QoS-Aware Dynamic RRH Allocation in a Self-Optimised Cloud Radio Access Network with RRH Proximity Constraint

M.Khan, *Student Member*, IEEE, R.S. Alhumaima, H.S. Al-Raweshidy, *Senior Member*, IEEE

***Abstract*—An inefficient utilisation of network resources in a time-varying traffic environment often leads to load imbalances, high call-blocking events and degraded Quality of Service (QoS). This paper optimises the QoS of a Cloud Radio Access Network (C-RAN) by investigating load balancing solutions. The dynamic re-mapping ability of C-RAN is exploited to configure the Remote Radio Heads (RRHs) to proper Base Band Unit (BBU) sectors in a time-varying traffic environment. RRH-sector configuration redistributes the network capacity over a given geographical area. A Self-Optimised Cloud Radio Access Network (SOCRAN) is considered to enhance the network QoS by traffic load balancing with minimum possible handovers in the network. QoS is formulated as an optimisation problem by defining it as a weighted combination of new key performance indicators (KPIs) for the number of blocked users and handovers in the network subject to RRH sectorisation constraint. A Genetic Algorithm (GA) and Discrete Particle Swarm Optimisation (DPSO) are proposed as evolutionary algorithms to solve the optimisation problem. Computational results based on three benchmark problems demonstrate that GA and DPSO deliver optimum performance for small networks, whereas close-optimum is delivered for large networks. The results of both GA and DPSO are compared to Exhaustive Search (ES) and K-mean clustering algorithms. The percentage of blocked users in a medium sized network scenario is reduced from 10.523% to 0.421% and 0.409% by GA and DPSO, respectively. Also in a vast network scenario, the blocked users are reduced from 5.394% to 0.611% and 0.56% by GA and DPSO, respectively. The DPSO outperforms GA regarding execution, convergence, complexity, and achieving higher levels of QoS with fewer iterations to minimise both handovers and blocked users. Furthermore, a trade-off between two critical parameters for the SOCRAN algorithm is presented, to achieve performance benefits based on the type of hardware utilised for C-RAN.**

***Index Terms* - Base Band Unit (BBU), Cloud Radio Access Network (C-RAN), Discrete Particle Swarm Optimisation (DPSO), Genetic Algorithm (GA), Remote Radio Head (RRH), Self-Optimising Network (SON)**

I. INTRODUCTION

THE up-surging volume of data services and applications along with the accelerated growth in wireless access demands has posed significant challenges for the Next Generation of Mobile Networks (5G). According to [1], the amount of IP data driven by wireless networks is predicted to surpass 500 exabytes by 2020. Mobile Network Operators (MNOs) are facing significant challenges to maintain the performance and availability of their network with high levels of QoS which signals the dawn of a 5G era. Densifying the access networks using small cells is realised as a promising solution to increase capacity and coverage, especially at traffic *hot-spots*. However, this leads to even bigger challenges for the MNOs such as the significant increase in Capital (CAPEX) and Operational (OPEX) expenditures, inefficient utilisation of network resources due to traffic imbalances and increased signalling overhead caused by frequent handovers among small cells. Therefore, MNOs are required to devise innovative solutions beyond the bounds of conventional performance upgrades to achieve optimum returns on investment and maintaining high levels of QoS.

Network performance is highly degraded due to inefficient utilisation of resources and fail to produce maximum returns on expenditure if they are underutilised or remain idle. Network resources are often under-utilised during unbalanced traffic loads situations, particularly when some network cells may suffer from heavy loads causing a high number of blocked users, while others remain lightly loaded with their resources underutilised. Therefore, it is crucial to achieving self-optimisation in the network on varying traffic environment, especially when the load distribution among cells is not uniform. Inter-cell optimisation is a critical optimisation problem in Self Organising Networks (SON) for the Third Generation Partnership Project (3GPP) [2], [3]. Furthermore, SON contributes to managing complexity and enhancing network performance by minimising network-cost via simplified operational tasks and autonomous configuration or management functionalities such as self-healing, self-configuration, and self-optimisation [4].

Numerous studies and methods on self-optimisation have suggested addressing the problem of load balancing in cellular networks via SON. The aim is to autonomously adjust operation parameters when a traffic imbalance is detected among cells. As a result, users connected to BSs with high load cells are handed over to ones with under-loaded cells thereby achieving high capacity, enhanced throughput, and a balanced network load. This is accomplished by adjusting the operation parameters such as antenna angle (Antenna tilt) [5] and/or handover parameters [6] to reduce the coverage area so as to achieve Mobility Load Balancing (MLB) [7]. In MLB, the handover thresholds are adjusted following traffic conditions which result in expansion or contraction of virtual transfer areas among adjacent cells and thereby reducing or increasing users in the cells. However, incorrect adjustment of handover parameter settings can cause unnecessary transfers in the network which often leads to handover ping-pongs/delays and radio link failures. Mobility Robustness Optimisation (MRO) [8] is a SON function which aims to eliminate link failures and reduce unnecessary handovers caused by incorrect handover parameters. Power adaptation for load balancing is another technique to effectively change the cell coverage area which in return changes the association of all users in the coverage area. In LTE, Cell Range Expansion (CRE) [9] is a technique which allows Low Power Nodes (LPN) to expand their coverage area and take in users from the Macro Cell. Usually, users associate to the cell which provides the strongest signal. However, in CRE users connect to the LPNs despite receiving the strongest signal from the Macro cell. [10] provides a comprehensive survey on self-organisation in future cellular networks, which includes a detailed description of the schemes mentioned above along with hybrid approaches and other existing SON load balancing methods in the literature.

Moreover, SON architectures can be divided into three types - a) centralised, b) decentralised and, c) hybrid. In the decentralised and hybrid SON architectures, the SON algorithm partially runs on the network management level and partially in the network elements. Coordination of different SON functions, possibly having conflicting goals and operating on various time scales, is more challenging than in the centralised architecture [11]. In the centralised structure, a central Network Management System (NMS) or a SON server decides the network optimisation algorithms and the eNodeB parameter configuration [12]. The centralised SON architecture is more manageable regarding the implementation of SON Algorithms compared to distributed and Hybrid SON architectures. It enables the SON algorithms to jointly optimise multiple network parameters, therefore, allowing a globally tuned system. However, the centralised SON server in this approach requires strict latency and delay requirements regarding system KPIs and UE measurements for SON parameter updates, which restricts the applicability of a purely centralised SON architecture. Increase response time limits the network to adapt to changes and may cause instabilities.

