Multi-Objective Optimization for RRH Clustering in Cloud Radio Access Networks

Hussein Taleb*, Melhem El Helou*, Samer Lahoud*, Kinda Khawam† and Steven Martin‡
*Ecole Supérieure d'Ingénieurs de Beyrouth, Saint Joseph University of Beirut, Beirut, Lebanon
† University of Versailles, Versailles, France
‡ Laboratoire de Recherche en Informatique, University of Paris-Sud, Orsay, France

***Abstract*—Cloud Radio Access Network (C-RAN) has been proposed as a potential solution to reduce network power consumption, while providing acceptable quality of service (QoS). In this context, the conventional base station is separated into a Base Band Unit (BBU) and a Remote Radio Head (RRH). The BBUs are located in a cloud data center, whereas the RRHs are geographically distributed across multiple sites. To achieve statistical multiplexing gain, many RRHs may be clustered and associated with a single BBU. We formulate the RRH clustering as a multi-objective optimization problem, using the weighted-sum method and the ϵ-constraint method. Our objectives are to minimize network power consumption and transmission delay. As these formulations result in non-linear problems, exhaustive search is used to obtain optimal solutions. Simulation results compare our solutions against the no-clustering solution, where each RRH is associated with a separate BBU, and the grand coalition solution, where all RRHs are associated with a single BBU. We further investigate the trade-off between our two crucial but conflicting objectives.**

I. Introduction

In recent years, data traffic over cellular networks has been growing exponentially. It is estimated that monthly global mobile data traffic will be 49 exabytes by 2021 [1]. In this context, Cloud Radio Access Network (C-RAN) was introduced as a promising technology to meet quality of service (QoS) requirements, while reducing network power consumption. In this centralized architecture, a base station is separated into a Base Band Unit (BBU) and a Remote Radio Head (RRH). The BBUs are located in a cloud data center and connected with distributed RRHs via fronthaul links. This decoupling allows many RRHs to be mapped to a single BBU, sharing radio and computing resources. Such clustering reduces network capital and operational expenditures and improves user radio conditions.

A real challenge is to design a BBU to RRH mapping method, also known as RRH clustering method, that minimizes the network transmission delay and power consumption. To achieve this goal, RRHs are organized into disjoint clusters, in a way to minimize the number of active BBUs without degrading user QoS. As these two objectives are conflicting, we resort to multi-objective optimization methods, namely the weighted-sum method and the ϵ-constraint method, to formulate our problem.

II. Related Work

Power consumption and delay are two contradictory problems that need to be jointly addressed to improve network performance. In various literature, the two problems have been treated independently. Articles [2], [3] have focused on the power saving problem. In [2], the authors proposed an algorithm that reduces the number of active BBUs, where the objective is to minimize the network power consumption. To solve this, the underutilized BBU is switched off after offloading its traffic to another suitable BBU. The authors in [3] proposed a novel approach to save energy, by adaptively adjusting RRH transmission power according to current traffic conditions. In this work, they formulated a non-linear programming model to find the best possible topology which minimizes the network energy consumption. However, satisfying user QoS has not been considered in these papers. Furthermore, the works in [4], [5] have tackled the delay minimization problem. To minimize the delay that occurs in data transmission and reception, the authors in [4] proposed a new approach for BBU selection based on queuing theory in order to minimize the response time in the BBU pool. A cost-constrained delay minimization framework for C-RAN is formulated in [5] that aims to minimize the overall network delay. Yet, this works overlooks energy minimization. Compared to prior work in the state of the art taking into account the power saving and the user quality of service, the authors in [6] propose an optimization problem that jointly minimizes the network power consumption and transmission delay. Power saving is achieved by adjusting the RRH transmission power from high transmit power levels to low transmit levels or switched-off. Minimizing the transmission delay is achieved by selecting the best user association with the RRH. However, this work does not support the feature of tuning the weights associated with the two costs (*i.e.,* the power and the delay).

