number of the wireless network resources as well as the energy and time overhead.

In this paper, we consider there are  $\mathcal{N} = \{1, 2, \dots, N\}$  UEs, each with one antenna, deployed in the C-RAN. Also, we consider there are  $\mathcal{L} = \{1, 2, \dots, L\}$  RRHs, each of which has  $K \geq 1$  antennas, connecting to the BBU pool through high-speed fiber fronthaul link, as shown in Fig. 1. We consider the case that each mobile user already has one specific mobile clone, established in

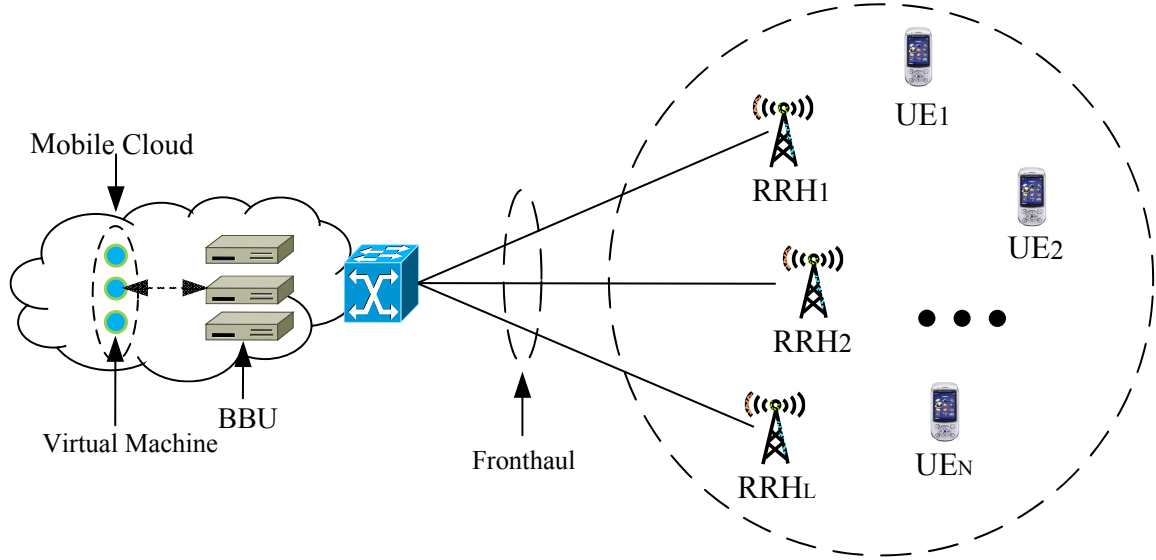


Fig. 1. A cloud radio access network with mobile cloud system.

the cloud, beside the BBU, and the mobile clone has the same software stack as its corresponding mobile user. Similar to [7] and [5], we assume that each of UE  $i$  has the computational intensive task  $U_i$  to be accomplished in the mobile clone as follows.

$$U_i = (F_i, D_i), \quad i = 1, 2, \dots, N \quad (1)$$

where  $F_i$  describes the total number of the CPU cycles needed to be completed for this computational task  $U_i$  for the  $i$ -th UE, while  $D_i$  denotes the whole size of the task's output data 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 [11].

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 energy consumption in which the UE transmits the indication signal and configuration information to the mobile clone to instruct the task to be executed. Also, we do not consider the time and energy consumption in the fronthaul link, but we will consider the the fonthaul constraints by using the transmitting data rate.

### B. Computation Model

In the mobile clone, the time spent to complete the task  $U_i$  is defined as follows

$$T_i^C = \frac{F_i}{f_i^C} \quad (2)$$

and the energy used in the  $i$ -th mobile clone is given as

$$E_i^C = \kappa_i^C (f_i^C)^{\nu_i^C - 1} F_i \quad (3)$$

where  $\kappa_i^C \geq 0$  is the effective switched capacitance,  $f_i^C$  is the computation capability of the  $i$ -th virtual machine serving UE  $i$  in the cloud and  $\nu_i^C \geq 1$  is the positive constant [12]. According to the realistic measurements,  $\kappa_i^C$  can be set to  $\kappa_i^C = 10^{-11}$  [13].

We also assume that different mobile clone may have different computational capacity and the constraint of the computation capacity  $f_i^C$  for the virtual machine is given by

$$f_i^C \leq f_{i,max}^C, \quad i = 1, 2, \dots, N \quad (4)$$

where  $f_{i,max}^C$  is the maximum computation capacity that the  $i$ -th virtual machine can achieve, as in the reality, the virtual machine cannot have unlimited computational capability.

### C. Network Model

After the mobile clone completes the task execution, the results will be returned to the mobile user through C-RAN. The received signal at the UE  $i$  under the complex baseband equivalent channel can be written as

$$y_i = \sum_{j \in \mathcal{C}} \mathbf{h}_{ij}^H \mathbf{v}_{ij} x_i + \sum_{k \neq i} \sum_{j \in \mathcal{C}} \mathbf{h}_{ij}^H \mathbf{v}_{kj} x_k + \sigma_i, \quad i = 1, 2, \dots, N \quad (5)$$

where  $\mathcal{C} \subseteq \mathcal{L}$  is the set of serving RRHs,  $\mathbf{h}_{ij} \in \mathbb{C}^{K \times 1}$  denotes the channel vector from RRH  $j$  to UE  $i$ , while  $\sigma_i$  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

$$\text{SINR}_i = \frac{|\sum_{j \in \mathcal{C}} \mathbf{v}_{ij}^H \mathbf{h}_{ij}|^2}{\sum_{k \neq i} \sum_{j \in \mathcal{C}} |\mathbf{v}_{kj}^H \mathbf{h}_{kj}|^2 + \sigma^2}, \quad i = 1, 2, \dots, N. \quad (6)$$

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

$$r_i = B_i \log(1 + \text{SINR}_i), \quad i = 1, 2, \dots, N \quad (7)$$

where  $B_i$  is the wireless channel bandwidth assigning to UE  $i$ .