However, Cloud Radio Access Network (C-RAN) is a novel paradigm which has the potential to resolve many challenges that the MNOs are experiencing today. According to [13], both operators and equipment vendors have proposed a C-RAN that possess a power efficient centralised processing infrastructure with real-time cloud computing and collaborative radio features. C-RAN aggregates the BBUs of typical base stations (BSs) to a centralised location called base band unit pool (BBU pool). The RRHs with simpler functions are left off on the cell sites and can be deployed densely with minimum cost. The RRHs collects radio signals from geographically distributed antennas and transmits them to the centralised BBU pool via an optical transmission network (OTN). A single BBU can serve multiple RRHs, and the distance between BBU and RRH is limited to 40 km due to propagation and processing delays [14]. C-RAN relieves the base stations (BSs) from maintaining 24/7 services by aggregating the BBUs in a remote data centre/BBU pool. Significant resource utilisation and power savings can be achieved by dynamically remapping the BBU-RRH configuration. C-RAN requires fewer BBUs to serve a geographical area compared to traditional RAN and saves operational and management cost to a great extent.

The main contribution of this paper is to present an efficient model for proper BBU-RRH mapping in C-RAN as one SON approach to achieve a self-optimising network structure and solving a load balancing problem. The self-optimising feature of SON combined with the capacity routing ability of C-RAN is explored to achieve a balanced system load with high levels of QoS. Network capacity is dynamically redistributed over a geographical area with respect to time-varying traffic. The BBU-RRH logical connections are adjusted by proper RRH assignment to BBU sectors via an intelligent algorithm. RRH-sector allocation is formulated as a linear integer-based optimisation problem with constraints. Two evolutionary algorithms, i.e., GA and DPSO, are considered to solve the optimisation problem. This paper presents not only load balancing in C-RAN but also the realisation of virtual small cells (supported by low-power RRHs), rather than Micro and Pico cells deployment at each antenna position. The cell split deployment scenario discussed in [15] is considered for virtual small cells. Note that, a BBU sector may change size (on time-varying traffic) based on the number of RRHs clustered together to support that sector.

The rest of the paper is organised as follows: Section II presents a survey of related work. Section III presents the self-organising C-RAN framework and the proposed system model. Section IV illustrates the formulation for dynamic RRH-sector allocation problem. Section V represent RRH clustering as a constraint for the optimisation problem. Section VI defines the SOCRAN algorithm. Computational results and complexity comparison of different algorithms are discussed in Section VII. Finally, the paper is concluded in Section VIII.

II. Related work

The MNOs together with academia have jointly initiated many experimental projects to explore the potential benefits of C-RAN. Next Generation Mobile Networks (NGMN) alliances project P-CRAN [16], European Commission's MCN [17], FP7-based projects such as HARP [18], iJOIN [19], and CROWD [20] are some of the major design and implementation initiatives for C-RAN. [21] summarises the ongoing work in C-RAN and examples of first field trials and prototypes along with innovative end-to-end solutions for practically implementable C-RANs. Moreover, [14], [22], [23] provides a comprehensive survey on C-RAN and highlights the challenges, advantages, and implementation issues regarding different deployment scenarios. Also, an in-depth review of the principles, technologies and applications of C-RAN describing innovative concepts regarding physical layer, resource allocation, and network challenges together with their potential solutions are highlighted in [24], [25].

Most of the existing research on C-RAN focuses on mapping between User Equipment (UE) and the RRH, whereas limited work on BBU-RRH mapping is addressed. Some recent studies on the connection between UE and RRH are described in [26]–[28]. In [26], the authors attempt to solve a joint RRH and precoding optimisation problem which aims to minimise network power consumption in a MIMO based user-centric C-RAN. In line with this work, the authors of [27] propose a weighted minimum mean square error (WMMSE) approach to solving the network-wide beamforming vector optimisation problem for RRH-UE clusters formation. The BBU scheduling is then formulated as a bin packing problem for energy efficient BBU utilisation in a heterogeneous C-RAN environment. The authors of [28] propose a cross-layer framework that jointly optimises physical and network layer resources to improve throughput performance. Also, RRHs beamforming vectors, user RRH association, and network coding based routing are optimised in an overall design.

Studies regarding BBU pooling in C-RAN are discussed in [29]–[33]. BBU utilisation significantly affects network throughput and transmission efficiency in C-RAN. Therefore, BBU behaviour is necessary to consider in a resource management design. In [29], a dynamic BBU-RRH mapping scheme is proposed using a borrow-and-lend approach in C-RAN. Overloaded BBUs switch their supported RRHs to underutilised BBUs for a balanced network load and enhanced throughput. The authors in [30] initially proposed semi-static and adaptive switching schemes to adjust BBU-RRH configuration based on peak hour traffic loads for all RRHs within a given time interval. Minimum possible BBUs are allocated to RRHs based on traffic load. The work of [31] then proposed a lightweight, scalable framework that utilises optimal transmission strategies via BBU-RRH reconfiguration to cater dynamic user traffic profiles. The work of [32] studies traffic adaptation and energy saving potential of TDD-based heterogeneous C-RAN by adjusting the logical connections between BBUs and RRHs. The authors of [33] recently investigated an RRH clustering design and proposed a spectrum allocation genetic algorithm (SAGA) to improve network QoS via efficient resource utilisation.

Regarding other related work, research initiatives are taken to develop Network Function Virtualisation (NFV) and Software Defined Network (SDN) solutions for C-RAN [34]–[36]. NFV is an architectural framework that provides a virtualised network infrastructure, functions and NFV orchestrator for control and management [37]. However, SDN is a concept related to NFV. SDN decouples data and control plane to enable directly programmable control plane while abstracting underlying physical infrastructure from applications and services [38]. Although SDN and NFV are not the prime focus of this paper, they are presented in this section for completeness of the C-RAN introduction and are important concepts that can help to implement virtualisation of baseband resources.