In this paper, the main contributions can be summarized as follows:

- We formulate the problem of power-delay minimization in cloud radio access networks. The transmission data reflects user QoS and is defined as the sum of data unit transmission.
- Two different methods, the weighted-sum method and the ϵ-constraint method, are proposed to formulate the multi-objective problem. Thereafter, exhaustive search is used to obtain optimal solutions.

The rest of this article is organized as follows. Section III described the system model. In Section IV, we formulated the weighted-sum problem. The ϵ-constraint problem is presented in Section V. Simulation results are discussed in Section VI, and concluding remarks are given in Section VII.

III. System Model

Consider R RRHs, denoted by the set $\mathcal{R} = \{r|1 \leq r \leq R\}$, that are randomly scattered in the region $\mathcal{S}$. We denote by $\mathcal{B} = \{b|1 \leq b \leq B\}$ the set of available BBUs located in a centralized pool. We define Δ^r and Ψ^b as two binary variables, which are equal to one if RRH r and BBU b are in active mode respectively, and zero otherwise. We further denote by $\mathcal{U} = \{u|1 \leq u \leq U\}$ the set of users in region $\mathcal{S}$. We also define two sets of decision variables:

- $\Phi_{u,r}$ that represents user association and is defined as follows:

$$\Phi_{u,r} = \begin{cases} 1, & \text{if user } u \text{ is attached to RRH } r, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

In this work, we assume that user u is associated with its best received RRH.

- $\Gamma_{r,b}$ that represents BBU-RRH mapping and is defined as follows:

$$\Gamma_{r,b} = \begin{cases} 1, & \text{if RRH } r \text{ is attached to BBU } b, \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

The signal-to-interference-plus-noise ratio achieved by user u, attached to RRH r, that is mapped to BBU b can be expressed as:

$$\Upsilon_{u,r,b} = \frac{\pi_r G_{u,r}}{N_0 + \sum_{r' \neq r}(1 - \Gamma_{r',b})\pi_{r'} G_{u,r'}}, \tag{3}$$

where π_r is the transmit power of RRH r. $G_{u,r}$ is the channel gain of user u when associated with RRH r and N_0 denotes the thermal noise power. Particularly, $\sum_{r' \neq r}(1-\Gamma_{r',b})\pi_{r'} G_{u,r'}$ represents inter-cluster interferences, caused by the RRHs that are not associated with BBU b.

We denote by $\widehat{R_{u,r,b}}$ the instantaneous peak throughput achieved by user u associated to RRH r, that is mapped to BBU b. It can be computed using Shannon's formula as follows:

$$\widehat{R_{u,r,b}} = W \log_2(1 + \Upsilon_{u,r,b}), \tag{4}$$

where W is the total system bandwidth.

A. Delay Model

We denote by $R_{u,r,b}$ the average throughput perceived by user u from RRH r, that is mapped to BBU b. Assuming a fair resource sharing, $R_{u,r,b}$ can be written as follows:

$$R_{u,r,b} = \frac{\widehat{R_{u,r,b}}}{\sum_r \sum_u \Phi_{u,r}\Gamma_{r,b}}, \tag{5}$$

where $\sum_r \sum_u \Phi_{u,r}\Gamma_{r,b}$ represents the number of users belonging to BBU b. Note that $R_{u,r,b}$ depends on user geographical position and radio conditions.

Thereafter, we define by $T_{u,r,b}$ the amount of time necessary to send a data unit to user u from RRH r, that is mapped to BBU b. In fact, the delay needed to transmit a bit for a given user is the inverse of the average throughput perceived by this user. Thus,

$$T_{u,r,b} = \frac{1}{R_{u,r,b}} = \frac{\sum_r \sum_u \Phi_{u,r}\Gamma_{r,b}}{\widehat{R_{u,r,b}}}. \tag{6}$$

B. Power Consumption Model

Using the model proposed in [7], the power consumed in a C-RAN is the sum of two terms: the power consumed by all the BBUs at the baseband processing pool and the power consumed by all the RRHs. Thus, the total C-RAN power consumption can be expressed as:

$$P_{total} = \sum_{b \in \mathcal{B}} P_b + \sum_{r \in \mathcal{R}} P_r, \tag{7}$$

where P_b and P_r respectively denote the power consumed by BBU b and that consumed by RRH r .

