Computer Networks 182 (2020) 107445

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet

# A fully distributed approach for joint user association and RRH clustering in cloud radio access networks

Hussein Taleb [a,*], Kinda Khawam [b], Samer Lahoud [a], Melhem El Helou [a], Steven Martin [c]

[a] Ecole Supérieure d'Ingénieurs de Beyrouth, Saint Joseph University of Beirut, Beirut, Lebanon
[b] University of Versailles, Versailles, France
[c] Laboratoire de Recherche en Informatique, University of Paris-Sud, Orsay, France

ARTICLE INFO

*Keywords:*
Cloud radio access networks (C-RAN)
Call Detail Records (CDR)
User association (UA)

ABSTRACT

Cloud radio access network (C-RAN) is a new centralized architecture to meet the exponential growing of demand of mobile traffic in 5G cellular wireless networks. However, C-RAN requires an efficient mechanism for the joint user association and the Remote Radio Head (RRH) clustering to improve network performance. In this paper, we investigate the problem of joint user association (UA) and RRH clustering (RC) in C-RAN. Our objective is to maximize the network utility function incurred by both network power consumption and total user throughput for both streaming and elastic traffic. The formulation of the joint optimization problem is a mixed-integer non-linear programming problem (MINLP), which is NP-hard and usually has no feasible solution. To solve it, we propose to decouple the joint problem into two sub-optimization problems: the user association (UA) sub-problem and the RRH clustering (RC) sub-problem. These two sub-problems are sequentially and iteratively solved until convergence is reached. Leveraging on the information delivered by the Call Detail Records (CDR), simulation results reveal the effectiveness of our heuristic solution for the RC sub-problem in enhancing network utility and adapting to the traffic load variation for both elastic and streaming traffic. It outperforms the performance of the state-of-the-art algorithms for RRH clustering solutions, including no-clustering and grand coalition methods. Moreover, the results show that our approach for the UA sub-problem provides close performance to the optimal UA sub-problem.

## 1. Introduction

The exponential growth of mobile data demand is one of the major challenges of the fifth generation (5G) cellular network. To cope with this demand, network densification is considered as a solution for increasing system capacity. However, a large number of small cells located close to each other increase significantly both interference and power consumption in the network, leading to low energy efficiency and throughput. To tackle the aforementioned challenges, a novel mobile network architecture called Cloud-RAN is considered as a promising solution for 5G [1,2]. As illustrated in Fig. 1, this architecture consists of two main components: the Base Band Unit (BBU) and the Remote Radio Read (RRH). The BBUs and the RRHs are physically separated. The BBUs are placed in a central location (*i.e.,* data center). As we consider, in this work, a MAC-PHY split architecture [3], the BBUs are responsible of full resource allocation management and most baseband signal processing. While the RRHs, geographically scattered across multiple sites, are responsible for user functions (signal processing). The MAC-PHY split architecture requires lower capacity fronthaul links, in comparison with the PHY split architecture and the fully centralized architecture. The BBUs and the RRHs are interconnected via optical fronthaul link. During day time, the users move between different locations, leading to non-uniform network traffic load. Typically, the base stations are dimensioned to support the peak traffic load, which introduced a considerable wastage in power consumption and resource utilization during off-peak hours. However, the physical separation between BBUs and RRHs in C-RAN context enables RRH clustering (RC), where several RRHs can be mapped to a single BBU, achieving statistical multiplexing gain. More precisely, a flexible BBU-RRH reconfiguration can be done according to the network traffic load to support efficiently the non-uniform traffic. Consequently, the number of active BBUs may be reduced, which decreases network power consumption. In addition, the user association mechanism (UA), that consists in selecting the serving RRHs for each user, plays a pivotal role in enhancing both spectrum and energy efficiencies through adequate network load balancing. Common baseline algorithm for user association is the SINR-based algorithm, where the user connects to the

* Corresponding author.
*E-mail address:* hussein.taleb@net.usj.edu.lb (H. Taleb).

https://doi.org/10.1016/j.comnet.2020.107445
Received 6 April 2020; Received in revised form 5 July 2020; Accepted 26 July 2020
Available online 14 August 2020

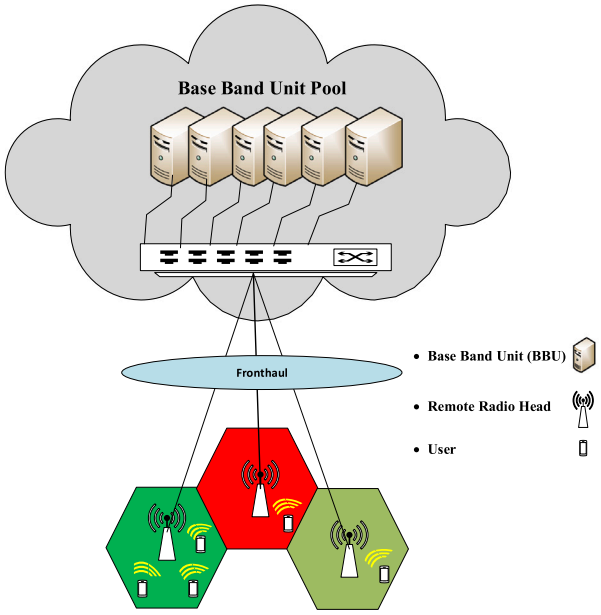


**Fig. 1.** The C-RAN architecture.

RRH with the best received radio signals. This strategy maximizes the achieved spectral efficiency but leads to imbalanced network load.

In the literature, user association and RRH clustering have been largely addressed independently. However, in practice, these two problems are mutually dependent. On the one hand, user association decisions depend on radio resource availabilities and user radio conditions. Both depend on the RRH clusters that have been formed. For instance, when RRHs to which many users are connected have been associated with the same BBU, new arrivals avoid joining these RRHs. If they do, their connected users receive low average rates as the BBU resources are shared among a large number of users. However, RRHs that have been mapped to the same BBU suffer from relatively low interferences. In fact, the intra-cluster interference is eliminated as only one user per BBU is served at a time. Users, that are connected to these RRHs, will have favorable radio conditions. On the other hand, RRH clustering depends on RRH load conditions and consequently on user association decisions. Clustering decisions are ideally load-aware, so as to minimize the number of active BBUs while providing acceptable quality of service.

As these two mechanisms (UA and RC) are mutually dependent, we tackle jointly the user association and RRH clustering problems. Our objective is to maximize the network throughput, while minimizing the network power consumption. As this problem is a mixed integer non-linear programming problem, it can be solved through exhaustive search. However, the computational complexity becomes intractable as the network size increases. Therefore, we decouple our joint problem into two sub-problems: the user association (UA) sub-problem and the RRH clustering (RC) sub-problem. For a given RRH clustering configuration, we solve the UA sub-problem by resorting to non-cooperative game theory. Cooperative game theory is used to address the RC sub-problem, where a heuristic solution based on the merge-and-split rules is applied. These two sub-problems are iteratively solved until convergence. For a realistic simulation scenario, we utilize the Call Detail Records (CDR) presented in [4]. It is a data structure that contains spatial and temporal data information about users traffic (*i.e.,* user ID, service start time, service duration and user-RRH association). Using CDR data, our approach determines the best UA and RC solutions to maximize our network utility for both elastic and streaming traffic load.

## 2. Related work

User association and RRH clustering mechanisms are key mechanisms in 5G C-RAN to enhance network performance. In the literature, several research work have addressed the two problems independently. The work in [5,6] and [7] have focused on the user association problem. In order to reduce the RRHs power consumption, the authors in [5] solved the UA problem based on the nearest RRH association scheme, where each user is served by its nearby RRH. To improve user throughput, the conventional SINR-based scheme is applied in [6] to determine user-RRH associations, where each user is associated with the RRH emitting the best radio signal. In [7], the authors propose an energy efficient user association scheme. Three algorithms including nearest-RRH based user association, single-candidate RRH user association, and multi-candidate RRHs user association have been developed. The aim is to maximize power saving by allowing underutilized RRH to be switched off. In [8], the authors have studied the UA problem to minimize both the average delay and the power consumption of a downlink C-RAN. Due to the NP-hardness of such problem, they proposed a low-complexity heuristic solution according to the distance and caching of each RRH to find the solution. Moreover, to reduce the power consumption, the inefficient RRHs are turned into sleep mode. The study in [9] aims to maximize the non-convex energy efficiency in a downlink (DL) massive MIMO system. Thus, they proposed an energy-efficient low-complexity algorithm (EELCA). this approach guarantees an optimal power allocation solution based on Newtons methods and joint users association based on the Lagrange decomposition method.

The work in [10–12] and [13] have tackled the BBU-RRH association problem, also known as the RRH clustering problem. Three clustering techniques that group the RRHs into clusters have been proposed in [10]. Taking into account the users data rate requirement and the capacity of BBU, the objective was to minimize the number of activated BBUs to reduce the power consumption. In [11], the dynamic aspect of the BBU-RRH association is formulated as a Set Partitioning Problem (SPP). The aim is to jointly minimize the power consumption and the re-association rate of users experiencing a BBU change, without compromising user quality of service (QoS). A semi-static and adaptive BBU-RRH switching schemes for C-RAN are been proposed in [12], where the objective is to reduce the consumed network power by decreasing the number of active BBUs. The authors in [13] propose an energy-saving algorithm with joint RRH clustering and RRH activation problem under QoS constraints.

All cited papers attempt to minimize the energy consumption and improve user throughput either by optimizing the user association, or RRH clustering. However, the work in [14] investigated QoS-aware joint BBU-RRH mapping and user association problem in C-RAN. The purpose is to minimize the energy consumption by reducing the number of active RRHs and BBUs. Thus, they decomposed the joint problem into two sub-problems and designed a time-efficient algorithm to solve it. To reduce call blocking probability and improve the quality of service (QoS) of a user, the authors in [15] proposed a two-stage solution. In the first stage, the Markov decision process is used for user-RRH selection. In the second stage, the Ant Colony Optimization (ACO) method is applied to obtain the best BBU-RRH mapping. In our previous work [16], we considered the joint user association and RRH clustering problem in cloud radio access networks with the objective of maximizing the overall network throughput and reducing the network power consumption. First, the user association sub-problem is solved based on the SINR-based algorithm. Second, a low-complexity heuristic solution, based on the merge-and-split rules, is introduced to solve the RC sub-problem. However, such basic user association scheme (SINR-based) will lead to inefficient resource utilization and unbalanced network load. In addition, the real traffic variations during day time are ignored in [14,15] and [16], which have a serious impact on network performance as demonstrated in this article. Unlike the previous mentioned works, our study has the originality to consider the variation of traffic load conditions delivered by the CDR. The contributions of our work can be summarized as follows:

- We formulate the joint user association and RRH clustering problem in C-RAN. Our objective is to minimize network power consumption while maximizing network throughput for various traffic types (*i.e.,* elastic and streaming). We extended our previous work in [16] by proposing a novel solution for UA sub-problem relying on a game theoretic framework.
- We prove that the optimization problem is NP-hard.
- To reduce the high computational complexity of the joint problem, we decouple it into two sub-problems: the user association sub-problem and the RRH clustering sub-problem. The two sub-problems are solved in an iterative fashion until convergence is reached.
- We model the user association sub-problem as a non-cooperative game and prove that it has the Finite Improvement Path (FIP) property. Hence, a best-response algorithm permits attaining the Pure Nash Equilibrium (PNE) of the game.
- Cooperative game theory is used to address the RC sub-problem, where a heuristic solution based on the merge-and-split rules is applied. This solution can handle the non-uniform elastic and streaming traffic. It outperforms the other RC schemes, with significantly lower computational complexity.
- For a realistic scenario, we evaluate our proposed solution based on a real CDR dataset and study its convergence and complexity.
- We prove the global convergence of the decomposition approach for the joint user association and RRH clustering problem.
- By simulation, we proved the effectiveness of our solution for UA sub-problem, by comparing its performance with the optimal UA sub-problem.

