Optimized load balancing by dynamic BBU-RRH mapping in C-RAN architecture

Ermínio Augusto Ramos da Paixão, Rafael Fogarolli Vieira, Welton Vasconcelos Araújo, and Diego Lisboa Cardoso
Operational Research Laboratory
Federal University of Pará
Belem, Brazil
{hermespaixao,fogarollirafael,weltonmax007}@gmail.com
diego@ufpa.br

***Abstract*—Cloud Radio Access Network (C-RAN) is one of the key architectures for the next generation of mobile networks (5G) that aim at centralized processing and management, collaborative radios and cloud computing in real time. These features enable the architectures to make a rational adjustment to the connection between remote radio heads (RRHs) and baseband units (BBUs) dynamically. This is important since if this feature is neglected, it can cause difficulties such as blocked calls and low-quality connections. This study investigates this area and proposes an optimized mapping model for RRH-BBU that seeks a more equitable and effective balancing. The Key Performance Indicator (KPI) of blocked calls was used for this to measure the quality of service (QoS). A particle Swarm algorithm (PSO) was created to minimize the number of blocked calls and additionally balancing the processing load between the BBUs. Scenario from literature was employed that consists of 19 RRHs distributed in a geographic area, which can be mapped in a BBU pool that manages two BBUs with three sectors each. The initial configuration on average, led to 80 blocked calls. The results obtained by the PSO show that there was a reduction of up to 100% of blocked calls, as well as a more equitable load distribution between the BBUs. Additionally, realistic scenarios with different user profiles were also included, since they demonstrate that these factors have a direct impact on the load generated in the BBUs and hence, affect their balance.**

***Index Terms*—C-RAN, RRHs, BBUs, PSO, balance**

I. Introduction

According to Cisco Systems, the real impact of 5G technology on the growth of mobile technology will begin in 2020. This is one of the conclusions reached in the Visual Networking Index (VNI) report about global mobile networks traffic in the period 2016-2021. In its 11th edition, it stated that the extraordinary growth of mobile phone users, smartphones, use of Internet of Things (IoT) and energy consumption of mobile videos is expected to multiply mobile traffic data by seven in the next five years [1].

This exponential growth of data traffic is a determining factor for research which must find ways to meet this demand, by recommending improvements with regard to the availability of data, the performance of the network and QoS. The new 5G architecture has arisen as a means of meeting the demand for greater capacity and addressing the new needs of the users, since the 4G architecture is unable to support the growth in demand.

One of the new technologies that has emerged in this new architecture is C-RAN, which is designed for centralized processing, collaborative radio and cloud computing [2].

In the C-RAN model, shown in figure 1, the RRHs are linked to a BBU pool through an interconnection protocol, such as the Common Public Radio Interface (CPRI) [3]. The BBUs in the BBU pool are interconnected and connected to a server called the Host Manager, that verifies the load in each BBU and is responsible for providing a suitable configuration of BBU-RRH. Each BBU can work with the N sector and each sector contains multiple RRHs connected to it. The RRHs only belong to one sector in a given period of time.

The load in each BBU is determined by its total number of active users within the BBU sector. However, owing to both hardware and software restrictions, the number of active users in each sector is limited this problem is known as hard capacity (HC). An appropriate way of solving this is through the Self Organizing Network (SON), which will adjust the mapping and distribution of resources in the network [4].

In [5], a genetic algorithm was created to map and allocate resources for a Self-organised Cloud Radio Access Network, where the performance measure of the blocked Call Center KPI is defined; this measurement is used to evaluate the QoS of the scheme. Acceptable results were obtained by maximizing the RRH-BBU mapping process, while at the same time, causing a load imbalance in the BBUs.

The study in [4] describes the arrival of the new architectures of cellular networks and their challenges. A scheme was put forward to introduce network function virtualization (NFV) with software- defined networking (SDN) to carry out self-organizing planning for future networks with the aim of providing energy to improve efficiency and data loading methods in different scenarios to optimize user experience.