The time cost in sending the execution results back to UE  $i$  from the RRHs is given by

$$T_i^{Tr} = \frac{D_i}{r_i} \quad (8)$$

where  $D_i$  is the returning data, introduced by the first subsection. Also, we can assume the power to send this task by RRHs is  $p_i$ , then the energy consumed by the serving RRHs is

$$E_i^{Tr} = p_i \cdot T_i^{Tr} = \frac{p_i D_i}{r_i} \quad (9)$$

where  $p_i$  can be given as  $p_i = \sum_{j \in \mathcal{C}} |\mathbf{v}_{ij}|^2$ . Also, we can assume that each RRH  $j$  has its own power constraint as follows

$$\sum_{i=1}^N |\mathbf{v}_{ij}|^2 \leq P_j, \quad j = 1, 2, \dots, L. \quad (10)$$

#### D. Fronthaul Constraints

The fronthaul link can carry the task results from the mobile clone to the UE through C-RAN. Reference [14] uses  $l_0$ -norm to model the  $j$ -th fronthaul capability as

$$\bar{C}_j = \sum_{i=1}^N ||\mathbf{v}_{ij}|^2|_0, \quad j = 1, 2, \dots, L \quad (11)$$

where  $||\mathbf{v}_{ij}|^2|_0$  denotes the  $l_0$ -norm of vector  $|\mathbf{v}_{ij}|^2$  and can be explained as the number of nonzero entries in the vector and can be mathematically expressed as

$$||\mathbf{v}_{ij}|^2|_0 = \begin{cases} 0, & \text{if } |\mathbf{v}_{ij}|^2 = 0 \\ 1, & \text{otherwise} \end{cases}. \quad (12)$$

One can see that the number of non-zeros elements of the transmitting beamforming vector  $|\mathbf{v}_{ij}|^2$  also indicates the number of data symbol streams, carried by the fronthaul link from BBU to RRH  $j$  for the  $i$ -th mobile user. Reference [14] also assume that each fronthaul link is only capable of carrying at most  $\bar{C}_{j,max}$  signals for UEs as

$$\bar{C}_j \leq \bar{C}_{j,max}, \quad j = 1, 2, \dots, L. \quad (13)$$

Reference [15] goes a step further and assume that the fronthaul consumption is the accumulated data rates of the users served by RRHs and model the  $j$ -th fronthaul capability as

$$C_j = \sum_{i=1}^N ||\mathbf{v}_{ij}|^2|_0 \cdot r_i, \quad j = 1, 2, \dots, L. \quad (14)$$

In this case, the  $j$ -th fronthaul constraint can be modeled as the maximum data rates which can be allowed to transmitting through BBU to  $j$ -th RRH as  $C_j \leq C_{j,max}$ . We also use this fronthaul constraint in our paper.

### E. QoS Requirement

The qualify of service (QoS) can be given as the whole time cost for completing the required 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

$$T_i = T_i^{Tr} + T_i^C. \quad (15)$$

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

$$T_i \leq T_{i,max} \quad (16)$$

Also, the whole energy cost in executing this task and transiting the results back to  $i$ -th UE can be given as

$$E_i = E_i^C + \eta_i E_i^{Tr} \quad (17)$$

where  $\eta_i \geq 0$  is a weight to trade off between the energy consumptions in the mobile cloud and the C-RAN, and it can be also explained as the inefficiency coefficient of the power amplifier at RRH.

## III. PROBLEM FORMULATION AND SEPARATE SOLUTIONS

In this section, we provide the energy minimization problem formulation. Our design aims to minimize the energy cost while satisfying the time constraints. First, we formulate the energy minimization for the mobile clone and then we give the energy minimization formulation for C-RAN with the fronthaul constraints. Two separate solutions are also provided for the energy minimization to the mobile clone and C-RAN, respectively.



### A. Energy Minimization for Mobile Clone

We assume the time constraints for completing the task in mobile clone as  $T_{i,max}^C$ , then the energy minimization optimization problem for the mobile clone can be given as

$$\begin{aligned} \mathcal{P}1 : \quad & \underset{f_i^C}{\text{minimize}} \quad \sum_{i=1}^N E_i^C \\ & \text{subject to } T_i^C \leq T_{i,max}^C, f_i^C \leq f_{i,max}^C, i = 1, 2, \dots, N. \end{aligned} \quad (18)$$

Assume  $f_i^{C*}$  as the optimum solution for problem  $\mathcal{P}1$ . Then, if  $f_i^{C*} \leq f_{i,max}^C$  for  $i = 1, 2, \dots, N$ , the equality holds for the first constraints. Thus, the optimal solution can be given by

$$f_i^{C*} = \frac{F_i}{T_{i,max}^C} \quad i = 1, 2, \dots, N. \quad (19)$$

If  $f_i^{C*} > f_{i,max}^C$ , we assume there is no solution for the above problem. Therefore, the whole energy cost is given by

$$\begin{cases} \sum_{i=1}^N \kappa_i^C \frac{F_i^3}{(T_{i,max}^C)^2}, & \text{if } f_i^{C*} \leq f_{i,max}^C \\ \text{no solution} & \text{if } f_i^{C*} > f_{i,max}^C. \end{cases} \quad (20)$$

### B. Energy Minimization for C-RAN

We assume the time constraints for transmitting the task results through C-RAN to UE  $i$  as  $T_{i,max}^{Tr}$ . Then, the energy minimization optimization problem for the C-RAN transmission can be given as

$$\begin{aligned} \mathcal{P}2 : \quad & \underset{\mathbf{v}_{ij}, r_i, C}{\text{minimize}} \quad \sum_{i=1}^N E_i^{Tr} \\ & \text{subject to } \sum_{i=1}^N |\mathbf{v}_{ij}|^2 \leq P_j, \quad j = 1, 2, \dots, L \\ & B_i \log \left( 1 + \frac{|\sum_{j \in \mathcal{C}} \mathbf{h}_{ij}^H \mathbf{v}_{ij}|^2}{\sum_{k=1, k \neq i}^N |\sum_{j \in \mathcal{C}} \mathbf{h}_{ij}^H \mathbf{v}_{kj}|^2 + \sigma^2} \right) \geq r_i, \quad i = 1, 2, \dots, N \\ & \sum_{i=1}^N |\mathbf{v}_{ij}|^2 \cdot r_i \leq C_{j,max}, \quad j = 1, 2, \dots, L. \\ & T_i^{Tr} \leq T_{i,max}^{Tr} \end{aligned} \quad (21)$$

Problem  $\mathcal{P}2$  is a non-convex optimization and NP-hard, which will be solved in the next section.