To sum up, the existing resource allocation mechanisms in C-RAN does not take full advantage of the concept of centralised BBU pool. In this paper the scope of C-RAN is further extended by developing a dynamic BBURRH mapping scheme in C-RAN considering 'blocking probability triggered load balancing' which has never been considered for C-RAN or LTE before. The primary objective is to enhance network QoS and to decrease the blocked users, especially when the user distribution is not uniform. Note that, load balancing schemes can be divided into two categories, i.e., 'blocking probability triggered load balancing' [39] and 'utility aware load balancing' [40]. Blocking probability based schemes decrease the blocking probability in the network regardless of proportional fairness among users, whereas utility based schemes serve users in a fair manner while keeping the system throughput balanced. However, the main complication with utility based schemes is the tendency to achieve a global network proportional fairness. This requires access to information of every individual user in the network which makes these schemes complicated and impractical for large networks. Load in utility based schemes is defined as a function of network resources (PRBs), whereas blocking probability based schemes defines load as a function of the number of connected users.

III. Self-Optimising Cloud Radio Access Network Framework

This paper proposes a self-organising framework that is applicable for short and long term dynamics of C-RAN. The framework maximises the overall QoS of the system while considering a network load balance. The framework is based on self-optimisation to exploit the benefits of capacity routing in C-RAN. It maximises QoS levels regarding desired KPIs. Note that, several performance indicators (KPIs) can be considered to measure the network QoS, so the framework is modelled as a general multi-objective optimisation problem including several criteria. Many other criteria may be included to tailor several other optimisation objectives subject to specific operator policy requirements. However, the scope of this paper focuses on QoS evaluation based on KPIs relevant for blocking probability triggered load balancing.

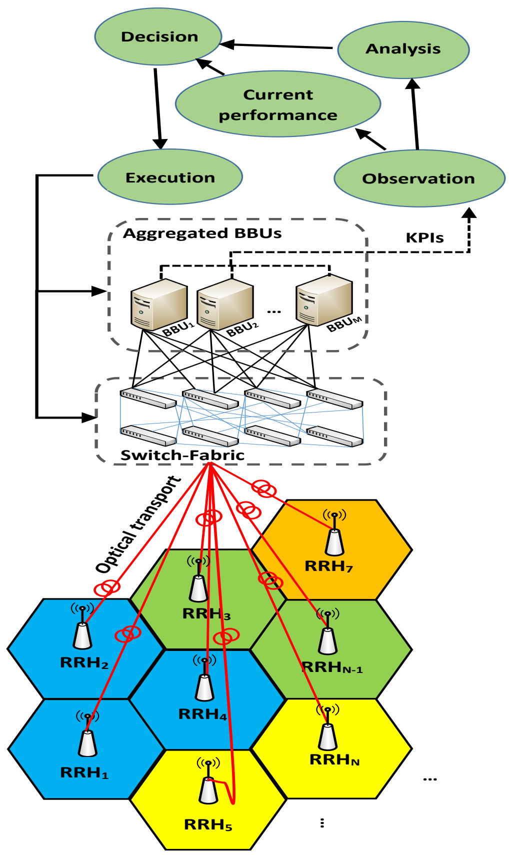

Fig. 1. Genetric concept of a Self organisation in C-RAN

Fig. 1 shows a generic concept of a self-organisation in C-RAN. The RRHs are connected to the aggregated BBU pool via optical transport network (often termed as *front-haul*) which may consist of a switch fabric (including a network of switches, optical splitters, multiplexers) [31] and low latency, high bandwidth fibre optic links. Note that, there are various possibilities of front-haul deployment in C-RAN architecture [41]. The RRHs are equipped with omnidirectional antennas and rests at the centre of small virtual-cells (micro-cells). The self-organisation concept is explained in phases as shown in Fig. 1, where the observation and analysis phases are utilised to detect the performance of current network deployment (BBU-RRH configuration), and then an optimal implementation is identified for performance comparison. KPIs are used to monitor network status for current and optimal deployment settings. Based on the chosen KPIs, an algorithm decides the best system configuration, and finally, the new topology (BBU-RRH setting) is enforced in the execution phase (if necessary).

A. *Proposed system model*

This paper introduces a SON server/controller inside the BBU pool to realise the self-organising concept as shown in Fig. 2. The SON server/controller hosts an intelligent algorithm to identify proper network setting dynamically. The primary objective of SON server/controller in the C-RAN architecture is as follows: (i) to compile required metric by discovering the status of each KPI and (ii) to produce a decision and enforce it. Fig. 3 provides a logical block diagram for a multiple objective decision making performed by the SON server/controller. The BBUs feed the system KPIs to the main multi-objective decision-making algorithm hosted by SON server/controller. The weights or priority levels are then applied to each KPI for decision making. The corresponding weight of a KPI defines its preference value and is set according to network operator's preferences.

In traditional cellular systems, a Macro-cell may divide into multiple sectors, and each sector may have its set of frequency channels. In this paper, the small micro virtual-cells are sectored dynamically such that each sector satisfies its hard-capacity (i.e., the maximum number of connected users). Furthermore, a group of RRHs (one cell per RRH) adjacent to each other forms a sector. A sector is a cluster or group of compact RRHs served by the same BBU, as shown in Fig. 2. Note that, the virtual-cells presented in this paper can support on-demand capacity by routing BBU resources to remote RRHs which enables BBU resources virtualisation at individual independent RRHs. Since each RRHs can access entire BBU cloud resources, the need for higher capacity by adding new base stations and bandwidth in traffic hot-spots may become unnecessary. Note that, each RRH can be allocated to only one sector at a particular time period. A different colour represents each sector in Fig. 2. Each BBU serves multiple sectors with multiple RRHs within these sectors, independently. The sectors served by each BBU can be identified by the colour assigned to each BBU, as shown in Fig. 2. The SON server/controller is responsible for proper RRH-sector allocation, and the switching fabric is in charge of realising these configurations via server commands in real time. Note that, the switching technology itself is challenging