The rest of the paper is organized as follows: Section 3 describes the system model. Section 4 formulates the joint optimization problem of user association and RRH clustering and discusses its complexity. Sections 5 and 6 describe our approaches to solve the user association and the RRH clustering sub-problems. Section 7 describes the implementation of user association and RRH clustering sub-problems. Simulation results are given in Section 8. Section 9 concludes the paper.

## 3. System model

We assume there are $U$ active users (UEs) and $R$ RRHs where each UE $u$ has its own throughput demand $d_u$ and is associated with at most one RRH $r$ (*i.e.,* single connectivity mode). We further suppose there are $Z$ BBUs. Each RRH $r$ can be associated with at most one BBU $z$. $Z$ is at most equal to the number of RRH $R$. We deem by $\mathcal{U} = \{u_1, u_2, \dots, u_U\}$, $\mathcal{R} = \{r_1, r_2, \dots, r_R\}$, and $\mathcal{Z} = \{z_1, z_2, \dots, z_Z\}$ the sets of UEs, RRHs and BBUs respectively. $X_{u,r}$ and $Y_{r,z}$ are two binary variables that define the user association and the RRH clustering respectively. $X_{u,r}$ is set to 1 when UE $u$ is associated to RRH $r$, and zero otherwise. $Y_{r,z}$ is set to 1 when RRH $r$ is associated to BBU $z$, and zero otherwise. Moreover, $t_r$ and $t_z$ are two binary decision variables with value 1 if RRH $r$ and BBU $z$ are turned on respectively, and zero otherwise.

The SINR of user $u$ attached to RRH $r$ and mapped to BBU $z$ is denoted by $SINR_{u,r,z}$. It can be represented as follows:

$$SINR_{u,r,z} = \frac{P_r G_{u,r}}{N_0 + \sum_{r' \neq r} (1 - y_{r',z}) P_{r'} G_{u,r'}}, \tag{1}$$

where $P_r$ is the transmit power of RRH $r$, $G_{u,r}$ is the channel gain between user $u$ and RRH $r$, and $N_0$ is the thermal noise power. In addition, $\sum_{r' \neq r} (1 - y_{r',z}) P_{r'} G_{u,r'}$ represents the inter-cluster interference caused by the RRHs that are not mapped to BBU $z$. Due to Orthogonal Frequency Division Multiple Access (OFDMA) technique applied at a cell level, the RRHs associated to the same BBU share orthogonal radio resources and act as a single cell equipped with a Distributed Antenna System (DAS), leading to intra-cluster interference cancellation among RRHs assigned to the same BBU.

Let $\hat{\eta_z}$ denote the average spectral efficiency per BBU $z$. $\hat{\eta_z}$ is given by:

$$\hat{\eta_z} = \frac{1}{u_z} \sum_{r \in \mathcal{R}} \sum_{u \in \mathcal{U}} X_{u,r} Y_{r,z} \log(1 + SINR_{u,r,z}), \tag{2}$$

where $u_z$ represents the number of users sharing the radio resources of BBU $z$ and expressed as:

$$u_z = \sum_{r \in \mathcal{R}} \sum_{u \in \mathcal{U}} X_{u,r} Y_{r,z}. \tag{3}$$

### 3.1. Throughput model

In this section, we present the throughput expression for both streaming and elastic traffics.

#### 3.1.1. Streaming traffic

Streaming traffic is designed to support real-time applications, such as video streaming and voice over IP services. The average throughput achieved by the user $u$ is limited to its demand $d_u$. More precisely, even if a user can allocate more radio resources, its achieved throughput is limited by its throughput demand.

We denote by $d_r$ the throughput demand of RRH $r$. It is defined as the sum of the throughput demands of users served by RRH $r$ and expressed as follows:

$$d_r = \sum_{u \in \mathcal{U}} X_{u,r} d_u. \tag{4}$$

We consider that radio resources in BBU $z$ are shared amongst its associated RRHs proportionally to their throughput demands. If the maximum throughput achievable by the BBU is greater than the total throughput demands of their RRHs, than the RRHs get their required number of resource blocks $RBs$ to fulfill their throughput demands. This implies that the users served by these RRHs achieve their total demands. Otherwise, when the maximum throughput of BBU is less than the RRHs throughput demands, then each RRH gets $RBs$ proportionally to its throughput demands, which ensures fairness between users with respect to their throughput demands. We denote by $T_{r,z}$ the average throughput achieved by RRH $r$ that is mapped to BBU $z$. It is expressed as follows:

$$T_{r,z} = \left[\frac{T_z^m}{\max\left(T_z^m, \sum_{r \in \mathcal{R}} Y_{r,z} d_r\right)}\right] \cdot d_r, \tag{5}$$

where $T_z^m$ is the maximum throughput achieved by BBU $z$ and is defined as follows:

$$T_z^m = W_z \cdot \hat{\eta_z}, \tag{6}$$

where $W_z$ is the channel bandwidth in BBU $z$.

We denote by $T_z$ the average throughput achieved in BBU $z$. It is the sum of the throughputs achieved by the RRHs that are attached to BBU $z$ and can be expressed as follows:

$$T_z = \sum_{r \in \mathcal{R}} Y_{r,z} T_{r,z}. \tag{7}$$

Assuming a fair resource sharing model, the average throughput achieved by user $u$, attached to RRH $r$ and mapped to BBU $z$, is expressed as:

$$T_{u,r,z} = \frac{T_{r,z}}{u_z}. \tag{8}$$

#### 3.1.2. Elastic traffic

Elastic traffic is designed to support non real-time applications, such as file transfer, email, and web applications. The average throughput achieved by a user $u$ adapts to resource availability and consequently depends on how resources are allocated among the users that are associated to the same BBU. Thus, assuming a fair resource sharing

model, and a full buffer traffic model, the average throughput achieved by user $u$ attached to RRH $r$ and mapped to BBU $z$, is expressed as:

$$T_{u,r,z} = \frac{\hat{T}_{u,r,z}}{u_z}, \tag{9}$$

where $\hat{T}_{u,r,z}$ is the peak throughput of a user $u$ attached to RRH $r$ and mapped to BBU $z$ (*i.e.,* the throughput receives by the user $u$ when associated alone to BBU $z$). It expressed as:

$$\hat{T}_{u,r,z} = W_z \log_2(1 + SINR_{u,r,z}). \tag{10}$$

Consequently, the total throughput achieved in the network, denoted $T_{total}$, is defined as the sum of the users throughputs as presented in (8) and (9). Thus, $T_{total}$ can be written as:

$$T_{total} = \sum_{z\in Z}\sum_{r\in R}\sum_{u\in U} X_{u,r}Y_{r,z}T_{u,r,z}. \tag{11}$$

### 3.2. C-RAN power consumption model

According to [17], the power consumption within the C-RAN architecture, denoted by $P_{total}$, is modeled as the sum of two terms: the power consumed by all active BBUs, coupled with the power consumed by all active RRHs. Thus, $P_{total}$ expressed as:

$$P_{total} = \sum_{z\in Z} P_z + \sum_{r\in R} P_r, \tag{12}$$

where $P_z$ and $P_r$ respectively denote the power consumed by BBU $z$ and that consumed by RRH $r$.

For streaming traffic, the power consumption at BBU $z$, $P_z$ is assumed to be a linear function of the throughput achieved at BBU $z$ [18]. It can be expressed as:

$$P_z = \begin{cases} \lambda + \mu \cdot T_z, & \text{if } t_z = 1, \\ 0, & \text{otherwise}, \end{cases} \tag{13}$$

where $\lambda$ represents the power consumption of BBU $z$ in active mode, and $\mu$ is the variation coefficient of $P_z$ as a function of $T_z$.

Further, the power consumption within a BBU $z$ in elastic traffic only depends on the number of active BBUs. Thus, it expressed as:

$$P(z) = \begin{cases} \lambda, & \text{if } t_z = 1, \\ 0, & \text{otherwise}. \end{cases} \tag{14}$$

Besides, the power consumption at RRH $r$ for both streaming and elastic traffics can be expressed as:

$$P(r) = \begin{cases} P^0 + \delta P_r, & \text{if } t_r = 1, \\ P^s, & \text{otherwise}, \end{cases} \tag{15}$$

where $\delta$ is the power amplifier efficiency, $P_r$ is the transmit power of RRH $r$, and $P^0$ and $P^s$ are the additional power consumed by RRH $r$ independently of $P_r$ in active mode and in sleep mode respectively.

### 3.3. Throughput user satisfaction model

We denote by $S(u)$ the user satisfaction model in term of throughput for both streaming and elastic traffic.

For streaming applications, such as video streaming and voice over IP services, a minimum throughput requirement is needed to achieve acceptable performance. Referring to [19], the user throughput satisfaction for streaming traffic can be modeled by a strictly increasing concave functions as illustrated in Fig. 2. In particular, this function is convex but not concave in a neighborhood around zero. Under a minimum acceptable throughput, and a maximum throughput, a Sigmoid (S-shape) function that has all the characteristics mentioned above can be used to represent the user satisfaction model for streaming traffic. It can be written as:

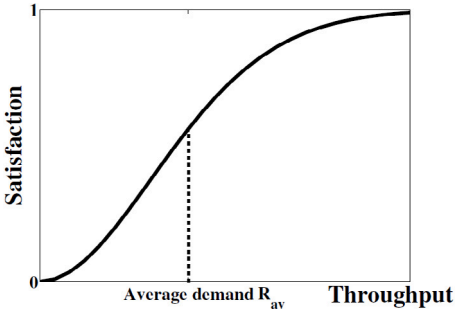


**Fig. 2.** Streaming sessions: Throughput satisfaction function.

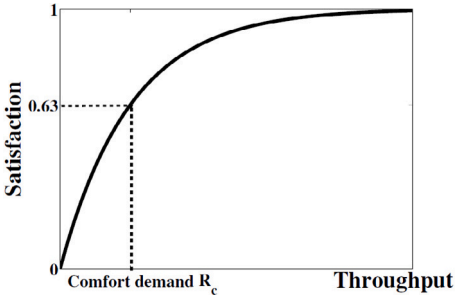


**Fig. 3.** Elastic sessions: Throughput satisfaction function.

$$S(u) = \begin{cases} 0, & T_{u,r,z} \le d_u^{\min}, \\ 2\left(\frac{T_{u,r,z}-d_u^{\min}}{d_u^{\max}-T_{u,r,z}}\right)^2, & d_u^{\min} < T_{u,r,z} < d_c, \\ 1-2\left(\frac{T_{u,r,z}-d_u^{\min}}{d_u^{\max}-T_{u,r,z}}\right)^2, & d_c < T_{u,r,z} < d_u^{\max}, \\ 1, & T_{u,r,z} \ge d_u^{\max}, \end{cases} \tag{16}$$

where $d_c = (d_u^{\min} + d_u^{\max})/2$.

Besides, the elastic traffic is designed to support traditional data services (*i.e.,* file transfer, email and web traffic). Such applications use the TCP protocol that adjusts the sending data rate to the resource availability (*i.e.,* load conditions of the network). Since the elastic traffic does not require a minimum throughput, it can be ignored in the user satisfaction model. According to [19], the user satisfaction model of elastic traffic is an increasing, strictly concave and continuously differentiable function. Their approximated function is displayed in Fig. 3. Intuitively, they also appear to have decreasing marginal improvement due to incremental increase in throughput. This means that when the realized throughput exceeds a particular value, the user satisfaction will increase slowly as a function of throughput. This value is the comfort throughput $R_c$ beyond which user satisfaction exceeds 63% of maximum satisfaction as shown in Fig. 3. An exponential function that has all the characteristics mentioned above can be used to represent the user satisfaction model of elastic traffic. It is expressed as follows:

$$S(u) = 1 - \exp\left(\frac{-d_u^{\max}}{R_c}\right), \tag{17}$$

where $d_u^{\max}$ is equal to the average throughput achieved by user $u$ in Eq. (9).