It is also of interest in exploring the power minimization optimization for C-RAN. In this case,

the equality holds for the last constraints of  $\mathcal{P}2$  and then, the minimum transmission data rate can be given by

$$r_i \geq \frac{D_i}{T_{i,max}^{Tr}}. \quad (22)$$

Therefore, the power minimization optimization can be written as

$$\begin{aligned} \mathcal{P}3: \quad & \underset{\mathbf{v}_{ij}, r_i, \mathcal{C}}{\text{minimize}} && \sum_{i=1}^N \sum_{j \in \mathcal{C}} |\mathbf{v}_{ij}|^2 \\ & \text{subject to} && \sum_{i=1}^N |\mathbf{v}_{ij}|^2 \leq P_j, \quad j = 1, 2, \dots, L \\ & && B_i \log \left( 1 + \frac{|\sum_{j \in \mathcal{C}} \mathbf{h}_{ij}^H \mathbf{v}_{ij}|^2}{\sum_{k=1, k \neq i}^N |\sum_{j \in \mathcal{C}} \mathbf{h}_{ij}^H \mathbf{v}_{kj}|^2 + \sigma^2} \right) \geq \frac{D_i}{T_{i,max}^{Tr}}, \quad i = 1, 2, \dots, N \\ & && \sum_{i=1}^N |\mathbf{v}_{ij}|^2 \cdot r_i \leq C_{j,max}, \quad j = 1, 2, \dots, L. \end{aligned} \quad (23)$$

As the arbitrary phase rotation of the beamforming vectors  $\mathbf{v}_{ij}$  does not affect  $\mathcal{P}3$ , the second constraint of  $\mathcal{P}3$  can be rewritten as a second-order cone (SOC) constraint as [16]

$$\sqrt{1 - \frac{1}{2^{\frac{B_i \cdot T_{i,max}^{Tr}}{D_i}}}} \sqrt{\sum_{k=1}^N |\sum_{j \in \mathcal{C}} \mathbf{h}_{ij}^H \mathbf{v}_{kj}|^2 + \sigma^2} \leq \text{Re} \left( \left| \sum_{j \in \mathcal{C}} \mathbf{h}_{ij}^H \mathbf{v}_{ij} \right|^2 \right), \quad i = 1, 2, \dots, N. \quad (24)$$

Also, according to [17], the non-convex  $l_0$ -norm can be approximated by a convex reweighted  $l_1$ -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  $\rho_k$  is the corresponding weight. Following reference [15], the last constraint in  $\mathcal{P}3$  can be rewritten as follows

$$C_j = \sum_{i=1}^N \rho_{ij} |\mathbf{v}_{ij}|^2 \cdot r_i \leq C_{j,max}, \quad j = 1, 2, \dots, L \quad (25)$$

where

$$\rho_{ij} = \frac{1}{|\mathbf{v}_{ij}|^2 + \epsilon} \quad j = 1, 2, \dots, L. \quad (26)$$

and  $\epsilon$  is a small positive factor to ensure stability. Then  $\mathcal{P}3$  can be transferred to

$$\begin{aligned}
\mathcal{P}4 : \quad & \underset{\mathbf{v}_{ij}, r_i, \mathcal{C}}{\text{minimize}} \quad \sum_{i=1}^N \sum_{j \in \mathcal{C}} |\mathbf{v}_{ij}|^2 \\
& \text{subject to} \quad \sum_{i=1}^N |\mathbf{v}_{ij}|^2 \leq P_j, \quad j = 1, 2, \dots, L \\
& \sqrt{1 - \frac{1}{2^{\frac{D_i}{B_i \cdot T_{i,max}^r}}}} \sqrt{\sum_{k=1}^N \left| \sum_{j \in \mathcal{C}} \mathbf{h}_{ij}^H \mathbf{v}_{kj} \right|^2 + \sigma^2} \leq \text{Re} \left( \left| \sum_{j \in \mathcal{C}} \mathbf{h}_{ij}^H \mathbf{v}_{ij} \right|^2 \right), \quad i = 1, 2, \dots, N \\
& C_j = \sum_{i=1}^N \rho_{ij} |\mathbf{v}_{ij}|^2 \cdot r_i \leq C_{j,max}, \quad j = 1, 2, \dots, L.
\end{aligned} \tag{27}$$

Note that  $\mathcal{P}4$  without the fronthaul constraint is an SOC problem, which can be solved by the interior-point method [18], while  $\mathcal{P}4$  including the fronthaul constraint can be addressed by the iterative solution, as shown in [15]. Therefore we can give the iterative Algorithm 1 to deal with  $\mathcal{P}3$  as follows.

---

**Algorithm 1** Proposed iterative algorithm for  $\mathcal{P}3$

---

**Initialize:**  $m = 1, \rho_{ij}^{(0)}, r_i^{(0)}, i = 1, 2, \dots, N, j = 1, 2, \dots, L;$

**Repeat:**

- 1: Solve the SOCP optimization  $\mathcal{P}4$  using interior-point method, obtaining the optimal beamforming vector  $\mathbf{v}_{ij}^{(m)}$ ;
- 2: Update  $r_i^{(m+1)} = r_i^{(m)}$  according to (6) and (7);
- 3: Update  $\rho_{ij}^{(m+1)} = \rho_{ij}^{(m)}$  according to (26);
- 4:  $m = m + 1$ ;

**Until** convergence.

---

#### IV. JOINT OPTIMIZATION SOLUTION

In this section, we are interested in solving the energy minimization and resource allocation optimization jointly between the mobile cloud and mobile network. The objective is to minimize the total energy consumption in mobile cloud for executing the task and in C-RAN for transmitting the processing results back to the mobile user. We assume that the task has to be completed in the given total time constraint, including the executing time plus the transmitting

time. Therefore, the joint energy optimization problem can be given as

$$\begin{aligned}
\mathcal{P5} : \quad & \underset{f_i^C, r_i, \mathbf{v}_{ij}, \mathcal{C}}{\text{minimize}} \quad \sum_{i=1}^N E_i^C + \eta_i E_i^{Tr} \\
& \text{subject to} \\
& \sum_{i=1}^N |\mathbf{v}_{ij}|^2 \leq P_j, \quad j = 1, 2, \dots, L, \\
& r_i = B_i \log \left( 1 + \frac{|\sum_{j \in \mathcal{C}} \mathbf{h}_{ij}^H \mathbf{v}_{ij}|^2}{\sum_{k=1, k \neq i}^N |\sum_{j \in \mathcal{C}} \mathbf{h}_{ij}^H \mathbf{v}_{kj}|^2 + \sigma^2} \right), \quad i = 1, 2, \dots, N, \\
& r_i \geq R_{i,min}, \quad i = 1, 2, \dots, N, \\
& f_i^C \leq f_{i,max}^C, \quad i = 1, 2, \dots, N, \\
& T_i^C + T_i^{Tr} \leq T_{i,max}, \quad i = 1, 2, \dots, N, \\
& \sum_{i=1}^N |\mathbf{v}_{ij}|^2 \cdot r_i \leq C_{j,max}, \quad j = 1, 2, \dots, L.
\end{aligned} \tag{28}$$

where  $R_{i,min}$  is the minimum achievable rate for UE  $i$ , and other constraints in  $\mathcal{P5}$  have been introduced in the last sections. The above  $\mathcal{P5}$  is non-convex problem and is difficult to solve. In the next subsections, we will provide the iterative algorithms based on weighted minimum mean square error (WMMSE) solution to deal with it.