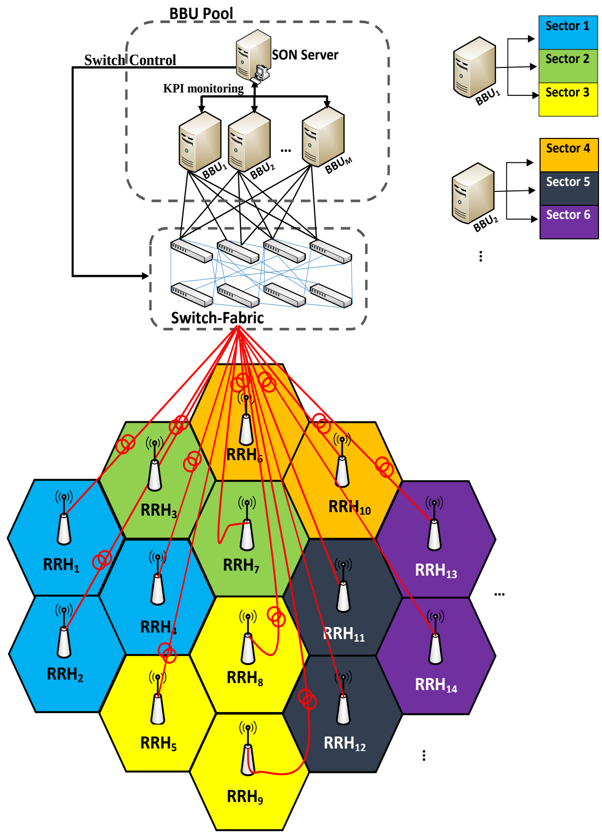

Fig. 2. Structure of a Cloud Radio Access Network represented as SON

in C-RAN. Optical switches are advantageous over electronic switches regarding cost, power, and data rate. However, they may incur significantly longer reconfiguration times. Discussion on the main challenges and potential solutions for front-haul deployments in C-RAN is explained in [41]. Note that, the hexagonal cell layouts shown in Fig. 1, Fig. 2, and Fig. 4 are considered only because of their well-defined shape and the fact that it uniformly covers the entire coverage area. They are merely used in Fig. 1, Fig. 2 and Fig. 4 to understand and evaluate the proposed concept.

B. *System model constraints*

This paper presents a system model designed as a centralised-SON architecture for C-RAN, which allows for more efficient resource utilisation through centralised control across aggregated BBU resources. However, the model is constrained in the following ways:

- Since the SON server/controller is in charge of monitoring the BBU-cloud, the whole network may collapse in case of server/controller failure.
- Coarse time-scales may limit the optimisation process due to interface-latency between SON controller and the BBUs, along with the front-haul latency.
- Depending on the front-haul technology used, the front-haul must support enough bandwidth for delivering delay sensitive signals, and the switching elements used to effect the BBU-RRH configurations

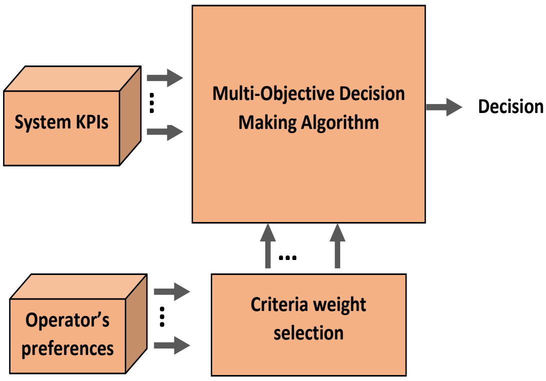

Fig. 3. Block diagram of multi-objective decision making logic for SON server

must not affect the sub-frames time scale (i.e., the 1ms duration of each subframe in LTE)

Potential solutions to the challenges/limitations mentioned above are discussed in [42] and [43].

IV. Dynamic RRH-sector re-allocation and formulation

Traffic variations occur at RRH coverage area in different periods of time. It is unavoidable to map the RRHs to BBU sectors in such a way that the RRH-sector allocation satisfies network limitations. Hard-capacity is one important limitation which is defined as the number of simultaneously connected users to an eNodeB or the maximum number of channel elements assigned to an eNodeB. Every user requires a Radio Resource Control (RRC) to access network services. However, a user registered on an LTE network can have 2 RRC states, i.e., RRC-idle and RRC-connected. In the RRC-idle state, the user equipment (UE) monitors control channels to determine whether any data is scheduled for communication while no data transfer takes place from the UE. However, in the RRC-connected state, transfer of data to/from the UE takes place. The number of connected users, i.e., RRC-Connected users per cell or sector is limited to hardware and software licensing limitations. This paper defines the hard-capacity of a sector as the number of RRC-connected users.

Network Performance indicates its QoS, which is determined by various Key Performance Indicators (KPIs). Based on these KPIs, the network must react pro-actively to avoid performance- threatening events that can cause unavailability of network services and infrastructure. Vendors evaluate their network performance using a different set of objectives mapped to a pre-defined QoS metrics. However, a weighted normalised function is required when multiple objectives are considered. This paper presents a QoS function by defining new KPIs for C-RAN, based on which the SON controller/server can identify optimum RRH-sector configuration and perform load balancing in the network.

When the RRH-sector configuration at time t is known, then finding the optimum RRH-sector configuration at time $t+1$ is the main objective to balance the load. Consider N number of BBUs in a BBU pool serving M number of RRHs distributed over a geographical area divided into S sectors. Each BBU serves multiple sectors. Let SOS_n be a set of sectors served by BBU_n, i.e., $|\mathrm{SoS}_n| = 3$, if BBU_n serves 3 sectors. Similarly, SOR_s represents the set of RRHs occupied by Sector_s. Let the RRH-sector allocation at time t is represented by a vector $\mathbf{R}_s^t = \{\mathrm{R}_{1s}^t, \mathrm{R}_{2s}^t, \mathrm{R}_{3s}^t, ..., \mathrm{R}_{\mathrm{M}s}^t\}$, where $s = 1, ...\mathrm{S}$, then finding the new RRH-sector allocation vector at time $t+1$ $\left(\mathbf{R}_s^{t+1} = \{\mathrm{R}_{1s}^{t+1}, \mathrm{R}_{2s}^{t+1}, \mathrm{R}_{3s}^{t+1}, ..., \mathrm{R}_{\mathrm{M}s}^{t+1}\}, s = 1, ..., \mathrm{S}\right)$ is the main objective. The binary variables R_{is}^t and R_{is}^{t+1} $(i = 1, ..., \mathrm{M}$ and $s = 1, ..., \mathrm{S})$ indicates the RRH assignment to a sector at a particular time period. For example $\mathrm{R}_{is}^t = \mathrm{R}_{is}^{t+1} = 1$, when RRH_i is assigned to Sector_s at both time period t and $t+1$. It is assumed that each RRH coverage area has U_i $(i = 1, ..., \mathrm{M})$ number of connected users at time period t and $t+1$. Notice that, a user U_A is associated with RRH_B only if the Uplink power received from U_A at RRH_B is higher than in all other existing RRHs. If the probability of users transition from RRH_i to RRH_j is ρ_{ij}, then the handovers from RRH_i to RRH_j is represented as $\mathrm{H}_{ij} = \rho_{ij} U_i$. The estimation for real-time ρ_{ij} is not considered in this paper as it is not the main objective. However, many models on terminal mobility to observe ρ_{ij} exist in literature [44] [45] and can be utilised to determine ρ_{ij}. This paper model ρ_{ij} to be inversely proportional to the distance between RRH_i and RRH_j i.e., $\left(\rho_{ij} = \frac{1}{D_{ij}}\right)$ because a uniform user distribution is considered within each RRH coverage area. Notice that, to produce a non-uniform user distribution within the entire network coverage area, a different number of users per RRH is considered.