### 3.4. Network utility function

We introduce a network utility function denoted by $U$. It is defined as a weighted linear combination of the total network throughput $T_{total}$

and the total network power $P_{total}$:

$$U = \alpha\alpha' T_{total} - \beta\beta' P_{total}, \tag{18}$$

where $\alpha'$ and $\beta'$ are two scaling factors applied to normalize the elements of $U$ to the same range ([0, 1]) , and $\alpha$ and $\beta$ are the weights attributed to the total network throughput and the total network power respectively. Note that $\alpha$ and $\beta$ are between 0 and 1, and $\alpha + \beta = 1$.

## 4. Optimization problem formulation

The optimization problem ($\mathcal{P}$) aims to find the optimal user association and RRH clustering decisions, that maximizes the network utility function $U$:

$$\underset{X,Y}{\text{maximize}} \quad U(X,Y) \tag{19}$$

$$\text{subject to} \quad \sum_{r\in\mathcal{R}} X_{u,r} \leq 1, \ \forall u \in \mathcal{U} \tag{20}$$

$$\sum_{z\in\mathcal{Z}} Y_{r,z} \leq 1, \ \forall r \in \mathcal{R} \tag{21}$$

$$X_{u,r} \leq t_r, \ \forall r \in \mathcal{R} \tag{22}$$

$$Y_{r,z} \leq t_z, \ \forall z \in \mathcal{Z} \tag{23}$$

$$X_{u,r}, Y_{r,z}, t_r, t_z \in \{0,1\}. \tag{24}$$

Constraints (20) imply that each user $u$ can only be associated with one RRH $r$. Constraints (21) imply that each RRH can at most be attached to one BBU. Constraints (22) indicate that RRH $r$ is turned on only when at least one user $u$ is associated with it. Constraints (23) indicate that BBU $z$ is activated only when at least one RRH $r$ is attached to it, and finally constraints (24) indicate that $X_{u,r}$, $Y_{r,z}$, $t_r$ and $t_z$ are binary variables.

### 4.1. Optimal solution complexity

The optimization problem ($\mathcal{P}$) of our approach is a mixed-integer non-linear programming problem (MINLP) , which is NP-hard (please see the proof in Appendix A). The optimal solution can be obtained through exhaustive search. However, all possible user-RRH associations should be considered, which requires $O(R^U)$ operations. Moreover, all possible number of RRH-BBU configurations should be explored, which is given by the $R$th Bell number and denoted by $B_R$. Consequently, the computational complexity for obtaining the optimal solution is in $O(B_R.R^U)$. Nevertheless, when the number of parameters (*i.e.,* the number of users and RRHs) is large, this problem becomes intractable. To overcome this complexity, we decompose the joint problem into two sub-problems including the user association (UA) sub-problem and the RRH clustering (RC) sub-problem (*i.e.,* the complexity of the optimal UA and RC sub-problems is given in Appendix B). First, the UA sub-problem is modeled as a non-cooperative game among competing UEs to choose a suitable RRH. When the Nash Equilibrium of the UA sub-problem is attained, cooperative game theory is used to address the RC sub-problem, where a heuristic solution based on the merge-and-split rules is applied. Further, depending on the clusters that have been recently formed, user associations may be reconsidered. The two sub-problems are sequentially repeated until convergence is reached, as portrayed in Algorithm 1.

**Algorithm 1** Iterative Approach for the Joint User Association and RRH Clustering Problem

1: **Requires:** RRH clusters, user and RRH positions.
2: **Initialize:** Solve the UA sub-problem based on non-cooperative game theory.
3: **Repeat:**
4: Solve the RC sub-problem using the merge-and-split rules.
5: Re-associate users according to the new RRH clustering.
6: **Until**: No more UA and RC need to be modified.

## 5. UA sub-problem

In this section, we investigate the UA sub-problem under dynamic traffic load variations. In particular, we aim to solve the re-association between UEs and BBUs that occur when network load conditions vary. More precisely, the fluctuation in the number of users during the day time impacts BBUs load conditions. On the one hand, UEs connected to the crowded BBUs might affect the users QoS. On the other hand, UEs connected to the less crowded BBUs might move to different ones to turn off lowly loaded BBUs to reduce the consumed power. Since UA decisions impact directly the users QoS and the number of active BBUs, we modeled such sub-problem as a non-cooperative game among competing UEs. This solution can be implemented in a fully distributed manner, which enables fast adaptation to traffic and channel fluctuations while reducing the amount of signaling load overhead and complexity. In such game, each user chooses a suitable strategy that minimizes the data transmission delay (*i.e.,* depicted the QoS) and the power consumption (*i.e.,* represented the number of active BBUs). This leads to balance the BBUs charge, so that the network resists to any possible fluctuation in the load conditions.

First, each user $u$ chooses a candidate set of two RRHs belonging to different BBUs (RRH $r' \in$ BBU $z'$ and RRH $r'' \in$ BBU $z''$), such that those RRHs endow user $u$ with the strongest downlink signal. Then, user $u$ singles out the BBU (among BBU $z'$ and BBU $z''$) that minimizes its cost function. The distributed user association will be portrayed as an unweighted crowding game.

Non-cooperative game theory models the interactions between players competing for a common resource. Hence, it is well adapted to model the user association scheme. We define a multi-player game $\mathcal{G}_{\mathcal{U},\mathcal{A}}$ between users which are assumed to make their decisions without knowing the decisions of each other.

The formulation of this non-cooperative game $\mathcal{G}_{\mathcal{U},\mathcal{A}} = \langle \mathcal{U}, S, \mathbf{C} \rangle$ can be described as follows:

- The finite set of users $\mathcal{U}$.
- The space of pure strategies $S$ formed by the Cartesian product of each set of pure strategies $S = S_1 \times S_2 \times \cdots \times S_U$, where the strategy space of any user $u$ is $S_u = \{BBU\ z', BBU\ z''\}$ with $z', z'' \in \mathcal{Z}$, where RRH $r'$ and RRH $r''$ are the two best detected antennas by user $u$ such as RRH $r' \in$ BBU $z'$ and RRH $r'' \in$ BBU $z''$.
  Finally, if user $u$ selects BBU $z'$, it is associated with RRH $r'$, and hence $X_{u,r'} = 1$ and $X_{u,r\neq r'} = 0$.
- We denote by $a_u = \{a_u^{z'}, a_u^{z''}\}$ the action taken by user $u$ which chooses one of available BBUs in $S_u$, where

$$a_u^z = \begin{cases} 1, & \text{if user } u \text{ chooses BBU } z \text{ comprising RRH } r \\ 0, & \text{otherwise} \end{cases} \tag{25}$$

- A set of cost functions $\mathbf{C} = (C_1, C_2, \ldots, C_U)$ that determine the players' preferences over the possible outcomes of the game that are given by the particular action $a_u$ chosen by user $u$ and the particular actions chosen by all other players.

### 5.1. Cost function

We denote by $C_u$ the cost function of user $u$ following its action $a_u$ and the action $a_{-u}$ of other users except $u$. It is defined as the weighted sum of the power consumed by a BBU chosen by a user and the delay endured by this user. More precisely, the power consumed by BBU $z$ serving RRH $r \in S_u$ chosen by user $u$ following its action $a_u$ can be written as:

$$P_u^z(a_u) = \sum_{z\in\mathcal{Z}} a_u^z P^z = a_u^{z'} P^{z'} + (1 - a_u^{z'}) P^{z''}, \tag{26}$$

where $P^z$ is defined in (13) and (14). Note that $P_u^z(a_u)$ does not depend on the action of other users.

Furthermore, $T_u(a_u, a_{-u})$ is the delay endured by user $u$ following its action $a_u$. It can be expressed as:

$$\begin{aligned} T_u(a_u, a_{-u}) &= \sum_{z\in\mathcal{Z}}\sum_{r\in\mathcal{R}} X_{u,r}Y_{r,z}\frac{N_z + X_{u,r}Y_{r,z}}{W_z \log_2(1+SINR_{u,r,z})} \\ &= X_{u,r'}Y_{r',z'}\frac{N_{z'} + X_{u,r'}Y_{r',z'}}{W_{z'}\log_2(1+SINR_{u,r',z'})} \\ &+ X_{u,r''}Y_{r'',z''}\frac{N_{z''} + X_{u,r''}Y_{r'',z''}}{W_{z''}\log_2(1+SINR_{u,r'',z''})} \\ &= \frac{a_u^{z'}(N_{z'} + a_u^{z'})}{\hat{T}_{u,r',z'}} + \frac{(1-a_u^{z'})(N_{z''}+1-a_u^{z'})}{\hat{T}_{u,r'',z''}}, \end{aligned} \tag{27}$$

where $N_z$ is the total number of users in BBU $z$ except user $u$.

Consequently, the cost function $C_u$ can be written as follows:

$$C_u(a_u, a_{-u}) = A \cdot P_u^z(a_u) + B \cdot T_u(a_u, a_{-u}), \tag{28}$$

where $A$ and $B \in [0, 1]$, and $A + B = 1$.

The cost sustained by a given user $u$ in any selected strategy depends upon the congestion impact inflected by other users $u'$ sharing the same radio resources of the common BBU. The latter is depicted by $T_u$ and acts as a load balancing function among active BBUs. Furthermore, it is related to the total power consumed within BBU $z$ chosen by user $u$ as expressed by $P_u^z$. Finally, user $u$ chooses to be associated with one of the best detected RRH served by the BBU that endows it with the lowest cost, which sets the variables $X_{u,r}$ for user $u$.

This UA game enhances both network load balancing and user radio conditions. Note that, at this stage, the RRH-BBU mappings are known (set by the RC sub-problem). Afterward, considering the output of the UA sub-problem (*i.e.,* user association variables $X_{u,r}$), the RRH clusters may change in a way to minimize both the total transmission delay and the network power consumption. This directly impacts user radio conditions and can lead to user re-associations.

### 5.2. Attaining the Nash equilibrium

In a non-cooperative game, an efficient solution is obtained when all players adhere to a Nash Equilibriums (NE). A NE is a profile of strategies in which no player will profit from deviating its strategy unilaterally. Hence, it is a strategy profile where each players' strategy is an optimal response to the other players' strategies.

Our game $\mathcal{G}_{\mathcal{UA}}$ is a finite game and in general such games are not guaranteed to have a Pure NE (PNE). Nevertheless, they possess a mixed NE where each player has to continually change its BBU selection according to a distribution probability over the strategy set. However, our game has the Finite Improvement Path (FIP) property which is a valuable property. In fact, for such games, simple dynamics converge to PNE [20,21]. The common feature of such games is the existence of a potential function which represents commonly the quality of the various strategy profiles for all players [22]. The unilateral change of one user strategy results in a change of its cost function that is equal to the change of the so-called potential function.

**Proposition 1.** *The game $\mathcal{G}_{\mathcal{UA}}$ has the FIP property.*

**Proof.** The game $\mathcal{G}_{\mathcal{UA}}$ is an unweighted crowding game, as it is a normal-form game in which users share a common set of actions and the cost function of user $u$ for choosing a particular action is player specific and a non-increasing function of the total number of users choosing that same action. Unweighted crowding games have PNE. Furthermore, when players have only two strategies (choosing between BBU $z'$ and BBU $z''$), the game has the Finite Improvement Path (FIP) property and hence a best-response algorithm permits attaining the PNE of the game [23].