### A. Problem Transformation

It is obvious that the equality of the time constraints holds for  $\mathcal{P5}$  [9] in relaxation. Therefore, by using (2) and (8), time constraints can be relaxed as

$$\begin{aligned}
T_{i,max} &= T_i^{Tr} + T_i^C \\
&= \frac{D_i}{r_i} + \frac{F_i}{f_i^C}.
\end{aligned} \tag{29}$$

Then,  $f_i^C$  can be given as

$$f_i^C = \frac{F_i}{T_{i,max} - \frac{D_i}{r_i}}. \tag{30}$$

Given that  $T_{i,max} > 0$ ,  $f_i^C > 0$  and  $f_i^C \leq f_{i,max}^C$ , one can get the minimum achievable rate as

$$r_i \geq R_{i,min} \tag{31}$$

where

$$R_{i,min} = \frac{D_i}{T_{i,max} - \frac{F_i}{f_{i,max}^C}}. \quad (32)$$

By using (30), (31) and (32),  $\mathcal{P}5$  can be simplified as

$$\begin{aligned} \mathcal{P}6 : \quad & \underset{r_i, \mathbf{v}_{ij}, \mathcal{C}}{\text{minimize}} \quad \sum_{i=1}^N \kappa_i^C \left( \frac{F_i}{T_{i,max} - \frac{D_i}{r_i}} \right)^{\nu_i^C - 1} F_i + \eta_i \frac{\sum_{j \in \mathcal{C}} |\mathbf{v}_{ij}|^2 D_i}{r_i} \\ & \text{subject to} \\ & \sum_{i=1}^N |\mathbf{v}_{ij}|^2 \leq P_j, \quad j = 1, 2, \dots, L, \\ & B_i \log \left( 1 + \frac{|\sum_{j \in \mathcal{C}} \mathbf{h}_{ij}^H \mathbf{v}_{ij}|^2}{\sum_{k=1, k \neq i}^N |\sum_{j \in \mathcal{C}} \mathbf{h}_{kj}^H \mathbf{v}_{kj}|^2 + \sigma^2} \right) \geq R_{i,min}, \quad i = 1, 2, \dots, N, \\ & C_j = \sum_{i=1}^N \rho_{ij} |\mathbf{v}_{ij}|^2 \cdot r_i \leq C_{j,max}, \quad j = 1, 2, \dots, L. \end{aligned} \quad (33)$$

Note that  $f_i^C$  does no longer exist in  $\mathcal{P}6$ . We denote  $\mathbf{v}_j = [\mathbf{v}_{1j}, \mathbf{v}_{2j}, \dots, \mathbf{v}_{Nj}]^H$ ,  $\mathbf{h}_j = [\mathbf{h}_{1j}, \mathbf{h}_{2j}, \dots, \mathbf{h}_{Nj}]^H$ ,  $\mathbf{v}_i = [\mathbf{v}_{i1}, \mathbf{v}_{i2}, \dots, \mathbf{v}_{iC}]^H$ ,  $\mathbf{h}_i = [\mathbf{h}_{i1}, \mathbf{h}_{i2}, \dots, \mathbf{h}_{iC}]^H$  for notation simplification. One can rewrite  $\mathcal{P}6$  as

$$\mathcal{P}7 : \quad \underset{r_i, \mathbf{v}_{ij}, \mathcal{C}}{\text{minimize}} \quad \sum_{i=1}^N \alpha_i(r_i) + t_i \quad (34)$$

subject to : Constraints of ( $\mathcal{P}6$ )

where

$$\alpha_i(r_i) = \kappa_i^C \left( \frac{F_i}{T_{i,max} - \frac{D_i}{r_i}} \right)^{\nu_i^C - 1} F_i, \quad (35)$$

$$t_i = \frac{\beta_i(\mathbf{v}_i)}{r_i} \quad (36)$$

and

$$\beta_i(\mathbf{v}_i) = \eta_i \mathbf{v}_i^H \mathbf{v}_i D_i. \quad (37)$$

Inspired by the solutions provided in [15], [19], [20],  $\mathcal{P}7$  can be rewritten as

$$\mathcal{P}8 : \quad \underset{r_i, \mathbf{v}_{ij}, \mathcal{C}}{\text{minimize}} \quad \sum_{i=1}^N \gamma_i(r_i) + \beta_i(\mathbf{v}_i) \quad (38)$$

subject to : Constraints of ( $\mathcal{P}6$ )

where

$$\gamma_i(r_i) = \alpha_i(r_i) - t_i \cdot r_i, \quad (39)$$

which can be solved by using iterative solution by updating  $t_i$  in each iteration, if  $r_i$  is a constant. Next, we will use WMMSE-based method to deal with the situation if  $r_i$  is a variable.

### B. WMMSE-based Solution

The WMMSE method is introduced by [21], [22] and use to address the weighted sum rate problem. One can see that the objective of  $\mathcal{P}8$  is an decreasing function of the mobile user's data rate  $r_i$ . Therefore, we can reformulate  $\mathcal{P}8$  as an equivalent WMMSE problem and use the block coordinate descent approach to solve it.

Assume 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. Thus, the corresponding MSE at UE  $i$  can be given as

$$\begin{aligned} e_i &= E [(\mathbf{u}_i y_i - x_i)(\mathbf{u}_i y_i - x_i)^H] \\ &= \sum_{i=1}^N \mathbf{u}_i^H (\mathbf{h}_i^H \mathbf{v}_i \mathbf{v}_i^H \mathbf{h}_i + \sigma_i^2) \mathbf{u}_i - 2 \operatorname{Re} [\mathbf{u}_i^H \mathbf{h}_i^H \mathbf{v}_i] + 1. \end{aligned} \quad (40)$$

Then, by fixing all the transmit beamforming vector  $\mathbf{v}_i$ , the optimal receive beamforming vector can be give by the well-known MMSE receiver as

$$\mathbf{u}_i = (\mathbf{h}_i^H \mathbf{v}_i) \cdot \left( \sum_{k=1}^N \mathbf{h}_i^H \mathbf{v}_k \mathbf{v}_k^H \mathbf{h}_i + \sigma_i^2 \right)^{-1}. \quad (41)$$

Next, letting

$$\tau_i(e_i) = \gamma_i(-B_i \cdot \log(e_i)), \quad (42)$$

one can see that  $\tau_i(e_i)$  is a strictly convex function under the constraints of  $\mathcal{P}8$  [15], [21]. Then, by fixing the transmit beamforming vector  $\mathbf{v}_i$  and the MMSE receiver  $\mathbf{u}_i$ , the corresponding optimal MSE weight  $\phi_i$  can be given by

$$\phi_i = \frac{\partial \tau(e_i)}{\partial e_i} = \frac{D_i \kappa_i^C (\nu_i^C - 1) \log(2) \left( \frac{B_i F_i \log(e_i)}{B_i T_{i, \max} \log(e_i) + D_i \log(2)} \right)^{\nu_i^C}}{B_i e_i \log^2(e_i)} + \frac{B_i t_i}{e_i \log(2)}. \quad (43)$$

Then, by fixing the optimal MSE weight  $\phi_i$  and MMSE receiver  $\mathbf{u}_i$ , the optimal transmit beamforming vector  $\mathbf{v}_i$  can be calculated by solving the following SOCP problem as

$$\mathcal{P}9 : \underset{r_i, \mathbf{v}_{ij}, \mathcal{C}}{\text{minimize}} \quad \sum_{i=1}^N \phi_i \cdot e_i + \beta_i(\mathbf{v}_i) \quad (44)$$

subject to : Constraints of ( $\mathcal{P}6$ ).