Following are the important KPIs considered for RRH-sector allocation problem. Fig. 4 shows an example of RRH-sector allocation at both time period t and $t+1$. The example consists of two BBUs and ten RRHs. Notice that, BBU_1 handles Sector_1 and Sector_2 whereas BBU_2 handles Sector_3 and Sector_4. The KPIs are calculated for both time period t and $t+1$.

A. Key Performance Indicator for blocked Users (KPI_{BU})

The number of users deprived of network services due to hard-capacity is considered as blocked users. This happens when the number of connected users in a Sector_s exceeds its hard-capacity (HC_s). The blocked users (BU) in the network at time $t+1$ can be calculated as follows:

$$\mathrm{BU} = \sum_s \max\left[\left(\left(\sum_i \mathrm{U}_i \mathrm{R}_{is}^{t+1}\right) - \mathrm{HC}_s\right), 0\right] \tag{1}$$

where $i = 1, 2, ..., \mathrm{M}$ and $s = 1, 2, ..., \mathrm{S}$. Then the KPI for blocked users ($\mathrm{KPI_{BU}}$) can be presented as

$$\mathrm{KPI_{BU}} = \begin{cases} 1 & \text{if } \mathrm{BU} = 0 \\ \frac{1}{\mathrm{BU}} & \text{otherwise} \end{cases} \tag{2}$$

where the binary variable $\mathrm{R}_{is}^{t+1} = 1$, if RRH_i belongs to Sector_s at time $t+1$. U_i represents the number of connected users served by RRH_i whereas HC_s represents the hard-capacity of Sector_s. Note that, if the number of

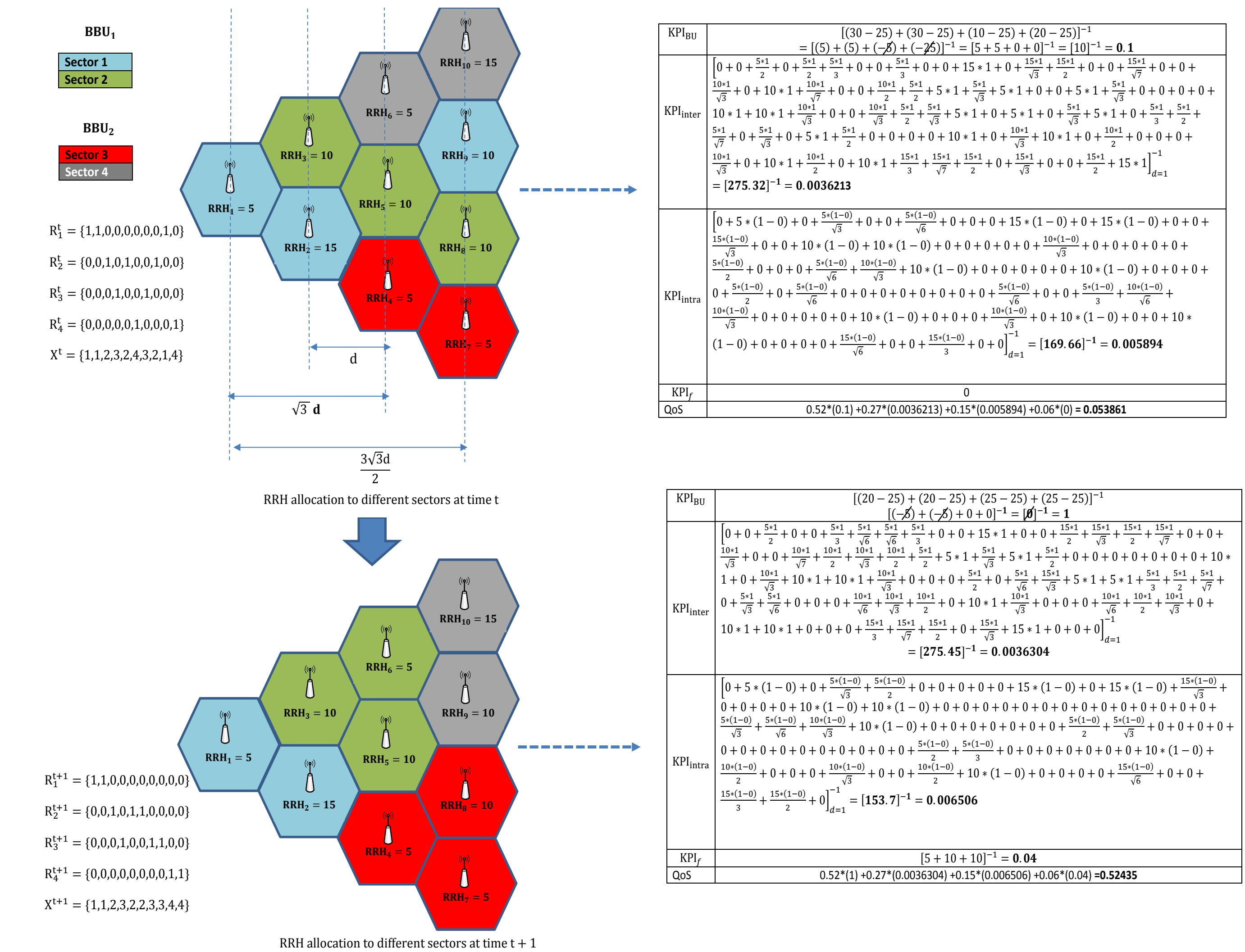

Fig. 4. An example of RRH allocation to different sectors at time t and $t+1$

connected users in a sector is lower than the its hard-capacity, then to avoid negative counting the function $\sum_s \max\left[\left(\left(\sum_i \mathrm{U}_i \mathrm{R}_{is}^{t+1}\right) - \mathrm{HC}_s\right), 0\right], \ \forall i, s$ is considered to skip counting the negative value. In Fig. 4, a hard-capacity of 25 is assumed for each sector. At time t, 10 blocked users are observed (i.e, $\mathrm{KPI}_{\mathrm{BU}} = [\frac{1}{10}] = 0.1$), however, at time $t+1$, the network is well balanced with no blocked calls. Note that, $\mathrm{KPI}_{\mathrm{BU}} = 1$, if there are no blocked users in the network.