Moreover, P_b can be expressed as:

$$P_b = \begin{cases} \lambda, & \text{if } \Psi^b = 1, \\ 0, & \text{otherwise,} \end{cases} \tag{8}$$

where λ represents the power consumption of BBU b when in active mode. Besides, P_r can be written as:

$$P_r = \begin{cases} \pi_r^0 + \delta\pi_r, & \text{if } \Delta^r = 1 \ , \\ \pi_r^s, & \text{otherwise,} \end{cases} \tag{9}$$

where δ is the power amplifier efficiency. π_r^0 and π_r^s are the power consumption in active and sleep mode respectively, and π_r is the transmit power of RRH r.

C. Network Cost

We denote by C the network cost function defined as the sum of the network power consumption and the total transmission delay. More precisely, the network power consumption P_{total} depends on the number of active BBUs and RRHs in the network, and is expressed as:

$$\begin{aligned} P_{total} &= \sum_{b \in \mathcal{B}} \Psi^b \lambda + \sum_{r \in \mathcal{R}} \Delta^r(\delta\pi_r + \pi^0) + \sum_{r \in \mathcal{R}}(1 - \Delta^r)\pi^s \\ &= \lambda \sum_{b \in \mathcal{B}} \Psi^b + \sum_{r \in \mathcal{R}}(\delta\pi_r + \pi^0 - \pi^s)\Delta^r + \sum_{r \in \mathcal{R}} \pi^s. \end{aligned} \tag{10}$$

Furthermore, the total transmission delay, denoted by T_{total}, is defined as the sum of data unit transmission delays of each user u as presented in Equation (5). Thus, T_{total} can be written as:

$$\begin{aligned} T_{total} &= \sum_{b \in \mathcal{B}} \sum_{r \in \mathcal{R}} \sum_{u \in \mathcal{U}} \Phi_{u,r}\Gamma_{r,b} T_{u,i,k} \\ &= \sum_{b \in \mathcal{B}} \sum_{r \in \mathcal{R}} \sum_{u \in \mathcal{U}} \Phi_{u,r}\Gamma_{r,b} \frac{\sum_r \sum_u \Phi_{u,r}\Gamma_{r,b}}{\widehat{R_{u,r,b}}} \\ &= \sum_{b \in \mathcal{B}} \sum_{r \in \mathcal{R}} \sum_{u \in \mathcal{U}} \Phi_{u,r}\Gamma_{r,b} \frac{\sum_i \sum_u \Phi_{u,r}\Gamma_{r,b}}{W \log_2(\frac{\pi_r G_{u,r}}{N_0 + \sum_{r' \neq r}(1-\Gamma_{r',b})\pi_{r'} G_{u,r'}})}. \end{aligned} \tag{11}$$

Consequently, C is given by:

$$\begin{aligned} C &= \alpha' P_{total} + \beta' T_{total} \\ &= \alpha' \lambda \sum_{b\in\mathcal{B}} \Psi^b + \sum_{r\in\mathcal{R}} (\delta\pi_r + \pi^0 - \pi^s)\Delta^r + \sum_{r\in\mathcal{R}} \pi^s \\ &+ \beta' \sum_{b\in\mathcal{B}} \sum_{r\in\mathcal{R}} \sum_{u\in\mathcal{U}} \Phi_{u,r}\Gamma_{r,b} \frac{\sum_i \sum_u \Phi_{u,r}\Gamma_{r,b}}{W \log_2(\frac{\pi_r G_{u,r}}{N_0 + \sum_{r'\neq r}(1-\Gamma_{r',b})\pi_{r'} G_{u,r'}})}, \end{aligned} \tag{12}$$

where α', β' are two normalizing constants.