Thus, we can deal with the overall optimization problem with WMMSE-based iterative method as in Algorithm 2, where  $Z^{(n)} = \sum_{i=1}^N \alpha_i(r_i^{(n)}) + t_i^{(n)}$  and  $\varepsilon$  is a small constant to guarantee convergence.

---

**Algorithm 2** Proposed iterative algorithm for joint optimization problem

---

**Initialize:**  $n = 1, t_i^{(0)}, \rho_{ij}^{(0)}, \mathbf{v}_{ij}^{(0)}, r_i^{(0)}, i = 1, 2, \dots, N, j = 1, 2, \dots, L;$

**Repeat:**

- 1: Obtain the receive beamforming vector  $\mathbf{u}_i^{(n)}$  according to (41) by fixing  $\mathbf{v}_{ij}^{(n-1)}$ ;
- 2: Obtain the MSE weight  $\phi_i$  according to (43) 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  $\mathcal{P}9$  by fixing  $\phi_i^{(n)}, \mathbf{u}_i^{(n)}$ ;
- 4: Update  $r_i^{(n+1)} = r_i^{(n)}$  according to (6) and (7);
- 5: Update  $t_i^{(n+1)} = t_i^{(n)}$  according to (36);
- 6: Update  $\rho_{ij}^{(n+1)} = \rho_{ij}^{(n)}$  according to (26);
- 7:  $n = n + 1;$

**Until**  $|Z^{(n+1)} - Z^{(n)}| < \varepsilon$

---

## V. SIMULATION RESULTS

In this section, simulation results are provided to show the effectiveness of the proposed joint energy minimization optimization. The simulation environment is shown as Fig. 2, in which we consider the C-RAN network with  $L = 4$  RRHs, each equipped with  $K = 2$  antennas. Also, we assume there are  $N = 5$  mobile users, each of which has only one antenna. We assume there are five mobile clones co-located with the BBUs, and each mobile clone has the same software stack as its corresponding mobile users and can execute the task for the mobile user.

Moreover, we assume the maximum transmit power for each RRH is 1 W, while the maximum computation capacity for each mobile clone is  $10^6$  CPU cycles per second. Similar with [23], we model the path and penetration loss as

$$p(d) = 127 + 25\log_{10}(d) \quad (45)$$

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 the energy tradeoff factor between the mobile clone and C-RAN as  $\eta_i = 10$  and the parameter for the cloud energy model  $\nu_i^C = 2$ . Also, we

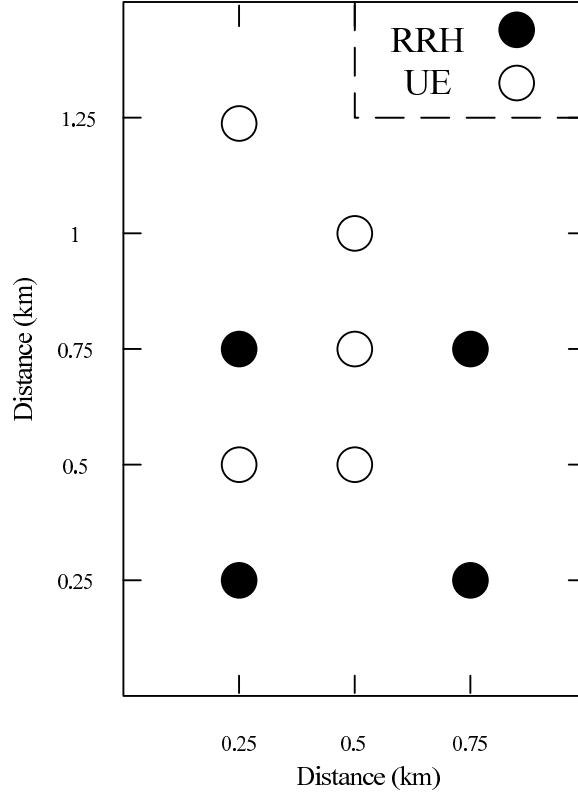


Fig. 2. C-RAN network with  $L = 4$  RRHs and  $N = 5$  UEs.

assume the wireless channel bandwidth as 10 MHz and the fronthaul capacity constraint as 10 Mbps.

In Fig. 3, we show the energy consumption for the whole system including mobile clone and C-RAN for different QoS requirement and different CPU cycles of the task. Transmission data  $D_i = 1000$  bits is set in this figure. One can see that with the increase of the CPU cycles of the task  $F_i$ , the energy cost rise correspondingly. Also, with the increase of the time constraint, the total energy decrease, as the mobile clone and the C-RAN can have more time to complete the task and return the result to the mobile user.

In Fig. 4, we show the total energy consumption for different QoS requirement and different data size of the transmission result.  $F_i = 1500$  CPU cycles is set in this figure. One can see that with the increase of the result data size  $D_i$  of the task, the energy cost increase correspondingly, but not as fast as Fig. 3. This is due to the tradeoff factors we set. Similarly, with the increase of the time constraint, the total energy cost decrease.