### 5.3. The best-response algorithm

As already mentioned, best-response dynamics permit attaining PNE for our UA game. Accordingly, at each round of the best-response algorithm, each user $u$ chooses to be associated with one of the BBU belonging to its strategy set that endows it with the lowest cost according to (28) until convergence (*i.e.,* when all users choose the same strategies as in the previous iteration). The complexity analysis of the Best-Response algorithm is studied in Appendix C.

### 5.4. Centralized approach for UA

In this section, we formulate the centralized optimization problem for UA sub-problem with aims at minimizing the total cost function $C_u$ for all users. Contrarily to our distributed approach, where the users selection is limited between two RRHs, each user $u$ can select the best RRH among all RRHs that minimizes its cost function $C_u$. Assuming a given RRH clustering (*i.e.,* RRH clustering variables $Y_{r,z}$ are known), the term $P_u$ in (26) becomes constant. By omitting this constant, the cost function $C_u$ of user $u$ in (28) can be expressed as follows:

$$\begin{aligned} C_u(x) &= B\sum_{z\in\mathcal{Z}}\sum_{r\in\mathcal{R}} X_{u,r}\frac{N_z + X_{u,r}}{W_z\log_2(1+SINR_{u,r,z})} \\ &= B\sum_{z\in\mathcal{Z}}\sum_{r\in\mathcal{R}} X_{u,r}\frac{\sum_{u'\neq u} X_{u',r} + 1}{W_z\log_2(1+SINR_{u,r,z})} \\ &= B\sum_{z\in\mathcal{Z}}\sum_{r\in\mathcal{R}} \frac{\sum_{u'\neq u} X_{u,r}X_{u',r} + X_{u,r}}{W_z\log_2(1+SINR_{u,r,z})} \end{aligned} \tag{29}$$

The optimal solution ($\tilde{C}$) consists in finding the user association that minimizes the total cost $C_u\ \forall u \in \mathcal{U}$. Therefore, ($\tilde{C}$) is given by:

$$\underset{x}{\text{minimize}} \quad \sum_{u\in\mathcal{U}} C_u(x) \tag{30}$$

$$\text{subject to} \quad \sum_{r\in\mathcal{R}} X_{u,r} \leq 1,\ \forall u \in \mathcal{U} \tag{31}$$

$$X_{u,r} \in \{0, 1\}. \tag{32}$$

This optimization problem is an integer non-linear problem (INLP) which is a complex problem. To simplify the resolution of such problem, we reformulate it as an integer linear problem (ILP) by replacing the product of the two binary variables $X_{u,r}$ and $X_{u',r}$ with a new binary variable $Z_{u,u',r}$, and add the new following constraints:

$$Z_{u,u',r} \leq X_{u,r} \tag{33}$$

$$Z_{u,u',r} \leq X_{u',r} \tag{34}$$

$$Z_{u,u',r} \geq X_{u,r} + X_{u',r} - 1 \tag{35}$$

$$Z_{u,u',r} \in \{0, 1\}. \tag{36}$$

Note that, the integer linear problem formulation can be easily solved using CVX tool in MATLAB.

## 6. RC sub-problem

The RC sub-problem determines the BBU-RRH association in a way to maximize the network utility. To solve the RC sub-problem, we propose a cooperative game among RRHs. More specifically, based on the merge-and-split rules, the RRHs cooperate and arrange themselves into disjoint independent clusters, to maximize the network utility:

- Clusters $\{c_1, c_2, \ldots, c_l\}$ are merged into one, if the resulting cluster provides a higher network utility:

$$U(\bigcup\nolimits_{i=1}^{l} c_i) > \sum_{i=1}^{l} U(c_i) \tag{37}$$

The merging process stops when no more preferable clusters can be further formed.

- A cluster $\bigcup_{i=1}^{l} c_i$ is splitted into smaller ones $\{c_1, \ldots, c_l\}$, if this provides a higher network utility:

$$\sum_{i=1}^{l} U(c_i) > U(\bigcup_{i=1}^{l} c_i) \tag{38}$$

The splitting process stops when no more improvement can be achieved by the formation of smaller clusters.

These two processes are sequentially repeated until convergence is reached (*i.e.,* no more merge-and-split can be further done). The complexity analysis of merge-and-split operations is presented in Appendix D. According to [24], since our heuristic consists of successive merge-and-split operations, it leads to a $\mathbb{D}_{hp}$-stable partition.

**Proposition 2.** *The decomposition problem of the joint UA and RC converges after a limited number of iterations.*

**Proof.** After a certain number of iterations, the association of UA and RC changes as follows:

$$\phi_0 \rightarrow \phi_1 \rightarrow \phi_2 \rightarrow \cdots \rightarrow \phi_l \tag{39}$$

where $\phi_0$ is the initial UA and RC associations. For iteration $l$, the association changes from $\phi_{l-1}$ to $\phi_l$. We denote by $U_l$ and $U_{l-1}$ the network utility for $l$ and $l-1$ iterations respectively. Therefore, we have $U_l > U_{l-1}$ (*i.e.,* the network utility $U$ increases for each iteration). This contradicts the fact that the network utility $U$ has an upper bound due to the limited network throughput and power consumption. Consequently, the problem converges after a finite number of iterations.

## 7. UA and RC sub-problems implementation

In this section, we describe the implementation of UA and RC sub-problems. We define two time scales: the large time scale, which corresponds to each hour of the day, and the small time scale that represents the user arrivals. Given the RRH clustering, at each user arrival, the incoming user will be associated based on the UA non-cooperative game described in 5. It chooses to be mapped with one of the best detected RRH served by the lowest loaded BBU. As the user will be served by the BBU with the higher resource availabilities, its throughput demands can be satisfied with high probability. In the worst case scenario, where user throughput is not satisfied, the user association and the RRH clustering will be updated at each hour of the day to enhance users QoS in terms of throughput demands. The joint execution of UA and RC sub-problems improves the network utility function defined in (18). In conclusion, the decoupling of the joint centralized problem into two sub-problems not only as a workaround to the NP-hardness of the original problem but to strike two other goals. The first goal concerns the network design as we intend to tackle the UA and RC problems in a distributed fashion by having recourse to game theory: UA is handled appropriately by the end-users and RC by the RRHs. The second goal reached through that the original problem splitting is enabling us to execute our two solutions for UA and RC in two different time scales which makes our approach applicable for real-time traffic that mandates low latency. More precisely, the UA solution is applied at each user arrival to immediately associate it to an active RRH, while the UA and the RC sub-problems will be jointly executed at each hour of the day to enhance the total network performance.

## 8. Simulation results

To show the effectiveness of our heuristic solution for the RC sub-problem in adapting to non-uniform traffic load for both streaming and elastic traffics, we compare it with the no-clustering solution, where one BBU is exclusively dedicated to each RRH, and the grand coalition, where all RRHs are associated with a single BBU. We also compare our distributed solution (best-response BR) for the UA sub-problem with the optimal solution in presence of elastic traffic.

**Table 1**
Simulation parameters.

| *Parameter* | *Value* |
|---|---|
| $P_r, \forall r$ | 10 W |
| $P^0$ | 6.8 W |
| $P^s$ | 4.3 W |
| $\delta$ | 4 |
| $\lambda$ | 40 W |
| Cell radius | 500 m |
| $W_z, \forall z$ | 20 MHz |
| $N_0$ | −174 dBm/Hz |

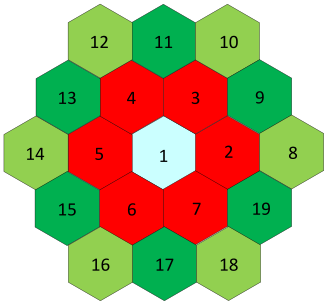


**Fig. 4.** Network with 19 RRHs.

The simulation results were obtained using Matlab software on a machine with Intel Core i5, 2.5 GHz Processor and 8 GB RAM. For illustration, we consider a network composed from 19 RRHs as presented in Fig. 4. The Cost-231 Hata model is applied to determine the channel gains. The simulations are run over 50 iterations, and the performance metrics are averaged and shown with 95% confidence intervals. We set the weighting factors $\alpha$, $\beta$, $A$ and $B$ of the network and user cost functions (cf. Eqs. (18), (28)) to 0.5. The rest of the simulation parameters are summarized in Table 1.

We utilize the Call Detail Records (CDR) presented in [4]. This is a data structure that contains spatial and temporal data information about users traffic and association. More precisely, for each user, we know:

- The user identity.
- The user demand as a function of the date and time.
- The user-RRH association.

### 8.1. Streaming sessions

We respectively depict in Figs. 5 and 6 the traffic load and the number of active BBUs during day time. Regardless of the traffic load in the network, the no-clustering solution for the RC sub-problem activates more BBUs in comparison with our heuristic and grand coalition solutions. In fact, each active RRH is mapped to one BBU. Thus, this number varies according to the number of active RRHs (cf. Fig. 7). However, for the same traffic load, our heuristic solution reduces the number of active BBUs in comparison with the no-clustering solution. We notice that at low traffic load (from 0:00 to 4:00), our solution activates only one or two BBUs to handle user demands. At high traffic load (at 10:00 and 17:00), up to 12 BBUs are activated to meet the increasing traffic demands. As a conclusion, our approach adapts the RRH clustering (*i.e.,* number of active BBUs) to support efficiently the non-uniform traffic load. Moreover, when the grand coalition solution is used to solve the RC sub-problem, only one BBU is activated. As a matter of fact, all RRHs are mapped to one BBU regardless of the traffic load in the network.