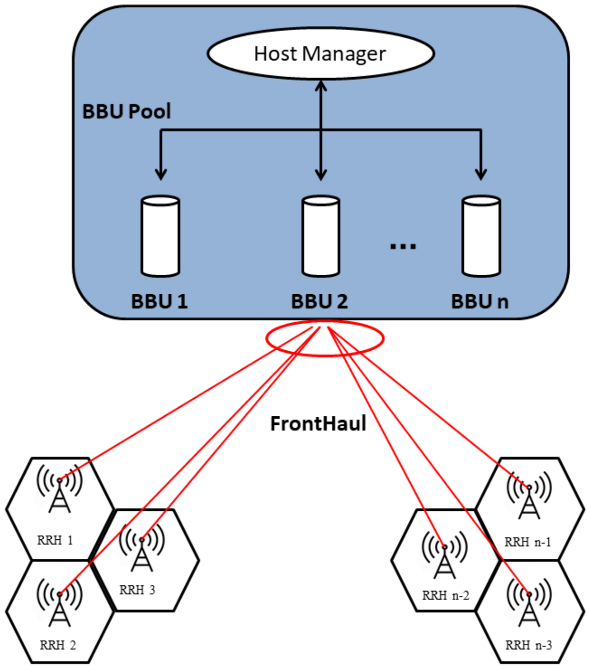


Fig. 1. C-RAN (Cloud Radio Access Network) architecture.

After the model has been implemented, a more cost-effective and malleable scenario can be achieved.

Finally, in [6] an optimal load-balancing model was created for C-RAN cellular systems, using an infinite-dimensional optimization method to obtain a better distribution of resources. The results show that this recommended technique allocated resources effectively without adding further costs to the operational sector.

In light of these studies, it is clear that there are still several gaps to be filled within the framework of the C-RAN architecture, especially with regard to load balancing in Self- Organized Cloud Radio Access networks. Thus, this work seeks to make an improvement to the balancing of resources between the sectors of more than one BBU, while also enabling load balancing in the BBUs.This can be a means of filling a gap in the literature, by reducing the number of blocked calls and, at the same time, balancing the load between the BBUs.

The rest of the article is structured as follows. In Section 2, there is an explanation of the BBU-RRH configuration and mathematical formulation. Section 3 outlines the allocation algorithm and there is a discussion about its optimization process. Section 4 displays and analyzes the results, Section 5 brings the study to a conclusion and and makes suggestions for future work.

II. BBU-RRH Mapping and formulation

As discussed in [5] a KPI (defined as the number of blocked calls) was used to represent C-RAN QoS. On the basis of this KPI, the host manager determines the appropriate settings for the BBU-RRH mapping balancing the resources of the BBUs.

Considering the same scenario, 19 RRHs were distributed randomly and managed by a BBU Pool, containing two BBUs with three sectors each. An imbalance is sought in the distribution of users among the available sectors of each BBU in an attempt to reduce the number of blocked calls. An imbalance can cause overhead in certain sectors and BBUs and, hence lead to blocked calls.

The objective function set out in [5] was changed, since there was also a need to distribute the number of users in accordance with the number of sectors available in the N BBUs (thus, balancing the network). To achieve this, a vector of users (RRHs load) is initially created by sector, as expressed in Equation 1.

$$U_s = \sum_{j=1}^{N} C_j {R_s}^{i+1}, S = 1, 2, ..., K \tag{1}$$

Where:
U_s is the user number in the sector;
N is the total RRHs;
K is the total sectors;
C is the amount of users connected in the RRH j;
R is a binary variable where it assumes the value 1 if RRH j is allocated in the sector s.

The U_s vectors is used in the objective function, which will vary according to the number of users in the network to ensure that the sectors will be balanced, and hence their respective BBUs as well.

All possible Us for all the K (sectors) will be tested to obtain the lowest possible value of (KPI_{min}). This involves reducing the number of blocked calls, and thus maximizing the QoS, as demonstrated in Equation 2.

$$KPI_{min} = \sum_{s=1}^{k} (U_s - HC), \tag{2}$$

$$\{0, if(U_s - HC) < 0\} or \{(U_s - HC) if (U_s - HC) >= 0\}.$$

The exit from the model suggested is given by a vector $S_j^{i+1} = \{S_1^{i+1}, S_2^{i+1}, ..., S_N^{i+1}\}$, which represents the sectors of the BBUs S_j^{i+1} and RRHs that were allocated to these sectors.

III. Proposed Allocation algorithm

The objective function suggested was implemented in a PSO. It was chosen because it is an evolutionary algorithm, that uses particles with a possible candidate solution matrix to solve optimization problems [7].