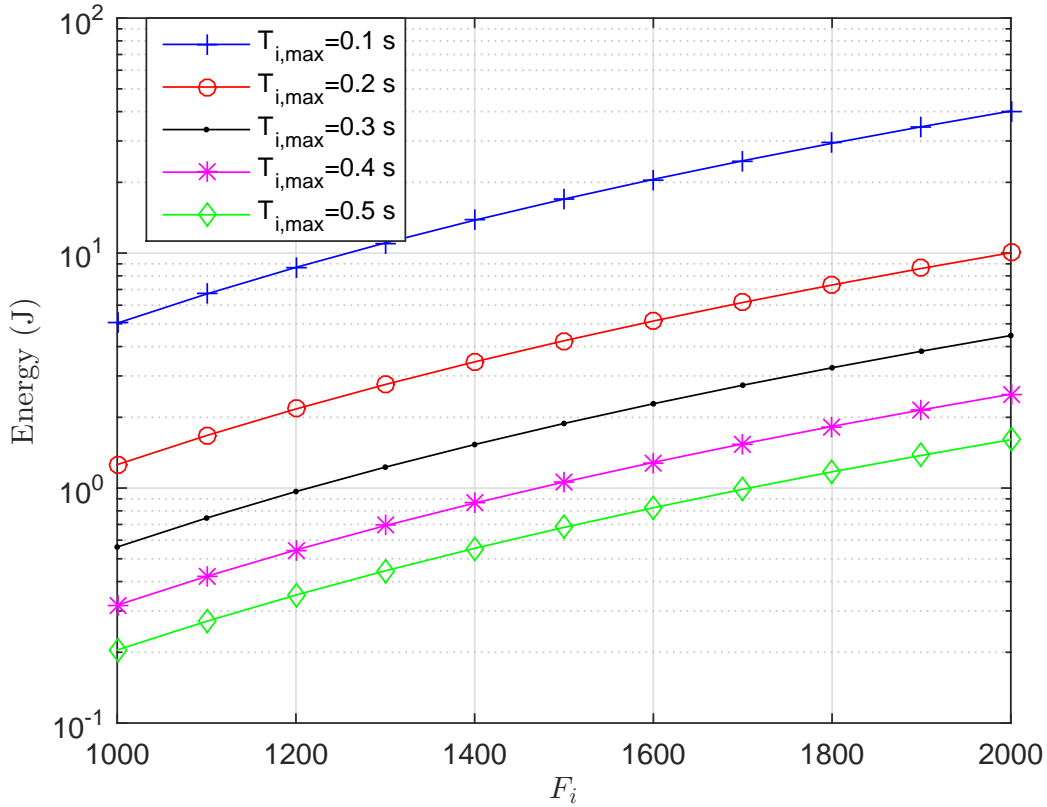


Fig. 3. Total energy consumption vs. CPU cycles under different  $T_{i,max}$  with  $D_i = 1000$ .

In Fig. 5, the relations between the total energy consumption and different time constraints are examined under different  $D_i$  with total CPU cycles  $F_i = 1500$ . One can see that with the increase of the time constraints, the energy consumption decreases, as expected. Also, with the increase of the data size, the energy increases, but the gap between them is small.

Similar with Fig. 5, Fig. 6 shows that the whole energy consumption of mobile cloud and C-RAN decreases either with the increase of the time constraints or with the decrease of the CPU cycles required by each task.

In Fig. 7 and Fig. 8, we compare the proposed joint energy minimization optimization with the separate energy minimization solutions, which has been used in some works such as [23], etc. 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  $T_{i,max}^{Tr} + T_{i,max}^C = T_{i,max}$ .  $T_{i,max} = 0.1$ s is set in both Fig. 7 and Fig. 8 while  $D_i = 1000$  and  $F_i = 1500$  are set in Fig. 7 and Fig. 8, respectively. One can see that the joint

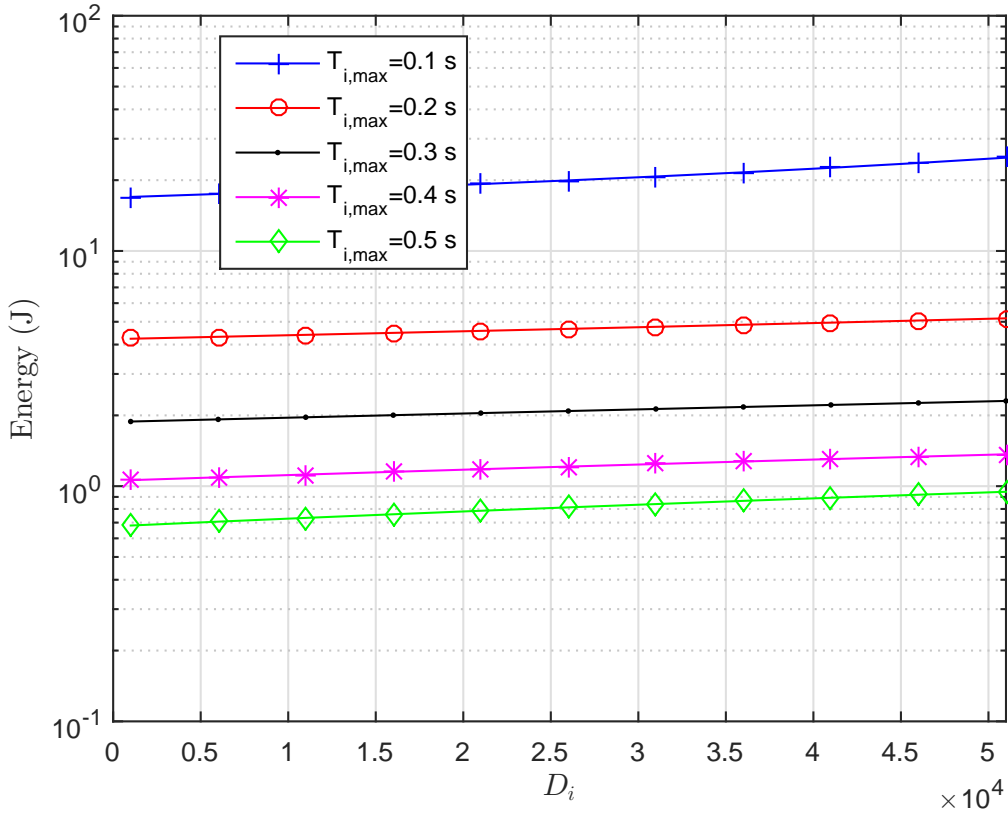


Fig. 4. Total energy consumption vs. data size under different  $T_{i,max}$  with  $F_i = 1500$ .

energy minimization achieves the best performance, followed by the second best solution when setting  $T_{i,max}^{Tr} = T_{i,max}^{Tr}/4$  in both Fig. 7 and Fig. 8. The performance of  $T_{i,max}^{Tr} = T_{i,max}^{Tr} * 3/4$  can be shown as the worst solution among the test ones in both figures. Therefore, the simulation results show that the proposed joint energy minimization outperforms the separate solutions in all the cases.