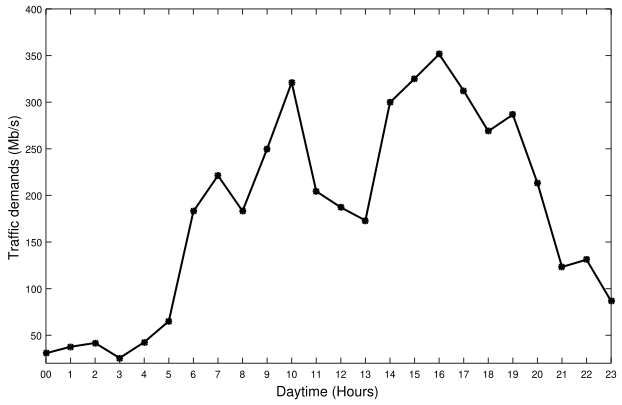


**Fig. 5.** Traffic load during day time (hours).

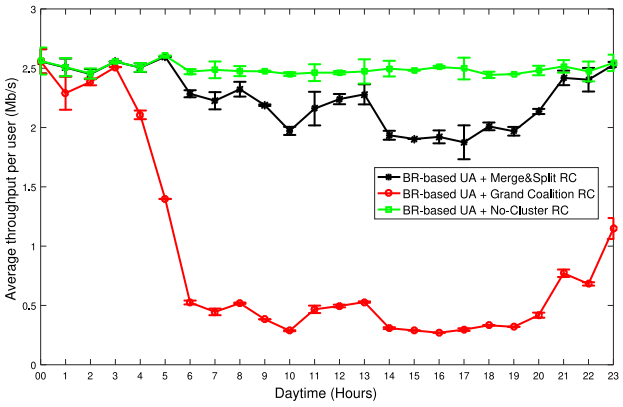


**Fig. 8.** Average user throughput: Streaming sessions scenario.

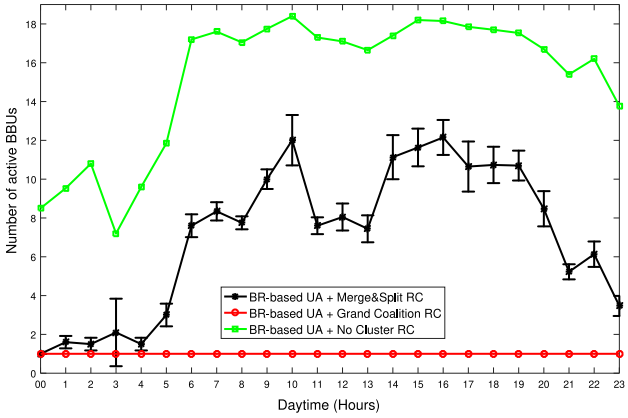


**Fig. 6.** Number of active BBUs: Streaming sessions scenario.

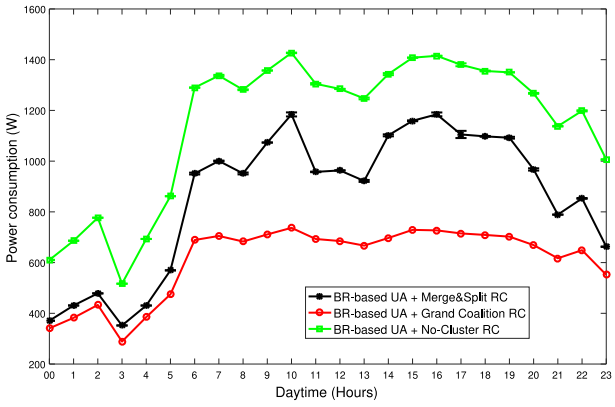


**Fig. 9.** Power consumption: Streaming sessions scenario.

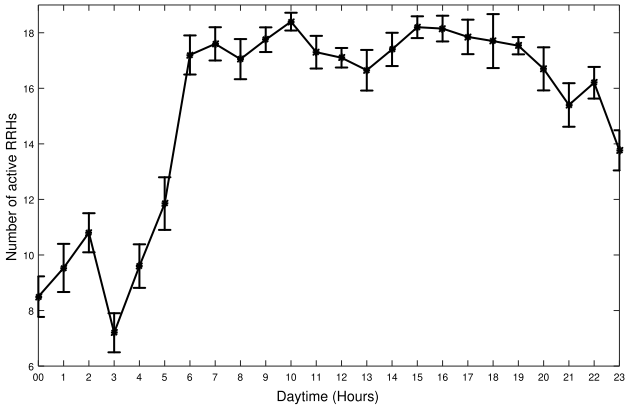


**Fig. 7.** Number of active RRHs: Streaming sessions scenario.

Fig. 8 shows the average user throughput during day time. Since the throughput achieved by a user depends on the number of available resources, the no-clustering solution for the RC sub-problem ensures the highest user throughput, equal to the users demand, owing to the availability of radio resources (as the number of active BBUs is maximized). Besides, at low traffic load (from 0:00 to 4:00), the grand coalition solution provides user throughput close to that of the no-clustering scheme. However, when the traffic load increases, the user throughput realized by this solution decreases dramatically. This is due to the limited number of resources provided by the grand coalition solution (*i.e.,* as only one BBU is activated) to support the growth of traffic load. Moreover, our heuristic solution ensures a sufficient number of resources to meet users demand under all traffic load conditions. As shown in Fig. 6, the number of active BBUs varies according to the traffic load in the network. As a result, it achieves user throughput close to that of the no-clustering solution most of the time.

Fig. 9 presents the power consumed by the network during day time. It depends on the number of active BBUs, RRHs, and their realized throughput (according to (13)). Since the no-clustering solution for the RC sub-problem achieves almost constant user throughput (*i.e.,* equal to users demand), the variation of power consumption follows the variation of activated BBUs. Thus, this leads to higher power consumption. On the other hand, the grand coalition for the RC sub-problem provides the lowest consumed power by activating solely one BBU. Further, this solution produces almost a constant user throughput between 6:00 and 23:00, leading to a constant power consumption in this interval. Moreover, our heuristic solution adapts the number of active BBUs according to the traffic load in the network (cf. Fig. 6). Consequently, it consumes less power than the no-clustering method under all traffic conditions.

Furthermore, Fig. 10 shows the throughput user satisfaction during day time. It is calculated using the sigmoid function presented in (16). When the no-clustering solution is used for the RC sub-problem, the throughput user satisfaction is almost 100% during all day time. In fact, this solution maximizes the availability of resources (the number of active BBUs) to meet the users demand, leading to the highest user satisfaction. Besides, when the grand coalition solution is applied, the users demand is fully satisfied at low traffic load. However, when the traffic load increases in the network, this satisfaction decreases dramatically to be 0% (between 6:00 and 20:00). As a matter of fact, the scarce amount of resources provided by activating one BBU fails to fulfill the minimum acceptable user throughput defined as 80% of users

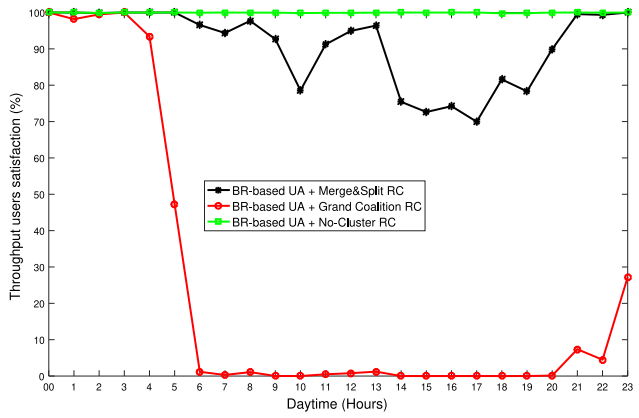


**Fig. 10.** Throughput user satisfaction: Streaming sessions scenario.

demand [25]. This leads to the lowest user satisfaction level. Moreover, at low traffic load, our heuristic solution ensures the maximum user satisfaction level with few active BBUs. Yet, when the total throughput demand increases, the satisfaction level decreases to reach almost 70% at 18:00. Typically, our approach activates more BBUs to support the increasing traffic (cf. Fig. 6), but not enough to fulfill the users demand as the case of the no-clustering solution. In fact, it provides a tunable trade-off between achievable user throughput and power consumption to maximize the network utility defined in (18).

Fig. 11 illustrates the network utility during day time. As the no-clustering solution activates the highest number of BBUs (cf. Fig. 6), it provides the highest throughput (cf. Fig. 8), but at the cost of higher power consumption (cf. Fig. 9). However, the grand coalition leads to the lowest power consumption, as only one BBU is active, but at the cost of lower throughput. This explains why, when a few active BBUs are enough to meet the users demand (from 00:00 to 04:00), the network utility achieved by the grand coalition solution outperforms that of the no-clustering solution. Moreover, when the throughput demand increases, more active BBUs are needed to support the traffic load. In this situation, the network utility provided by the no-clustering solution exceeds that of the grand coalition (beyond 6:00). In addition, when our heuristic solution is used, enough BBUs are activated to balance between high throughput and low power consumption. More precisely, by enabling few BBUs, our proposed solution provides very close network utility to that of the grand coalition at low traffic load. Furthermore, at high traffic load, additional BBUs are activated to meet the throughput demand. Nevertheless, our proposed solution consistently provides significantly lower power consumption in comparison with when the no-clustering solution (cf. Fig. 9). As a conclusion, our approach realizes a good trade-off between the user throughput and the overall power consumption to efficiently handle the non-uniform while maximizing the network utility. This leads to the highest network utility under all traffic conditions.

To illustrate the significant gain that can be achieved by our approach for the RC sub-problem, we present in Figs. 12 and 13 the cumulative power consumption and the cumulative network utility respectively. We can see that when our solution is used, the power consumption can be reduced to half in comparison with the no-clustering solution at the end of the day. Moreover, our solution can achieve up to 33% and 53% higher network utility in comparison with the no-clustering and the grand coalition methods respectively. Furthermore, the grand coalition solution consumes the lowest power, but provides the lowest utility at the end of the day.

### 8.2. Elastic sessions

We depict in Figs. 15 and 16 the number of active BBUs and RRHs in presence of elastic traffic during day time. In this context,