The PSO uses a Swarm of initial particles in which each of them corresponds to a candidate solution. These determine the area of the solution randomly having a different speed and position, after this process the members of the Swarm inform good positions among themselves and update their position and speed in areas defined as local best $Pbest$ and global Best $Gbest$, Directing the particles to their best aptitude values. The acceleration factor $Pbest$ and $Gbest$ are changed in a stochastic way in each iteration and its update stems from the experience of the best historical position of its particle and the best experience of neighboring particle position, that is, the best global best $Gbest$.

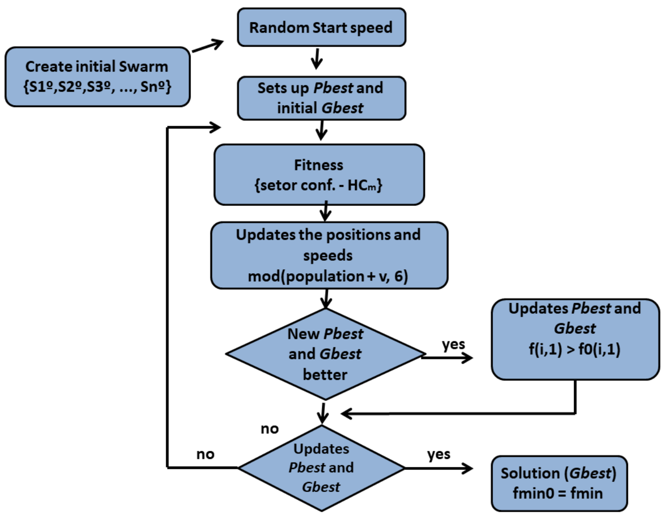


Fig. 2. Procedures of the PSO.

As result, the best particles is represented by a vector result $\{S_1^{i+1}, S_2^{i+1}, ..., S_N^{i+1}\}$ where each particle characterizes a possible combination of RRHs distributed in the sectors of BBUs. The parameters applied in the simulation are shown in Table 1.

TABLE I
PARAMETERS

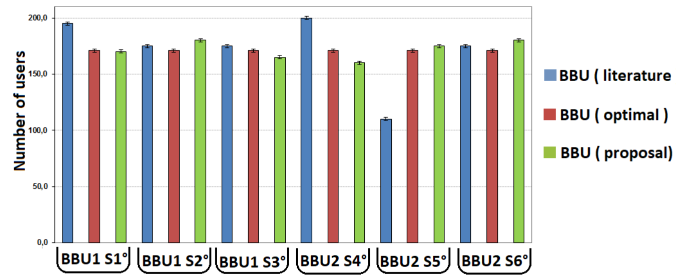
Parameter	Values
Local acceleration factor	1.8
Global acceleration factor	1.8
Population size	220
Limit number of interactions	100

IV. RESULTS

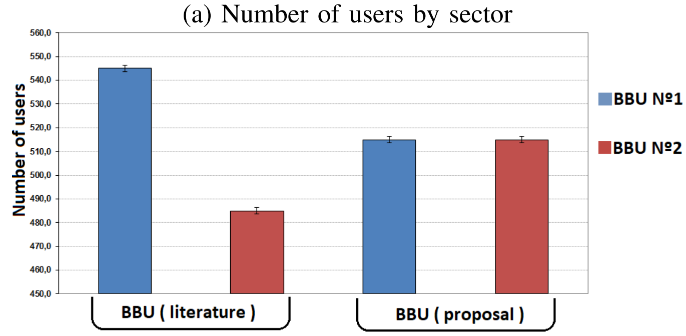
As stated earlier, for comparative purposes, the same scenario proposed in [5], was used. This consisted of 19 RRHs distributed randomly in one geographic area, that are managed by a BBU pool (with two BBUs that have three sectors each) and HC= 200 is stipulated for them. Additionally, it was implemented another approach called optimal, that simply divides the users between the BBU sectors without taking account of their distributions in the RRHs.

A. Homogeneous UEs Results

The results of these three approaches can be observed in figure 3(a), it should be noted that the method used in the literature provides sectors without blocked calls, but with a certain disparity in the allocation of their users. Moreover, the proposed model obtained no blocked calls and with sectors that are more equally distributed. In Diagram 3(b) it can be confirmed that the model used in the literature uses 90% of the total capacity of the first BBU and 81% of the charge capacity of the second BBU, while the new form, only uses 85% in each BBU, which is evidence of, a better rate of efficiency in balancing the load of the BBUs.