## VI. CONCLUSION

A novel C-RAN architecture with the mobile clone involved is proposed in this paper by taking full advantages of the two cloud-based techniques. In particular, we assume there is one task needed to be executed in the mobile clone for each UE and we model this task with two features, i.e, the total number of the CPU cycles required to complete this task and the data size required to transmit the result back to the UEs through C-RAN. We jointly minimize the whole

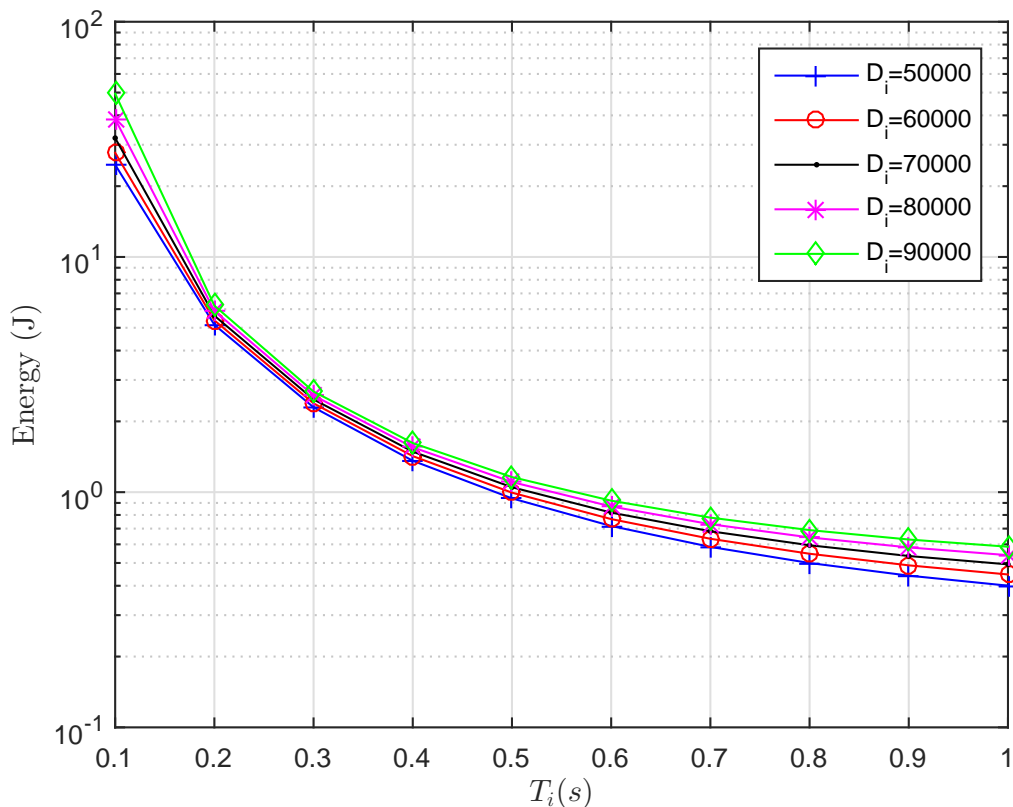


Fig. 5. Total energy consumption vs. time constraint under different data size  $D_i$  with  $F_i = 1500$ .

energy cost in mobile cloud and mobile network by modeling this problem into the optimization problem when taking the time constraints into consideration. Also, we consider the fronthaul constraints in C-RAN in order to get the sparse solutions. Numerical results are presented to show that the proposed energy minimization and resource allocation solution can improve the system performance and save energy.

## VII. ACKNOWLEDGEMENT

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

## REFERENCES

- [1] J. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. Soong, and J. Zhang, “What will 5g be?” *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, June 2014.

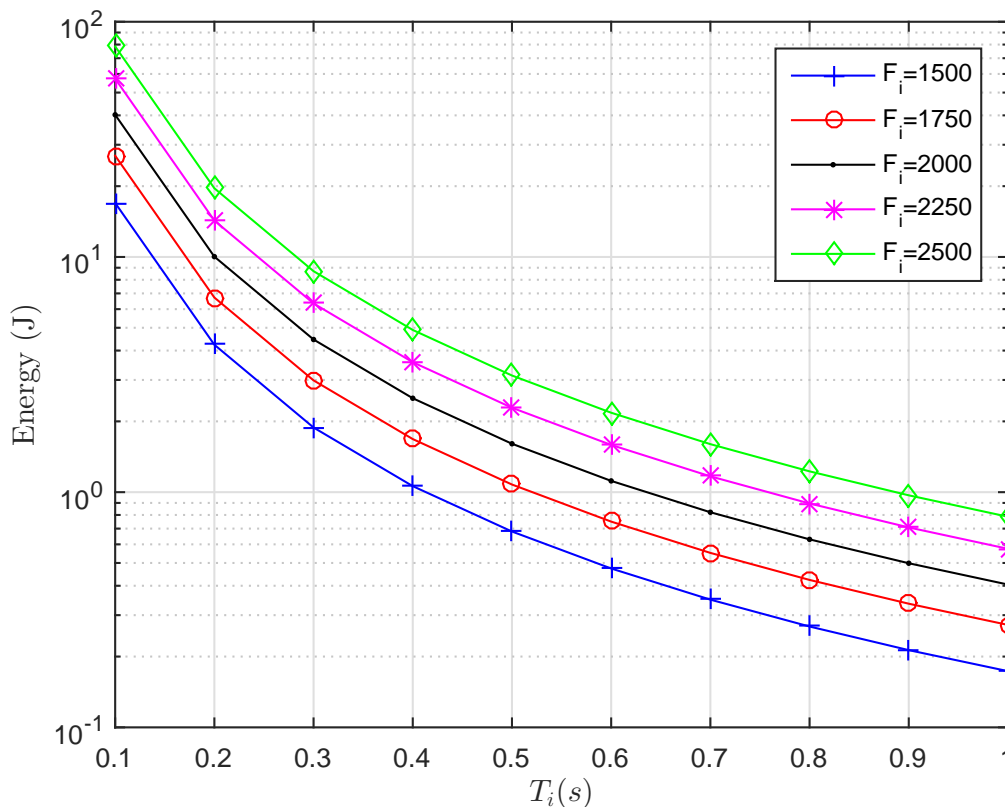


Fig. 6. Total energy consumption vs. time constraint under different CPU cycles  $F_i$  with  $D_i = 1000$ .