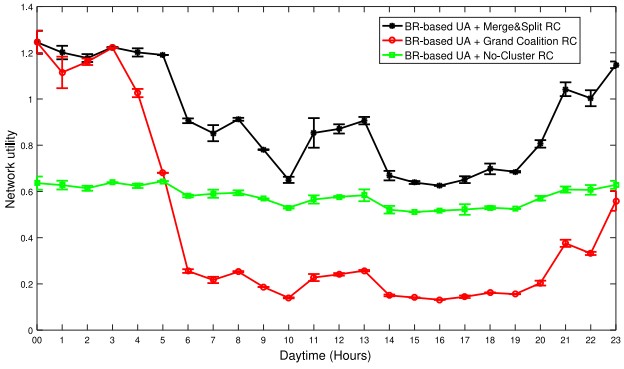


**Fig. 11.** Network utility: Streaming sessions scenario.

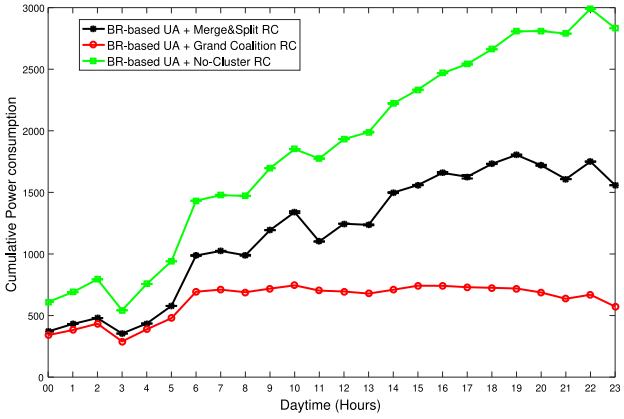


**Fig. 12.** Cumulative power consumption: Streaming sessions scenario.

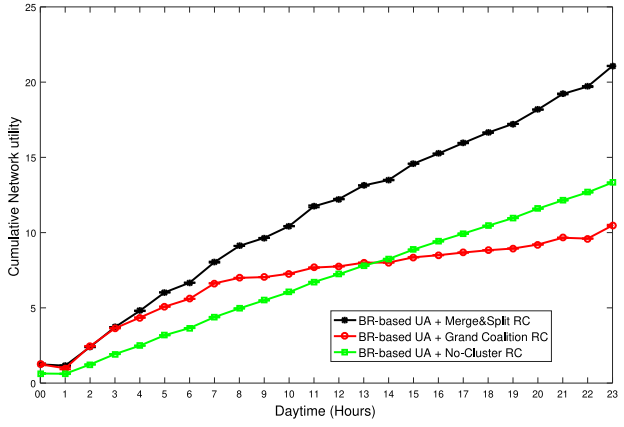


**Fig. 13.** Cumulative network utility: Streaming sessions scenario.

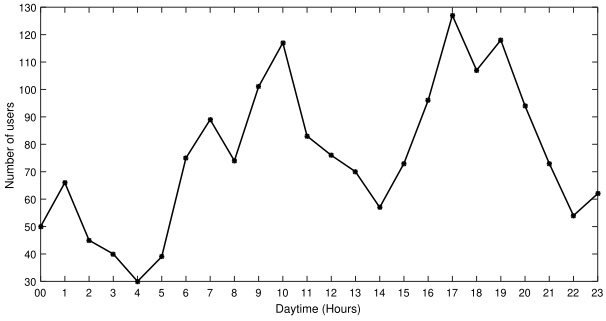


**Fig. 14.** Users number during day time (hours).

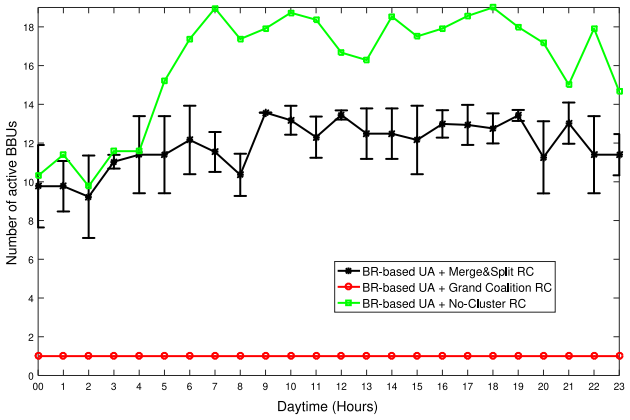


**Fig. 15.** Number of active BBUs: Elastic sessions scenario.

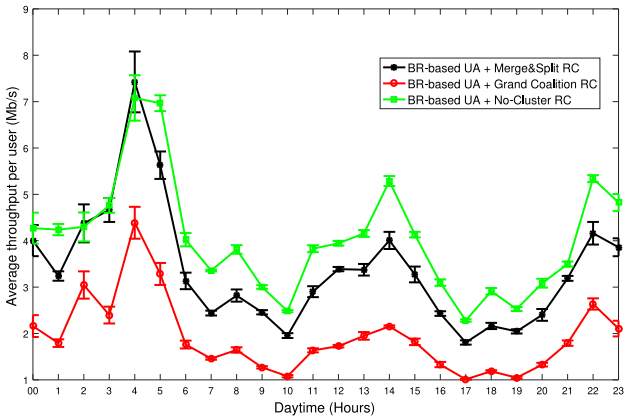


**Fig. 17.** Average user throughput: Elastic sessions scenario.

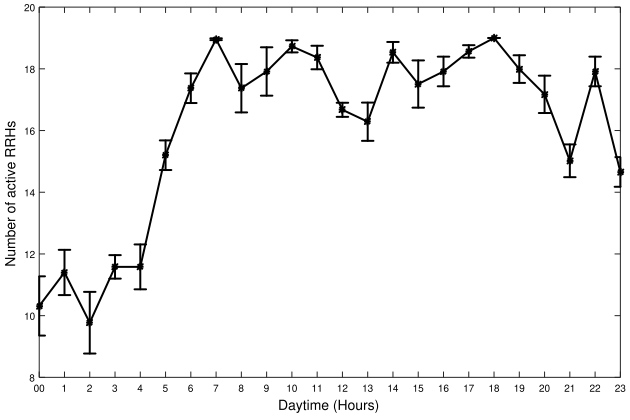


**Fig. 16.** Number of active RRHs: Elastic sessions scenario.

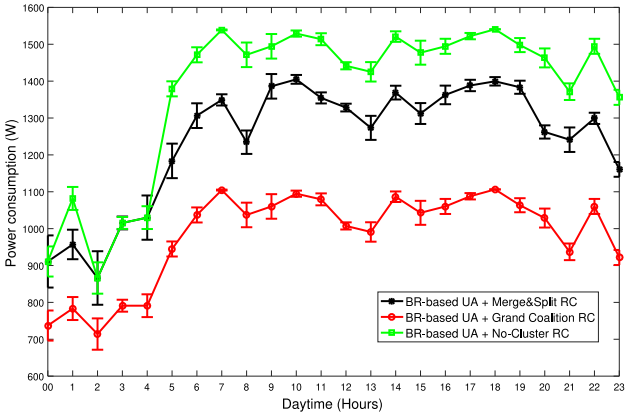


**Fig. 18.** Power consumption: Elastic sessions scenario.

the perceived user throughput depends on the availability of shared resources and is not constrained to users demand. Thus, when our heuristic solution is used to support elastic traffic, it almost activates the same number of BBUs given by the no-clustering solution when the number of RRHs is below 12. In fact, by maximizing the number of active BBUs (*i.e.,* maximizes the availability of shared resources), our proposed solution provides a close user throughput to the no-clustering solution as shown in Fig. 17. However, when the number of RRHs increases in the network, our solution seeks to find the best trade-off between increasing user throughput and decreasing power consumption. This explains the slight variation in the number of active BBUs during the day contrarily to streaming traffic scenario where our solution adapts the number of active BBUs to the traffic load condition (cf. Fig. 6). Besides, the number of activated BBUs given by the no-clustering depends exclusively on the number of active RRHs. More specifically, this number is equivalent to the number of serving RRHs as shown in Fig. 16. Moreover, regardless of the number of users in the network, the number of activated BBUs given by the grand coalition scheme is unchanged (*i.e.,* equal to 1), since all RRHs are mapped to one BBU.

In Figs. 17 and 18, we present the user throughput and the power consumption during day time. We note that, the perceived user throughput for elastic traffic depends on the availability of resources shared by all users in the network. Thus, when the number of users increases, the user throughput decreases in all solutions used to solve the RC sub-problems. However, the no-clustering solution provides the highest throughput in comparison with our heuristic and the grand coalition solution (cf. Fig. 17). Nevertheless, this comes at the cost of higher power consumption as shown in Fig. 18. Yet, when the grand coalition is used for the RC sub-problem, it leads to the lowest power consumption, as a single BBU is activated. Also, it provides a lower throughput due to the limited number of resources. In comparison with the no-clustering scheme, our proposed solution provides lower power consumption most of the time with close user throughput (cf. Fig. 17). On the other hand, it ensures higher user throughput but also higher power consumption in comparison with the grand coalition.

Fig. 19 shows the throughput user satisfaction in presence of elastic traffic during day time. It is calculated using the exponential function presented in (17). We assume that the comfort throughput $R_c$ is equal to 1.5 Mb/s, and $d_u^{\max}$ is the average user throughput illustrated in Fig. 17. We note that, when $d_u^{\max} = R_c$, the user satisfaction level is equal to 63%. Since the grand coalition solution for the RC sub-problem frequently achieves an average throughput below the comfort throughput $R_c$, the satisfaction level is less than 63% most of the time. Particularly, at 8:00 and 11:00, the grand coalition provides an average throughput equal to $R_c$, that is why the satisfaction level is equal to 63% at these periods. Besides, the no-clustering solution ensures an average throughput much higher than $R_c$, which leads to fulfilling user satisfaction all day time. Moreover, the user satisfaction level given by our heuristic solution is always higher than when the grand coalition is used. In addition, it provides an average throughput higher than $R_c$ (cf. Fig. 17), leading to a satisfaction level up to 63% during all day time.

Furthermore, Fig. 20 illustrates the network utility in presence of elastic traffic during day time. When our heuristic is adopted, it consumes less power compared to the no-clustering solution (cf. Fig. 18), while providing a close average throughput (cf. Fig. 17). As a consequence, the network utility given by our approach outperforms

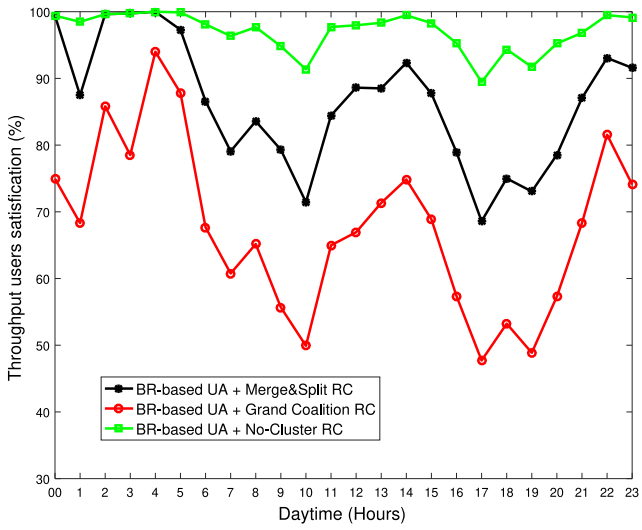


**Fig. 19.** Throughput user satisfaction: Elastic sessions scenario.

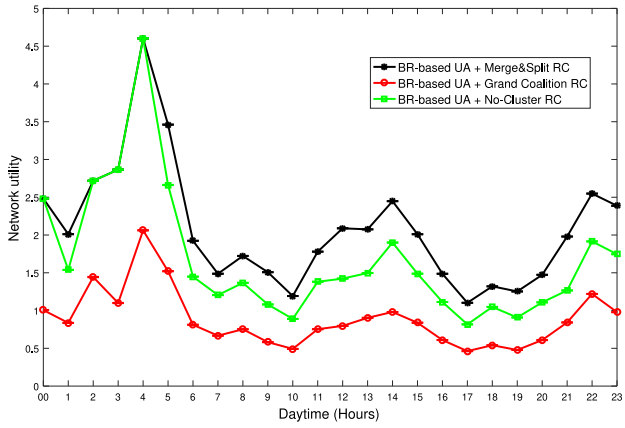


**Fig. 20.** Network utility: Elastic sessions scenario.

that of the no-clustering solution. Besides, in comparison with our proposed solution and the no-clustering solution, the grand coalition provides lower power consumption but also lower average throughput. As a result, it leads to the lowest network utility during all day time. Moreover, at the end of the day, our approach introduces a significant performance gain in term of network utility in comparison with the two other solutions for the RC sub-problem, as presented in Fig. 21. We can see that our approach can increase the network utility up to 25% compared to the no-clustering method (*i.e.,* the network utility increases from 35 to 47), and up to 57% compared to the grand coalition (*i.e.,* the network utility increases from 20 to 47).

Figs. 22 and 23 show the number of iterations required to reach convergence and the execution time necessary for both UA sub-problem and UA & RC sub-problems (BR + Merge and Split) respectively. Our solution for UA sub-problem, that is employed at each user arrival, converges in average of 2 iterations in almost 0.4 ms as displayed in Fig. 22 and Fig. 23. Moreover, our iterative approach for UA sub-problem and RC sub-problem converges in a maximum of 9 iterations in almost 9.5 ms (cf. Figs. 22 and 23). Note that this iterative approach is executed once each hour of the day (*i.e.,* long-term period). The results show that our iterative approach has the potential to be applied for real-time traffic that requires low latency in two different time-scale: large time scale and small time scales.

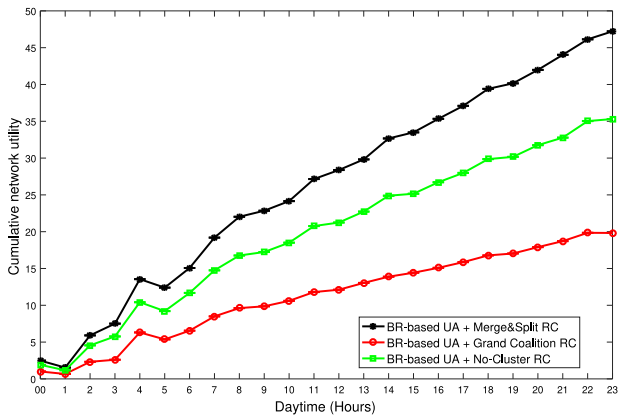


**Fig. 21.** Cumulative network utility: Elastic sessions scenario.

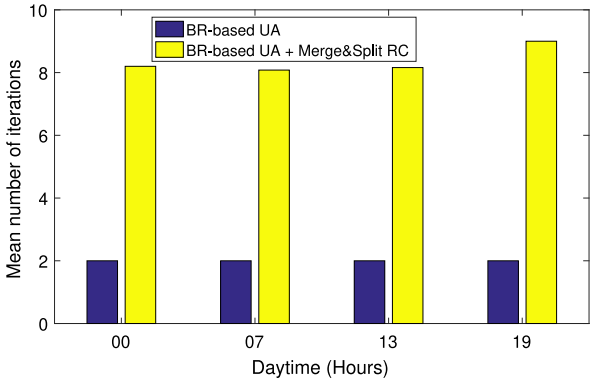


**Fig. 22.** Number of iterations of our iterative approach: Elastic sessions scenario.

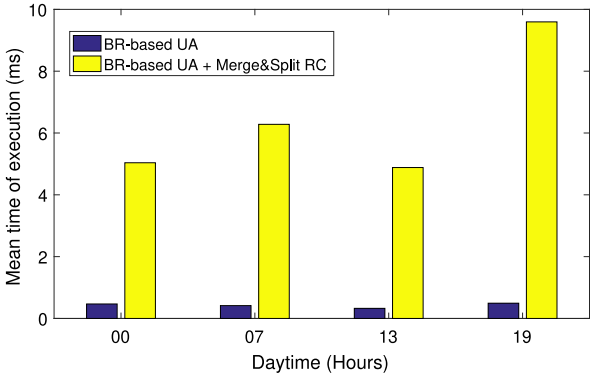


**Fig. 23.** Time of executions: Elastic sessions scenario.

### *8.3. BR vs optimal solution for UA sub-problem*

In this section, we investigate the performance of our distributed solution (best-response BR) for the UA sub-problem in presence of elastic traffic. To reveal the effectiveness of our proposed solution, we compare its results with the optimal solution for the UA sub-problem.