IV. Weighted-Sum Problem Formulation

The weighted-sum method is frequently used to formulate and solve multi-objective optimization problems. The idea is to convert multi-objective problems into single-objective problems by aggregating the objective functions. It consists in associating a weighting factor to each component of the multiple-objective optimization problem. The weighting factors allow to investigate the trade-offs between the network performance indicators, making the solution flexible. By applying the weighted-sum method, our network cost C can be rewritten as:

$$C = \alpha\alpha' P_{total} + \beta\beta' T_{total} \tag{13}$$

where α and β are the weighting factors associated with C. Note that α and $\beta \in [0, 1]$ and $\alpha + \beta$=1. Particularly, when α equals 1 and β equals 0, we ignore the effect of delay. Moreover, when α decreases and β increases, more emphasis is put on the delay component.

Consequently, our optimization problem $(\mathcal{P})$ consists in finding an optimal RRH clustering that minimizes the network cost C. Therefore, $(\mathcal{P})$ can be written as follows:

$$\underset{\Gamma}{\text{minimize}} \quad C(\Gamma) \tag{14}$$

$$\text{subject to} \quad \sum_{b\in\mathcal{B}} \Gamma_{r,b} \leq 1, \ \forall r \in \mathcal{R} \tag{15}$$

$$\Gamma_{r,b} \leq \Psi^b, \ \forall b \in \mathcal{B} \tag{16}$$

$$\Gamma_{r,b}, \Psi^b \in \{0, 1\}, r \in \mathcal{R}, b \in \mathcal{B} \tag{17}$$

Constraints (15) ensure that each RRH is attached to one BBU. Constraints (16) ensure that BBU b is turned on if at least one RRH is mapped to it, and finally constraints (17) indicate that all the decision variables, namely $\Gamma_{r,b}$ and Ψ^b, are binary.

V. ϵ-Constraint Problem Formulation

The ϵ-constraint method is based on the optimization of one selected objective component, while considering the other objectives as constraints bounded by the epsilon level ϵ. Consequently, our network cost C can be transformed from a multi-objective problem to a single-objective problem. The objective is to minimize the total transmission delay under the total power constraint. Thus, the optimization problem $(\mathcal{P})$ can be reformulated as:

$$\underset{\Gamma}{\text{minimize}} \quad T_{total}(\Gamma) \tag{18}$$

$$\text{subject to} \quad (15) - (16) - (17) \tag{19}$$

$$P_{total}(\Gamma) \leq \epsilon P_{max} \tag{20}$$

where P_{max} is the maximum power consumption when all BBUs and RRHs are activated.

VI. Simulation Results

In this section, we use MATLAB for simulation. For illustration, we consider a small network consisting of 7 cells, where the center cell is surrounded by 6 interfering cells as shown in Fig. 1. We take 250 simulation snapshots. We assume that users are uniformly distributed in the network, and each user is associated with the RRH whose radio signal is the best received. Performance metrics are averaged and shown with 95% confidence interval. In order to show the effectiveness of our approach, we compare the numerical results with the two state-of-the-art methods, namely no-clustering and the grand coalition.Table I illustrates the main parameters considered in our simulation.

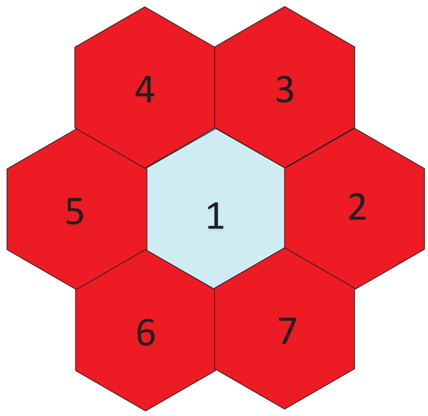

Fig. 1. Network topology

TABLE I
Simulation Parameters

Parameter	***Value***
π_r, $\forall i$	10 W
π^0	6.8 W
π^s	4.3 W
δ	4
λ	40 W
Cell radius	500 m
W	20 MHz
N_0	−174 dBm/Hz
P_{max}	607.6 W

In this section, we solve the ϵ-constraint problem. By adjusting the value of ϵ associated with the total power constraint, different operator policies can be applied. We consider three different strategies:

- Strategy 1: ϵ = 0.25.
- Strategy 2: ϵ = 0.5.
- Strategy 3: ϵ = 0.75.