B. Key Performance Indicators for Handovers

Different KPIs are considered for following types of handovers:

1) Inter-BBU handovers: Handovers are necessary functions provided by the network to maintain the QoS of ongoing user sessions and to associate users with the best possible eNodeBs. In LTE/LTE-A, inter-eNodeB handovers are performed based on X2 interface between the eNodeBs, where users move from one eNodeB to another eNodeB, both connected to the same MME. However, if serving and target eNodeBs are not attached to the same MME, then S1 based inter-eNodeB handovers are performed. Detailed information on X2 and S1 based inter-eNodeB handovers are presented in a particular section in [46]. The same concept holds true for inter-BBU handovers in this paper. Due to the structural difference between C-RAN and LTE/LTE-A, the inter-eNodeB handovers and intra-eNodeB handovers are referred to as inter-BBU handovers and intra-BBU handovers for C-RAN, respectively.

Let Y_{ijn}^{t+1} be a binary variable such that $\mathrm{Y}_{ijn}^{t+1} = 1$ when both RRH_i and RRH_j are served by BBU_n at time period $t+1$ i.e., $\mathrm{Y}_{in}^{t+1} = \mathrm{Y}_{jn}^{t+1} = 1$, also $\mathrm{Y}_{in}^{t+1} = \sum_{s \in \mathrm{SOS}_n} \mathrm{R}_{is}^{t+1}$. Inter-BBU handover cost variable is measured by using a binary variable Y_{ij}^{t+1} such that $\mathrm{Y}_{ij}^{t+1} = 1 - \sum_n \mathrm{Y}_{ijn}^{t+1}$. $\mathrm{Y}_{ij}^{t+1} = 1$, when RRH_i and RRH_j are served by different BBUs at time period $t+1$. The inter-BBU handovers at time $t+1$ can now be calculated as:

$$\text{Inter-BBU}_{\text{HO}} = \sum_i \sum_{j \neq i} \mathrm{H}_{ij} \mathrm{Y}_{ij}^{t+1} \tag{3}$$

where $\mathrm{H}_{ij} = \rho_{ij} \mathrm{U}_i$ represents handovers from RRH_i to RRH_j and ρ_{ij} is modelled as the inverse of the distance between RRH_i and RRH_j (i.e., $\rho_{ij} = \frac{1}{\mathrm{D}_{ij}}$). Note that,

Y_{ij}^{t+1}=$f_1(Y_{ijn}^{t+1})$, $Y_{ijn}^{t+1} = f_2(Y_{in}^{t+1}.\ Y_{jn}^{t+1})$, Y_{in}^{t+1}=$f_3(R_{is}^{t+1})$, and Y_{jn}^{t+1}=$f_4(R_{js}^{t+1})$. Therefore, the variable Y_{ij}^{t+1} is a function of the main binary variable R_{is}^{t+1}, $i = 1,...,M$ i.e., $\left(Y_{ij}^{t+1}=f_1(f_2(f_3(R_{is}^{t+1}), f_4(R_{is}^{t+1})))\right)$. The detailed form of Inter-BBU$_{HO}$ can be given as follows:

$$\text{Inter-BBU}_{HO} = \sum_i \sum_{j\neq i} H_{ij} Y_{ij}^{t+1}$$

$$\text{Inter-BBU}_{HO} = \sum_i \sum_{j\neq i} \rho_{ij} U_i \left(1 - \sum_n Y_{ijn}^{t+1}\right)$$

$$\text{Inter-BBU}_{HO} = \sum_i \sum_{j\neq i} \rho_{ij} U_i \left(1 - \sum_n \left(Y_{in}^{t+1}.Y_{jn}^{t+1}\right)\right)$$

$$\text{Inter-BBU}_{HO} = \textstyle\sum_i \sum_{j\neq i} \frac{U_i}{D_{ij}} \left(1 - \left(\sum_n \sum_{s\in SOS_n} \left(R_{is}^{t+1}.\ R_{js}^{t+1}\right)\right)\right) \tag{4}$$

where the dot(.) operators used in (4) show the logical AND operation. The KPI for inter-BBU handovers $(\text{KPI}_{\text{inter}})$ can now be given as:

$$\text{KPI}_{\text{inter}} = \begin{cases} 1 & \text{if Inter-BU}_{HO} = 0 \\ [\text{Inter-BBU}_{HO}]^{-1} & \text{otherwise} \end{cases} \tag{5}$$

2) Intra-BBU handovers : Intra-eNodeB handovers are performed when users transition from one sector to another becomes necessary. Provided that the sectors involved in users transition are handled entirely within the eNodeB. Intra-eNodeB sector changes are not normally notified to the MME [46]. To compute users transition from one sector to another under the same BBU, a binary variable Z_{ijs}^{t+1} is introduced such that $Z_{ijs}^{t+1} = 1$, if both RRH$_i$ and RRH$_j$ belong to Sector$_s$ at time period $t+1$ $\left(\text{i.e., } R_{is}^{t+1} = R_{js}^{t+1} = 1\right)$. The cost variable for intra-BBU handovers can be determined by utilising two variables Z_{ij}^{t+1} and Y_{ij}^{t+1}. The variable $Z_{ij}^{t+1} = 1$ if RRH$_i$ and RRH$_j$ belong to different sectors at time period $t+1$ and can be presented as $Z_{ij}^{t+1} = 1 - \sum_s Z_{ijs}^{t+1}$. To solve if the RRHs are served by the same BBU, the binary variable Y_{ij}^{t+1} calculated previously, is used i.e., $Y_{ij}^{t+1} = 0$ if RRH$_i$ and RRH$_j$ are served by the same BBU. Therefore, the intra-BBU handovers (Intra-BBU$_{HO}$) at time $t+1$ can be computed as:

$$\text{Intra-BBU}_{HO} = \sum_i \sum_{j\neq i} H_{ij} \left(Z_{ij}^{t+1} - Y_{ij}^{t+1}\right) \tag{6}$$