For users to form their strategy set using BR-based UA solution, each RRH can send on its beacon channel the identity of its BBU. Hence, any user can verify that the second best detected RRH $r''$ does not belong to the same BBU as the best detected RRH $r'$.

We consider a network topology with only 7 cells: a central RRH is surrounded by a ring of 6 immediately adjacent RRHs. Moreover,

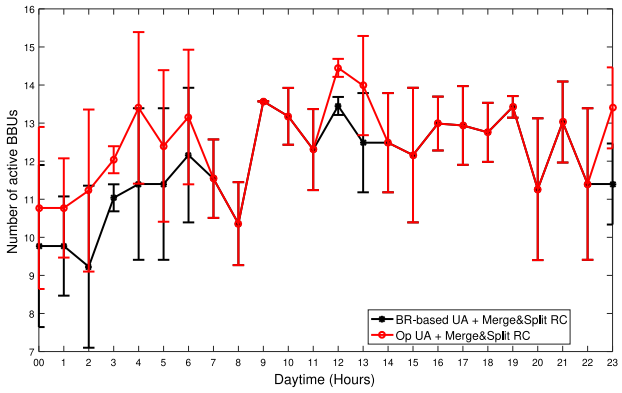


**Fig. 24.** Number of active BBUs: Elastic sessions scenario.

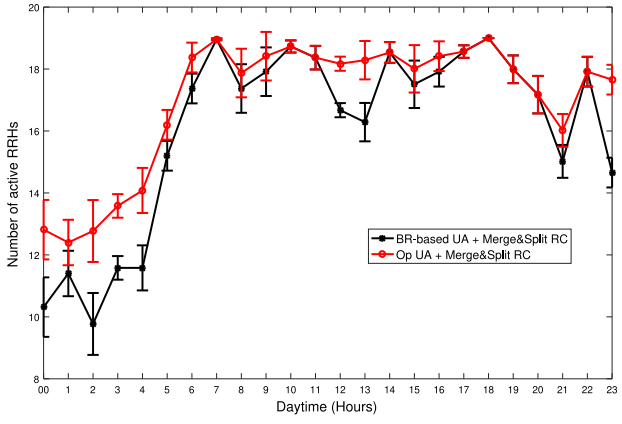


**Fig. 25.** Number of active RRHs as a function of the number of users.

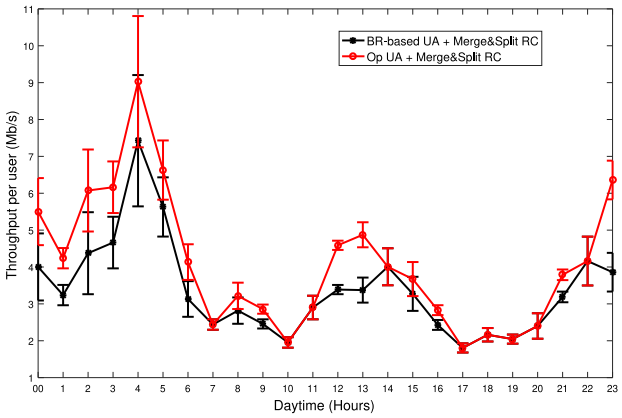


**Fig. 26.** Average user throughput: Elastic sessions scenario.

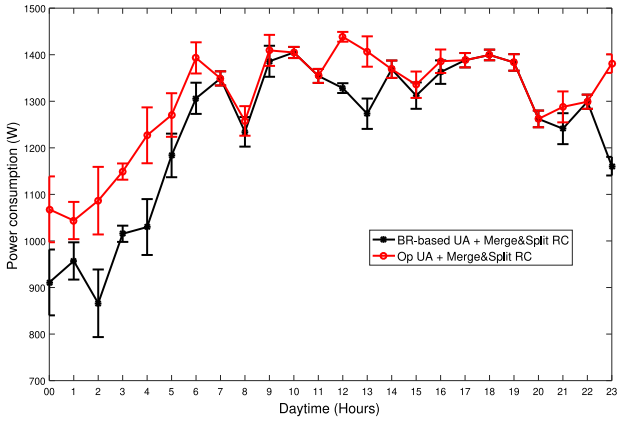


**Fig. 27.** Power consumption: Elastic sessions scenario.

to highlight the performance of our distributed approach for UA sub-problem, we used our approach based on the merge-and-split rules to solve the RC sub-problem for both BR and optimal UA sub-problem.

We respectively depict in Figs. 24 and 25 the number of active BBUs and RRHs during day time. When the users number is less than 70 users (cf. Fig. 14), the optimal UA sub-problem activates more BBUs and RRHs in comparison with the BR-based UA sub-problem. In fact, as each user $u$ seeks to minimize its own cost function described in (28), it chooses to be served by the BBU with the lowest number of users (*i.e.,* lowest load) and by the RRH belonging to this BBU that provides it with the best radio condition. Thus, the optimal UA sub-problem increases both the number of active BBUs and RRHs in comparison with the BR-based UA sub-problem. Besides, when the BR-based UA sub-problem is used, each user $u$ chooses only between two BBUs comprising RRHs that provide it with the strongest downlink signal. Moreover, when the users number increases in the network (above 70 users), more BBUs are activated to minimize its own cost function. As a result, both solution for UA sub-problem (*i.e.,* optimal and BR) activate the same number of BBUs and RRHs.

Figs. 26 and 27 respectively show the average user throughput and the C-RAN power consumption during day time. The optimal UA sub-problem activates more BBUs and RRHs in comparison with the BR-based UA sub-problem (cf. Figs. 24 and 25), leading to the highest user throughput (cf. 26). Nevertheless, this comes at the cost of higher power consumption, as shown in Fig. 27. As for BR-based UA sub-problem, it realizes much more power saving, mainly when the number of users is less than 70, while realizing user throughput close to that realized when the optimal UA sub-problem is used. Furthermore, by reducing the number of active BBUs and RRHs (cf. Figs. 24 and 25), the BR-based UA sub-problem consumes less power than the optimal UA. However, limiting the number of active BBUs provides the lowest user throughput as illustrated in Fig. 26.

As illustrated in Fig. 28, when the users number is greater than 70 users, the network utility achieved by our proposed algorithm (BR-based UA) is close to that when the optimal UA sub-problem is applied. In fact, the two solutions for UA sub-problem activate relatively the same number of BBUs and RRHs, leading to close user throughput and power consumption. However, when the users number decreases (less than 70 users), the gap between the two methods increases. In fact, our proposed algorithm (BR-based UA) realizes user throughput close to that realized when the optimal UA sub-problem is used with much more power saving (cf. Figs. 26 and 27).

## 9. Conclusion

In this paper, we have investigated the joint user association and RRH clustering problem in C-RAN. The aim is to maximize the network utility function incurred by both network power consumption and total user throughput for both streaming and elastic traffic. For a realistic simulation scenario, we utilize the Call Detail Records (CDR) provided by a real network operator. We formulated the problem as a mixed integer non linear optimization problem (MINLP). Since such a problem is NP-hard, we proposed to decouple it into two sub-problems: the user association (UA) sub-problem and the RRH clustering (RC) sub-problem. First, the UA sub-problem is modeled as an unweighted crowding. We proved that this game has a Finite Improvement Path (FIP) property and hence a best-response algorithm permits attaining the Pure Nash Equilibrium (PNE). Then, the RC sub-problem is modeled

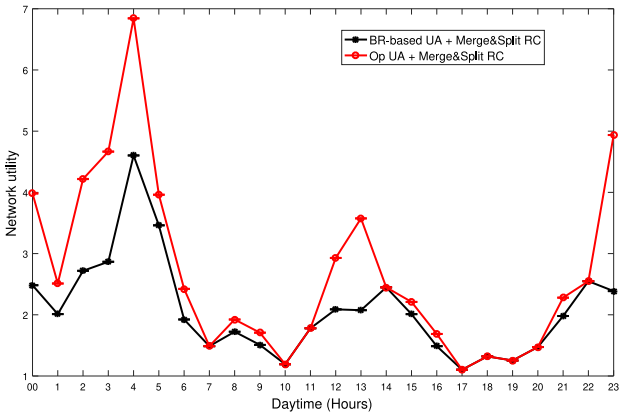


**Fig. 28.** Network cost as a function of the number of users.

as a cooperative game, where a heuristic solution, based on the merge-and-split rules, is applied to solve the RC sub-problem. These two sub-problems are sequentially and iteratively solved until convergence is reached. To show the effectiveness of our heuristic solution for the RC sub-problem in adapting to non-uniform traffic load, we compared it with the no-clustering and grand coalition solutions. Moreover, we highlighted the benefits of our approach for UA sub-problem by comparing its performance with the optimal UA sub-problem. We proved that, by applying our proposed solutions (*i.e.,* cooperative game theory for RC sub-problem and non-cooperative game theory for UA sub-problem), the network performance can be enhanced and adapts rapidly to the variation of traffic load during the day time.

## CRediT authorship contribution statement

**Hussein Taleb:** Conception and design of study, Acquisition of data, Analysis and/or interpretation of data, Writing - original draft, Writing - review & editing. **Kinda Khawam:** Conception and design of study, Analysis and/or interpretation of data, , Writing - review & editing. **Samer Lahoud:** Conception and design of study, Analysis and/or interpretation of data, , Writing - review & editing. **Melhem El Helou:** Conception and design of study, Analysis and/or interpretation of data, , Writing - review & editing. **Steven Martin:** Conception and design of study.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgment

All authors approved the version of the manuscript to be published.