Fig. 2 illustrates the number of active BBUs as a function of the number of users in the network. Since a single BBU is dedicated to each RRH, the number of active BBUs given by the no-clustering method increases with the number of active RRHs. In fact, as the number of users increases, the number of active RRHs increases, leading to more active BBUs. In addition, the grand coalition activates only one BBU as all RRHs are clustered together. Moreover, as ϵ value increases, the power constraint is more relaxed. Consequently, more BBUs are activated so as to decrease the transmission delay (*i.e.,* improve user QoS).

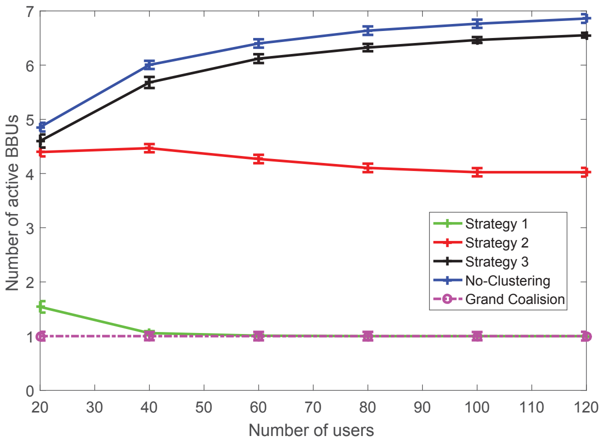

Fig. 2. Number of active BBUs as a function of the number of users

Fig. 3 shows the average interference level experienced by a user as a function of the number of users in the network. In the grand coalition scheme, users do not experience any interference since all RRHs are mapped to one BBU. For the no-clustering scheme, the inter-cluster interference increases with the number of users. In fact, the higher the number of users is, the higher the number of activated BBUs is so as to cope with the increased traffic load. However, activating more BBUs introduces more inter-cluster interference. Furthermore, for $\epsilon = 0.25$, which corresponds to strategy 1, the inter-cluster interference can be eliminated (*i.e.,* above 40 users) as only one BBU is switched on (cf. Fig. 2). In addition, by increasing ϵ value, the interference level in the network increases as more BBUs are activated (cf. Fig. 2).

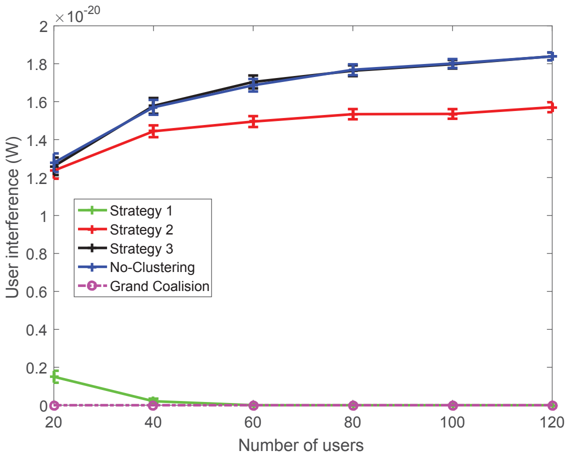

Fig. 3. Interference level per user as a function of the number of users

Figures 4 and 5 show the user transmission delay and the C-RAN power consumption as a function of the number of users. The grand coalition and the optimal solution applying strategy 1 benefit from very limited radio resources and consequently provide the highest user delay, albeit the lowest power consumption. Contrarily, the no-clustering solution and the optimal solution applying strategy 3 increase the number of radio resources available to users, leading to the lowest user delay at the cost of the highest inter-cluster interference (cf. Fig. 3) and power consumption. However, the optimal solution applying strategy 2 provides very close user delay to that of the no-clustering method, especially for low number of users, yet with lower power consumption. Further, as the number of users in the network increases, the gap between the two schemes increases. Note that strategy 3 achieves around 15% power saving in comparison with the no-clustering solution, while achieving the same user delay (cf. Fig. 4).