(a) Number of users by sector



(b) Number of users by BBU

Fig. 3. BBUs balancing

Table 2 shows the measurements of both works [5], where the balance between the sectors can be observed, with regard to the capacity restrictions, although without any concern about the sectoral distribution of loads, which in high load conditions, can compromise the performance of the BBUs. The new scheme has proved to be efficient and can provide a fairer balance, since, the sectors have more equitable averages.

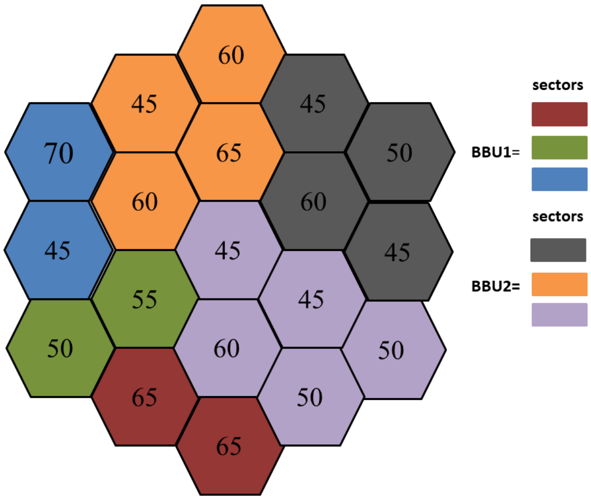
TABLE II
AVERAGE NUMBER OF USERS

Measurements (amt users)	S1	S2	S3	S4	S5	S6
Medium (lit.)	195	175	175	199	110	174
Std deviations (lit.)	0.96	0.65	0.74	0.64	0.65	0.45
Average (proposal)	170	180	164	160	174	180
Std deviation (proposal)	0.40	0.36	0.41	0.57	0.70	0.67

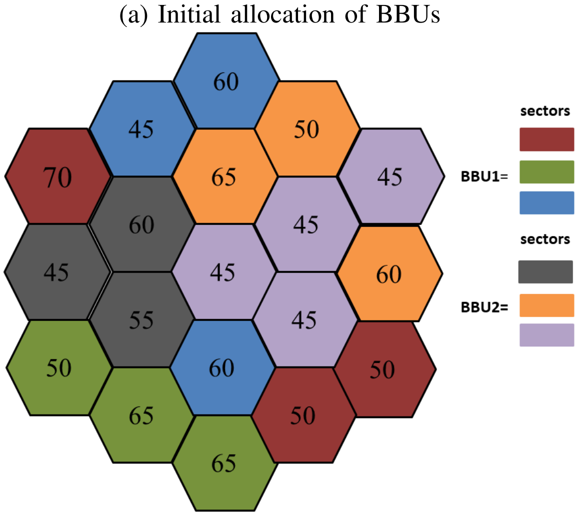
The distribution of the BBU sectors can be observed in figures 4(a) and (b). In the figure 4(a) the sectors are distributed randomly, which can cause a possible imbalance. After the execution of the algorithm, its final allocation can be noted in figure 4 (b). can be noted. As well as being able to mitigate the problem of blocked calls, it can have a more equal load distribution in the BBU sectors.

B. Heterogeneous UEs Results

Users have traditionally formed different profiles of their applications, especially when they are at work or at home [8], which means that it is unrealistic to treat all of them in a uniform way. In light of this, an additional restriction