## Appendix A. Proof of NP-hardness

In this section, we convert our optimization problem (19) to a maximum weight clique problem which has been proved to be NP-complete [26]. Let $\mathcal{A}$ be the set of all possible associations between users, RRHs, and BBUs to find the optimal solution (*i.e.,* $\mathcal{A} = \mathcal{U} \times \mathcal{R} \times \mathcal{Z}$). The power-set of $\mathcal{A}$, $\mathcal{P}(\mathcal{A})$, representing all possible associations between users, RRHs, and BBUs regardless of the constraints. Let $\mathcal{F}$ be the set of all feasible schedules satisfying the constraints, and it is obvious that $\mathcal{F} \in \mathcal{P}$. Let $\mathcal{S} = \{s_1, s_2, \dots, s_{|\mathcal{S}|}\} \in \mathcal{F}$ be any feasible association where $s_i \in \mathcal{A}, \forall i \leq |\mathcal{S}|$. Define $f : \mathcal{A} \rightarrow \mathbb{R}$ as a function mapping each individual association $s_i$ to the achievable utility. The original optimization problem (19) can then be reformulated as follows:

$$\text{maximize} \quad \sum_{i=1}^{|\mathcal{S}|} f(s_i) \tag{40}$$

$$\text{subject to} \quad \mathcal{S} \in \mathcal{F} \tag{41}$$

Consider the scheduling graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ where each vertex $v \in \mathcal{V}$ is an association between users, RRHs and BBUs, and the distinct vertices are connected by an edge in $\mathcal{E}$ if the constraints are satisfied. Denote by $\mathcal{C}$ the set of all possible cliques with degree $Z_{total}$. Then, the problem (40) can be written as a maximum weight clique problem in the following:

$$\begin{aligned} \mathbf{S}^* &= \underset{\mathbf{S} \in \mathcal{F}}{\text{argmax}} \quad \sum_{i=1}^{|\mathbf{S}|} f(s_i) \\ &= \underset{\mathbf{C} \in \mathcal{C}}{\text{argmax}} \quad \sum_{i=1}^{|\mathbf{C}|} w(v_i) \end{aligned} \tag{42}$$

where $\mathbf{C} = \{v_1, v_2, \dots, v_{|\mathbf{C}|}\} \in \mathcal{C}$ is a clique in the scheduling graph, and $w(v_i)$ is the weight of each vertex $v_i, \forall 1 \leq i \leq |\mathbf{C}|$. The optimal solution of the problem (40) is the maximum weight clique of degree $Z_{total}$ in the scheduling graph where the weight of each vertex $v_i \in \mathcal{V}$ is defined as the achievable utility of its corresponding association:

$$w(v_i) = f(s_i) \tag{43}$$

The maximum weight clique problem is known to be NP-complete [26], which means that the optimization problem of joint user association and RRH clustering (19) must be NP-hard, and cannot be solved optimally in polynomial time.

## Appendix B. Complexity analysis of optimal UA and RC sub-problems

In this section, we investigate the complexity of finding the optimal solution for UA and RC sub-problems. Given the RRH clustering (*i.e.,* RRH clustering variables $Y_{r,z}$ are known), finding the optimal user association can be determined by using an exhaustive search. Although decomposition of the joint problem can reduce the computational complexity to $O(R^U)$, it remains practically intractable for a large number of users. Moreover considering the user association decisions that have been taken by the UA sub-problem, the computational complexity to find the optimal RC solution, using exhaustive search, is reduced to $O(B_R)$ where $B_R$ is given by the $R$th Bell number and denotes the number of all RRH-BBU configurations should be explored. However, finding the optimal solution of RC sub-problem remains intractable for large $R$.

## Appendix C. Complexity analysis of the best-response algorithm

According to [23], a congestion game where players have only two strategies has the FIP and is hence isomorphic to a potential game. Further, it is well-known that the Best Response Algorithm converges in finite time to a pure NE in potential games [20]. Recently, the work in [27] proved that when players followed BR in turn (round robin), the mean number of iterations is $e^{\gamma} N + O(N)$ where $N$ is the number of players and $\gamma$ is the Euler constant.

## Appendix D. Complexity analysis of merge-and-split operations

The computational complexity is determined by the number of attempts for the merge-and-split operations. In the worst case scenario, the first RRH initiates $(R-1)$ merging attempts, the second initiates $(R-2)$ iterations, and so on. Consequently, the total number of merging attempts will be $(R(R-1)/2)$. Thus, the complexity of the merging process is in $O(R^2)$. In practice, the merging process requires a significantly lower number of attempts: once a cluster is formed, the merge

process stops. Furthermore, in the worst case scenario, the complexity of the splitting process is in $O(B_R)$. This process requires finding all possible partitions of $\mathcal{R}$. In practice, the splitting process is limited to the clusters that have already been formed. It is not performed over all the RRHs in $\mathcal{R}$. As the network utility takes into account the total network throughput, cluster sizes are usually kept small. As a result, the complexity of the splitting process is feasible in practical scenarios.

## References


[1] A. Checko, H.L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M.S. Berger, L. Dittmann, Cloud RAN for mobile networks a technology overview, IEEE Commun. Surv. Tutor. 17 (1) (2015) 405–426.

[2] M. Peng, Y. Sun, X. Li, Z. Mao, C. Wang, Recent advances in cloud radio access networks: System architectures, key techniques, and open issues, IEEE Commun. Surv. Tutor. 18 (3) (2016) 2282–2308.

[3] Small Cell Virtualization Functional Splits and Use Cases, Tech. Rep., Small Cell Forum, 2015.

[4] H. Taleb, Mobile user information, demand, and base station association for day time simulations, 2020, https://github.com/Hussein7082/Hussein-taleb.git.

[5] M. Peng, S. Yan, H.V. Poor, Ergodic capacity analysis of remote radio head associations in cloud radio access networks, IEEE Wirel. Commun. Lett. 3 (4) (2014) 365–368.

[6] A. Goldsmith, Wireless communications, 2005.

[7] J. Zuo, J. Zhang, C. Yuen, W. Jiang, W. Luo, Energy efficient user association for cloud radio access networks, IEEE Access 4 (2016) 2429–2438.

[8] Y. Xu, J. Zhang, S. Cai, Y. Cheng, H. Wang, User association algorithm of downlink C-RANs with edge caching and fronthaul compression, in: 2019 IEEE 19th International Conference on Communication Technology, ICCT, 2019, pp. 561–566.

[9] A. Salh, N.S.M. Shah, L. Audah, Q. Abdullah, W.A. Jabbar, M. Mohamad, Energy-efficient power allocation and joint user association in multiuser-downlink massive MIMO system, IEEE Access 8 (2020) 1314–1326.

[10] H. Hesham, R. Hesham, M. Ashour, Clustering of remote radio heads in C-RAN to minimize the number of base-band units, in: 2019 International Conference on Innovative Trends in Computer Engineering, ITCE, 2019, pp. 316–321.

[11] K. Boulos, M.E. Helou, K. Khawam, M. Ibrahim, S. Martin, H. Sawaya, RRH clustering in cloud radio access networks with re-association consideration, in: 2018 IEEE Wireless Communications and Networking Conference, WCNC, 2018, pp. 1–6.

[12] S. Namba, T. Warabino, S. Kaneko, BBU-RRH switching schemes for centralized RAN, in: 7th International Conference on Communications and Networking in China, 2012, pp. 762–766.

[13] H.M. Soliman, A. Leon-Garcia, QoS-aware joint RRH activation and clustering in cloud-RANs, in: 2016 IEEE Wireless Communications and Networking Conference, 2016, pp. 1–6.

[14] J. Yao, N. Ansari, Qos-aware joint BBU-RRH mapping and user association in cloud-RANs, IEEE Trans. Green Commun. Netw. 2 (4) (2018) 881–889.

[15] M. Mouawad, Z. Dziong, K. Addali, RRH selection and load balancing through dynamic BBU-RRH mapping in C-RAN, in: 2019 IEEE Canadian Conference of Electrical and Computer Engineering, CCECE, 2019, pp. 1–5.

[16] H. Taleb, M.E. Helou, K. Khawam, S. Lahoud, S. Martin, Joint user association and RRH clustering in cloud radio access networks, in: 2018 Tenth International Conference on Ubiquitous and Future Networks, ICUFN, 2018, pp. 376–381.

[17] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M.A. Imran, D. Sabella, M.J. Gonzalez, O. Blume, A. Fehske, How much energy is needed to run a wireless network? IEEE Wirel. Commun. 18 (5) (2011) 40–49.

[18] C. Liu, Y. Wan, L. Tian, Y. Zhou, J. Shi, Base station sleeping control with energy-stability tradeoff in centralized radio access networks, in: 2015 IEEE Global Communications Conference, GLOBECOM, 2015, pp. 1–6.

[19] S. Shenker, Fundamental design issues for the future Internet, IEEE J. Sel. Areas Commun. 13 (7) (1995) 1176–1188.

[20] D. Monderer, L. S. Shapley, Potential games, Games Econom. Behav. 14 (1996) 124–143.

[21] P. H. Young, Strategic Learning and Its Limits, 2004.

[22] R.W. Rosenthal, A class of games possessing pure-strategy Nash equilibria, Internat. J. Game Theory 2 (1973) 65–67.

[23] I. Milchtaich, Congestion games with player-specific payoff functions, Games Econom. Behav. 13 (1) (1996) 111–124, https://ideas.repec.org/a/eee/gamebe/v13y1996i1p111-124.html.

[24] K.R. Apt, T. Radzik, Stable partitions in coalitional games, 2006, arXiv:cs/060532v1.

[25] Broadband Services, Applications, and Networks: Enabling Technologies and Business Models: Comprehensive Report, International Engineering Consortium, 2004.

[26] R. Karp, Reducibility among combinatorial problems, Complexity Comput. Comput. 40 (1972) 85–103.

[27] S. Durand, B. Gaujal, Complexity and Optimality of the Best Response Algorithm in Random Potential Games, Vol. 9928, 2016, pp. 40–51.


**Hussein Taleb** received a diploma in Telecommunication Engineering from Islamic University of Lebanon in 2008 and a master's degree in Telecom Networks and Security from Saint Joseph University of Beirut in 2016. He started the Ph.D. in Communication network from CIMTI Research Center, University of Saint Joseph University of Beirut, in 2017. His research thesis focusing on 5G Communications.

**Kinda Khawam** got her engineering degree from Ecole Supérieure des Ingénieurs de Beyrouth (ESIB) in 2002, the Master's degree in computer networks from Telecom ParisTech (ENST), Paris, France, in 2003, and the Ph.D. from the same school in 2006. She was a post doctoral fellow researcher in France Telecom, Issy-Les-Moulineau, France in 2007. Actually, she is an associate professor and researcher at the University of Versailles in France. Her research interests include radio resource management, modeling and performance evaluation of mobile networks. She has co-authored more than 60 papers published in international journals and conference proceedings.

**Samer Lahoud** received the Ph.D. degree in communication networks from IMT Atlantique, Rennes, in 2006. After his Ph.D. degree, he spent one year at Nokia Bell Labs Europe. From 2007 to 2016, he was with the University of Rennes 1 and with IRISA Rennes as an Associate Professor. He is currently an Associate Professor with the Saint Joseph University of Beirut, where he lectures computer networking courses with the Faculty of Engineering, Ecole Supérieure dIngénieurs de Beyrouth (ESIB). His research activities focus on routing and resource allocation algorithms for wired and wireless communication networks.

**Melhem El Helou** received the engineer's and master's degrees in telecommunications and networking engineering from the Ecole Supérieure d'Ingénieurs de Beyrouth (ESIB), Faculty of Engineering, Saint Joseph University, Beirut, Lebanon, in 2009 and 2010, respectively, and the Ph.D. degree in computer and telecommunications engineering from IRISA Research Institute, University of Rennes 1, Rennes, France, and Saint Joseph University, in 2014. In 2013, he joined ESIB, where he is currently an Assistant Professor (fr: Maître de conférences). His current research interests include wireless networking, Internet of Things, and quality of service.

**Steven Martin** received his Ph.D. degree from INRIA, France, in 2004. Since 2005, Steven MAR- TIN is working at Paris-Sud University. Leading the research group "Networking and Optimization" at LRI (the Laboratory for Computer Science at Paris-Sud University, joint with CNRS), his research interests include quality of service, wireless networks, network coding, ad hoc networks and real-time scheduling. He is involved in many research projects and had the lead of the activity "Personal safety in digital cities of the future" in EIT ICT Labs (the European Institute of Innovation and Technology) from 2010 to 2014. He is the author of a large number of papers published in leading conference proceedings and journals. He has served as TPC member for many international conferences in networking.