Note that, the variables $Z_{ij}^{t+1} = f_5(Z_{ijs}^{t+1})$ and $Z_{ijs}^{t+1} = f_6(R_{is}^{t+1}, R_{js}^{t+1})$. Therefore, the binary variable Z_{ij}^{t+1} is a function of the main binary variable R_{is}^{t+1}, $i = 1,2,...,M$. i.e., $\left(Z_{ij}^{t+1} = f_5(f_6(R_{is}^{t+1}, R_{js}^{t+1}))\right)$. The detailed form of Intra-BBU$_{HO}$ can be given as follows:

$$\text{Intra-BBU}_{HO} = \sum_i \sum_{j\neq i} H_{ij} \left(Z_{ij}^{t+1} - Y_{ij}^{t+1}\right)$$

$$\text{Intra-BBU}_{HO} = \textstyle\sum_i \sum_{j\neq i} \rho_{ij} U_i \left[\left(1 - \sum_s Z_{ijs}^{t+1}\right) - \left(1 - \sum_n Y_{ijn}^{t+1}\right)\right]$$

$$\text{Intra-BBU}_{HO} = \sum_i \sum_{j\neq i} \frac{U_i}{D_{ij}} \Bigg[\left(1 - \sum_s (R_{is}^{t+1}.\ R_{js}^{t+1})\right) - \left(1 - \sum_n \sum_{s\in SOS} \left(R_{is}^{t+1}.\ R_{js}^{t+1}\right)\right)\Bigg]$$

$$\text{Intra-BBU}_{HO} = \sum_i \sum_{j\neq i} \frac{U_i}{D_{ij}} \Bigg[\sum_n \sum_{s\in SOS} (R_{is}^{t+1}.\ R_{js}^{t+1}) - \sum_s \left(R_{is}^{t+1}.\ R_{js}^{t+1}\right)\Bigg] \tag{7}$$

where the dot(.) operators used in (7) show the logical AND operation. An important constraint here is that each RRH is served by a single BBU at time $t+1$ i.e., $\sum_{n=1}^{N} R_{in}^{t+1} = 1$. This indicates that RRHs in the same sector cannot be served by multiple BBUs at any given time t. The KPI for intra-BBU handovers (KPI$_{\text{intra}}$) can now be given as:

$$\text{KPI}_{\text{intra}} = \begin{cases} 1 & \text{if Intra-BU}_{HO} = 0 \\ [\text{Intra-BBU}_{HO}]^{-1} & \text{otherwise} \end{cases} \tag{8}$$

3) Forced handovers : The primary objective of this paper is to find a new RRH configuration (RRH-sector allocation vector) that provides a higher QoS and a balanced network load at the cost of minimum possible handovers compared to existing RRH-sector configuration (or current RRH association to sectors). It means that an RRH might change its sector in time. Change in RRH's sector means all connected users in RRH coverage area are required to make a new sector transition. It is assumed that no call or session drops are experienced during user transitions due to a mechanism called soft handover. Since the BBUs are co-located in a common place and can communicate by exchanging data and control signals, it is possible to connect a user to multiple BBUs regardless of the modulation/access scheme. Soft handovers can be used not only for CDMA systems but non-CDMA systems as well [47]. In this procedure, the radio links assigned to users are attached and detached in such a manner that at least one radio link to the mobile network is kept active. It enables users to connect to multiple cell sectors during an active call session. To compute if the RRHs have changed their sectors, a new binary variable C_{is} is introduced. Where C_{is}= 1, if RRH$_i$ changes its current sector at time period t to Sector$_k$ at time period $t+1$ (i.e., $C_{is} = 1$, if $R_{is}^{t} = 0$ and $R_{is}^{t+1} = 1$). The forced handovers f_{HO} can then be presented as:

$$f_{\text{HO}} = \sum_s \sum_i C_{is} U_i \tag{9}$$

Note that, the binary varibale C_{is} is a function of the main binary variables R_{is}^{t} and R_{is}^{t+1}, $\forall i = 1,2,...,M$ i.e., $\left(C_{is} = f_7\left(R_{is}^{t}, R_{is}^{t+1}\right)\right)$. The detailed form of f_{HO} is given as:

$$f_{\text{HO}} = \sum_s \sum_i \left(R_{is}^{t} + R_{is}^{t+1}\right) U_i \tag{10}$$

where the operator (+) used in (10) shows the logical OR operation. The KPI for forced handovers (KPI_f) can now be represented as

$$\text{KPI}_f = \begin{cases} 1 & \text{if } f_{\text{HO}} = 0 \\ [f_{\text{HO}}]^{-1} & \text{otherwise} \end{cases} \tag{11}$$

V. RRH Proximity Constraint

RRH re-allocation to different BBU sectors in C-RAN provides enhanced flexibility in network management. However, an important limitation to consider is the reliable operation of C-RAN regarding RRH allocation to sectors for high network performances. Therefore, proper sectoring of RRHs (or microcells) is important to avoid cases like unbalanced traffic and increased user call blockings. Proper RRH sectoring means the RRH allocation to a sector should be connected. Notice that, all RRHs in a sector broadcasts radio signals simultaneously. RRH-sector configurations which produce disassociated RRH arrangement induce unnecessary handovers among the sectors. It is important for a sector to be consistent throughout its allocated RRHs. This compactness or consistency in sectors not only minimises unnecessary handovers among sectors but also minimises the interferences among them. The RRHs of a consistent sector have less common boundaries with the neighbouring sectors than the RRHs of an inconsistent sector. In Fig. 4, the detached RRH_9 of Sector_1 at time t is surrounded by RRHs allocated to other neighbouring sectors. Therefore, RRH_9 experienced high interferences due to common edges (boundaries) with other sectors. Furthermore, proximity and consistently connected RRH allocations in a sector not only decreases the number of handovers required for new RRH-sector allocation transition but also reduces the length of boundaries (cell edges) between sectors served by different BBUs. Therefore, the RRHs proximity is defined by introducing a binary variable A_{ij}, where A_{ij}= 1, if RRH_i and RRH_j are adjacent else A_{ij}= 0. Notice that, the proximity constraint requires every RRH to be allocated to a sector at time $t+1$ (i.e., $\Sigma_s \text{R}_{is}^{t+1} = 1$, $i = 1, 2, 3, ..., M$). If a sector has multiple RRHs, then the RRHs in that sector must be adjacent and connected. To formulate the connectedness and proximity of the RRHs, it is assumed that a Sector_s is connected. Let S1_s be any proper subset of the set of RRHs occupied by Sector_s (SOR_s), such that $\text{S1}_s \subset \text{SOR}_s$, $\text{S1}_s \neq \emptyset$, and $\text{S1}_s \neq \text{SOR}_s$. Let S2_s be another subset of SOR_s such that, $\text{S2}_s = \text{SOR}_s - \text{S}_{1s}$, i.e., S2_s is the complementary set of S1_s. To confirm that the RRHs in Sector_s are connected, the following property must be satisfied

$$\sum_{i \in \text{S1}_s} \sum_{j \in \text{S2}_s} \text{A}_{ij} \geq 1 \tag{12}$$