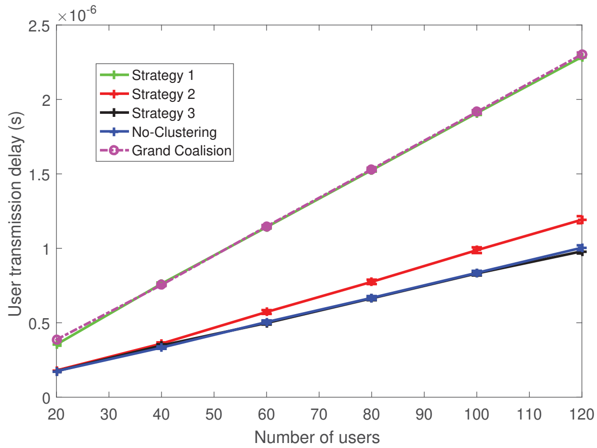

Fig. 4. Transmission delay per user as a function of the number of users

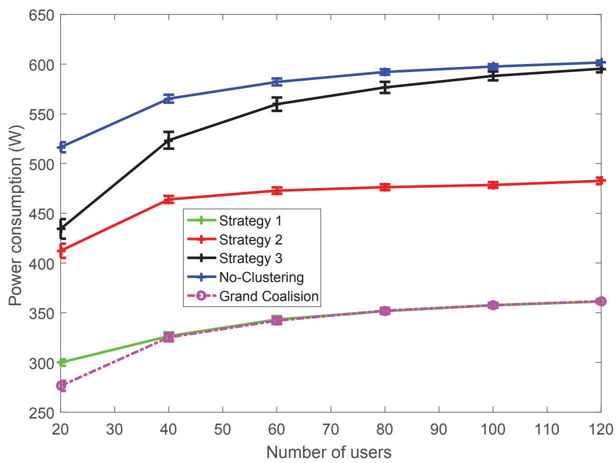

Fig. 5. Network power consumption as a function of the number of users

VII. Conclusion

In this paper, we have addressed the RRH clustering in cloud radio access networks. Using the weighted-sum method and the ϵ-constraint method, RRH clustering was formulated as a multi-objective optimization problem. Our objectives are to minimize network power consumption and transmission delay. Exhaustive search is used to obtain optimal solutions for the ϵ-constraint problem. Numerical results highlighted the optimal delay-power tradeoff that can be achieved under different power constraint strategies. In our future work, we will introduce heuristic methods to solve the problem in large network scenarios.

References

[1] Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021 Cisco, March 2017.

[2] M. Khan, R. S. Alhumaima, and H. S. Al-Raweshidy, "Reducing energy consumption by dynamic resource allocation in c-ran," in *2015 European Conference on Networks and Communications (EuCNC)*, June 2015, pp. 169–174.

[3] M. A. Yigitel, O. D. Incel, and C. Ersoy, "Dynamic base station planning with power adaptation for green wireless cellular networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2014, no. 1, p. 77, May 2014. [Online]. Available: https://doi.org/10.1186/1687-1499-2014-77

[4] A. Arfaoui, S. Hamouda, L. Nuaymi, and P. Godlewski, "Minimization of delays in multi-service cloud-ran bbu pools," in *2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC)*, June 2017, pp. 1846–1850.

[5] J. Wu and D. Ghosal, "Cost-constrained delay minimization of c-ran and its multiplexing gain," in *2016 IEEE International Conference on Communications (ICC)*, May 2016, pp. 1–6.

[6] F. Moety, S. Lahoud, B. Cousin, and K. Khawam, "Joint power-delay minimization in 4g wireless networks," in *2014 IFIP Wireless Days (WD)*, Nov 2014, pp. 1–8.

[7] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. A. Imran, D. Sabella, M. J. Gonzalez, O. Blume, and A. Fehske, "How much energy is needed to run a wireless network?" *IEEE Wireless Communications*, vol. 18, no. 5, pp. 40–49, October 2011.