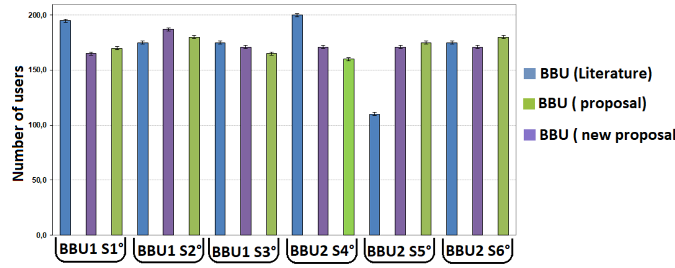


(a) Initial allocation of BBUs

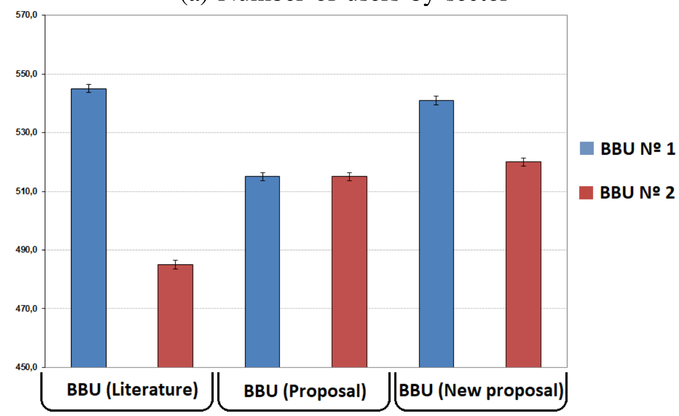


(b) Final allocation of BBUs

Fig. 4. 19 RRHs governed by 2 BBUs in the BBU pool, each BBU deals with 3 sectors



(a) Number of users by sector



(b) Number of users by BBU

Fig. 5. BBUs balancing with user differentiation

was proposed where the RRHs of the proposed scenario are divided into commercial and residential.30% of the antennas were considered commercial and the remaining, 70%, were defined as residential [6]. The demands generated by the RRHs installed in commercial areas are 10% higher than those in the residential scenarios.

Figure 5 (a) shows the results obtained for this new scenario. Although it involves a higher load, the objective function achieved the balance within the imposed restrictions, and showed better results than the approach adopted in the literature. It should be noted that in figure 1 the difference between the BBU loads is lower than in the literature approach, showing the effectiveness of this proposal.

V. Conclusion

The C-RAN architecture and its capacity to reconfigure the mapping of resource allocation between BBU-RRH is still an open problem. Several studies in the literature have already attempted to overcome this problem, by seeking to mitigate the blocked calls and obtain a fairer and more efficient balance. This study has set out a new system for modeling the problem which was implemented through a particle Swarm optimization algorithm (PSO).

The PSO can converge quickly and with fewer iterations than the algorithm recommended in the literature. As well as optimizing the QoS, it can achieve a fair balance in the BBU-RRH sectors. Through the results it was observed a better performance than presented in [8], even in different load scenarios.

With regard to possible future works, there is a need to exploit and analyze other performances measures, that can have an effect on load balance in networks. As well as this, an attempt should be made to make use of propagation models to determine the allocation between the UEs-RRHs, and thus make the scenarios more realistic. Additionally, it would be worth demonstrating the improvements proposed in simulated scenarios (i.e. using simulators of discrete event simulations), where other performances indicators can be measured and evaluated, such as the behavior of applications for the end user.

References

[1] Cisco, Worldwide mobile data traffic will increase 7 times between 2016 and 2021 , https://www.cisco.com/c/ptpt/about/press/news-archive-2017/20170208.html, November, 2017.

[2] Ericsson, Ericssons fiber fronthaul solution deployed for China Mobiles LTE C-RAN, http://www.ericsson.com/news/140707-ericssons-fiber-fronthaul-solutiondeployed244099436c, November, 2017.

[3] C. Specification, V4. 2(2010-09-29). Ericsson AB, Huawei Technol. Co Ltd, NEC Corp., 2009.
[4] Chen, Na et al. Self-organizing scheme based on NFV and SDN architecture for future heterogeneous networks. Mobile Networks and Applications, v. 20, n. 4, p. 466-472, 2015.
[5] Khan, M.; Alhumaima, R. S.; Al-Raweshidy, H. S. Quality of service aware dynamic BBU-RRH mapping in cloud radio access network. In: Emerging Technologies (ICET), 2015 International Conference on. IEEE, 2015. p. 1-5.
[6] Ran, Chen; Wang, Shaowei; Wang, Chonggang. Optimal load balancing in cloud radio access networks. In: Wireless Communications and Networking Conference (WCNC), 2015 IEEE. IEEE, 2015. p. 1006-1011.
[7] Couceiro, Micael; Ghamisi, Pedram. Particle Swarm optimization. In: Fractional Order Darwinian Particle Swarm Optimization. Springer International Publishing, 2016. p. 1-10.
[8] Holm, Henrik et al. Optimal assignment of cells in C-RAN deployments with multiple BBU pools. In: Networks and Communications (EuCNC), 2015 European Conference on. IEEE, 2015. p. 205-209.