- [2] 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.
- [3] C. M. R. Institute., "C-ran white paper: The road towards green ran. [online]," (Jun. 2014), Available: <http://labs.chinamobile.com/cran>.
- [4] 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.
- [5] 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.
- [6] W. Zhang, Y. Wen, K. Guan, D. Kilper, H. Luo, and D. Wu, "Energy-optimal mobile cloud computing under stochastic wireless channel," *IEEE Transactions on Wireless Communications*, vol. 12, no. 9, pp. 4569–4581, September 2013.
- [7] 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.
- [8] 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.
- [9] J. Tang, W. P. Tay, and T. Quek, "Cross-layer resource allocation in cloud radio access network," in *2014 IEEE Global*

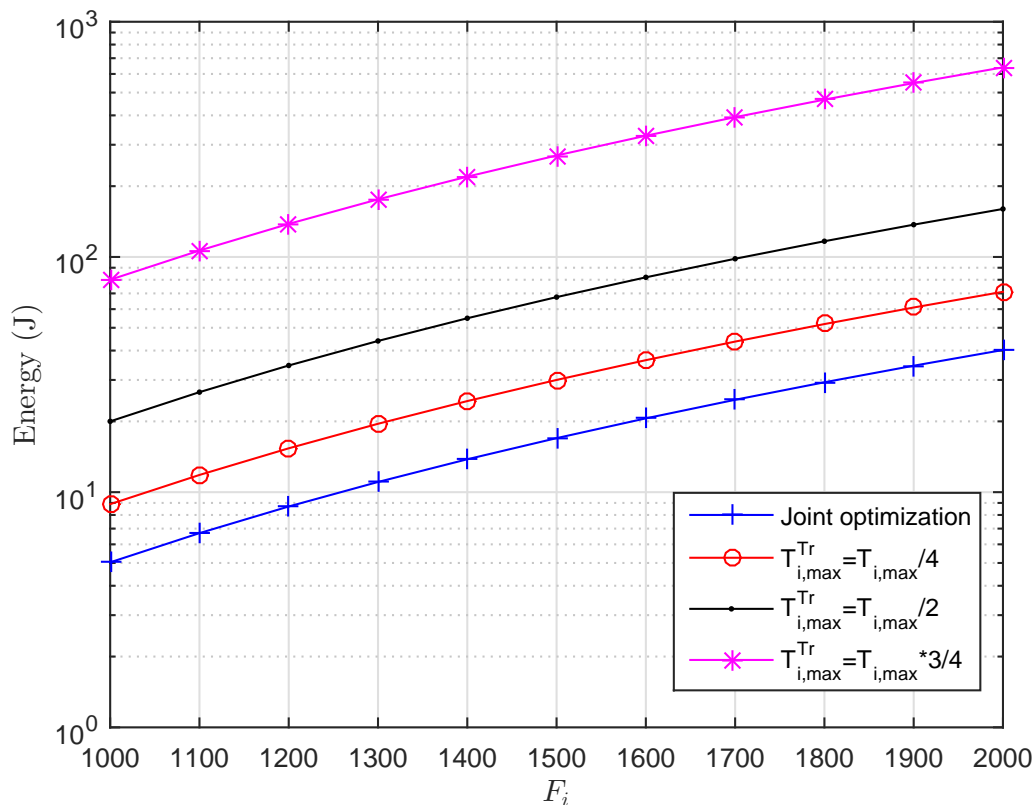


Fig. 7. Total energy consumption vs. CPU cycles under different  $T_{i,max}^{Tr}$  with  $D_i = 1000$ .

*Conference on Signal and Information Processing (GlobalSIP)*, Dec 2014, pp. 158–162.

- [10] M. Guazzone, C. Anglano, and M. Canonico, “Energy-efficient resource management for cloud computing infrastructures,” in *2011 IEEE Third International Conference on Cloud Computing Technology and Science (CloudCom)*, Nov 2011, pp. 424–431.
- [11] 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.
- [12] J. Tang, W. P. Tay, and Y. Wen, “Dynamic request redirection and elastic service scaling in cloud-centric media networks,” *IEEE Transactions on Multimedia*, vol. 16, no. 5, pp. 1434–1445, Aug 2014.
- [13] A. P. Miettinen and J. K. Nurminen, “Energy efficiency of mobile clients in cloud computing,” in *Proceedings of the 2nd USENIX conference on Hot topics in cloud computing*, 2010, p. 4.
- [14] 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.
- [15] 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.

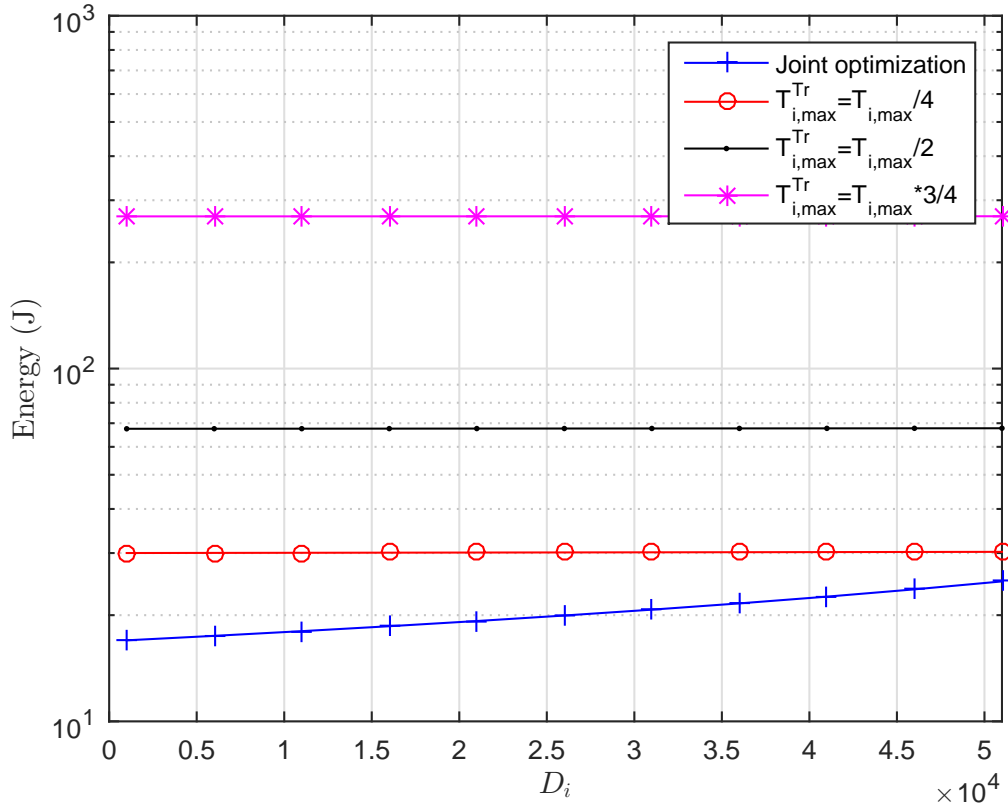


Fig. 8. Total energy consumption vs. data size under different  $T_{i,max}^{Tr}$  with  $F_i = 1500$ .

- [16] 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.
- [17] 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.
- [18] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [19] W. Dinkelbach, "On nonlinear fractional programming," *Manag. Science*, vol. 13, pp. 492–498, 1967.
- [20] M. Peng, K. Zhang, J. Jiang, J. Wang, and W. Wang, "Energy-efficient resource assignment and power allocation in heterogeneous cloud radio access networks," *IEEE Transactions on Vehicular Technology*, vol. PP, no. 99, pp. 1–1, 2014.
- [21] 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.
- [22] 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.
- [23] 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.