To maximise the network QoS, a weighted sum of all KPIs is taken to define the QoS function. Optimising the QoS function is the main objective

$$\begin{aligned} &\underset{\text{R}_s^T}{\text{M}ax}\ \text{QoS} = w_1 \text{KPI}_{\text{BU}} + w_2 \text{KPI}_{\text{inter}} + w_3 \text{KPI}_{\text{intra}} + w_4 \text{KPI}_{\text{f}} \\ &\text{subject to:} \\ &\sum_{s=1}^{\text{S}} \text{R}_{is}^{t+1} = 1, \forall i \\ &\sum_{n=1}^{\text{N}} \text{R}_{in}^{t+1} = 1, \forall i \\ &\sum_{i \in S1_s} \sum_{j \in S2_s} \text{A}_{ij} \geq 1, \forall i, j \end{aligned} \tag{13}$$

where w_1,w_2,w_3, and w_4 are the priority levels of the defined KPIs. Furthermore, the binary variables Y_{ij}^{t+1}, Z_{ij}^{t+1}, and C_{is} are all functions of the main binary variables R_{is}^{t+1}, where $i, j \in \{1, 2, ..., \text{M}\}, s \in \{1, 2, ..., \text{S}\}$, and $n \in \{1, 2, ...\text{N}\}$. Since the RRH allocation to a sector at time t is represented by $\mathbf{R}_s^t = \{\text{R}_{1s}^t, \text{R}_{2s}^t, \text{R}_{3s}^t, ..., \text{R}_{\text{Ms}}^t\}$, where the binary variables R_{is}^t inside the vector shows the existence ($\text{R}_{is}^t = 1$) and non-existence ($\text{R}_{is}^t = 0$) of RRH in Sector_s. The search space size thus becomes 2^{MS}. To decrease the search space size, we eliminate the first constraint from 13 (i.e., the constraint $\sum \text{R}_{is}^{t+1} = 1, \forall i$) by introducing a sector variable x_i^t, where $\mathbf{x}_i^t \in \{1, 2, ..., \text{S}\}$ and $i = 1, 2, ..., \text{M}$. This decreases the search space from 2^{MS} to S^{M}. The RRH-sector allocation vector at time $t+1$ $\left(\mathbf{R}_s^{t+1} = \{\text{R}_{1s}^{t+1}, \text{R}_{2s}^{t+1}, \text{R}_{3s}^{t+1}, ..., \text{R}_{\text{M}s}^{t+1}\}\right)$ is now translated into a new RRH-sector allocation vector $\mathbf{X}^{t+1} = \{\text{x}_1^{t+1}, \text{x}_2^{t+1}, ..., \text{x}_{\text{M}}^{t+1}\}$, where the sector variable x_i^{t+1} can be assigned only one sector at time t and $t+1$. $\text{x}_i^{t+1} = s$ indicates that RRH_i is assigned to Sector_s.

To find the optimal RRH-sector configuration, an exhaustive search for all possible RRH-sector combinations is required. This makes the size of search space 2^{MS}, where M and S represents the number of RRHs and sectors in the network, respectively. The search space size increases with the number of RRHs (M) and sectors (S). Based on the constraint presented in Section IV, each RRH has to be assigned to a single sector at time $t+1$ ($\sum_s \text{R}_{is}^{t+1} = 1, i = 1, 2, ..., \text{M}$). Therefore, the size of search space is reduced to S^{M} by introducing a sector variable x_i^t, where $\text{x}_i^t \in \{1, 2, ..., \text{S}\}$ and $i = 1, 2, ..., \text{M}$. The sector variable $\mathbf{X}^t = \{1, 1, 2, 3, 2, 4, 3, 2, 1, 4\}$ in Fig. 4 example shows that RRH_1, RRH_2, RRH_3, RRH_4, RRH_5, RRH_6, RRH_7, RRH_8, RRH_9, and RRH_{10} are allocated to Sector_1, Sector_1, Sector_2, Sector_3, Sector_2, Sector_4, Sector_3, Sector_2, Sector_1, and Sector_4, respectively. The QoS objective function in (13) can now be represented as:

$$\text{Max QoS} = w_1\text{KPI}_{\text{BU}} + w_2\text{KPI}_{\text{inter}} + w_3\text{KPI}_{\text{intra}} + w_4\text{KPI}_f$$
$$\text{X}^T$$
subject to:
$$\sum_{n=1}^{\text{N}} \text{R}_{in}^{t+1} = 1, \forall i$$
$$\sum_{i \in S1_s} \sum_{j \in S2_s} \text{A}_{ij} \geq 1, \forall i, j \tag{14}$$

The optimum RRH-sector allocation is identified by exhaustively searching the entire search space for all possible RRH-sector allocations. Since the number of possible RRH-sector allocations increases exponentially with the number of RRHs and sectors, the algorithm execution time also increases exponentially. Therefore, evolutionary algorithms are proposed in Section VI to solve the RRH-sector allocation as an optimisation problem. A detailed form of (14) is presented in (15), subject to the constraints provided in (2), (5), (8), (11), and (14).

VI. Self-optimised Cloud Radio Access Network (SOCRAN) Algorithm

A SOCRAN algorithm is suggested in this section that depends on the above intuitive analysis and is based on a centralised algorithm running on the SON server/controller. The SOCRAN algorithm utilises aggregate network gain information, which consists of network KPIs, to execute appropriate RRH allocation to BBU sectors. Fig. 5 describes the SOCRAN algorithm block diagram. Network information is utilised for KPI analysis and QoS measurement for current RRH-sector configuration in the first step. In the optimisation step, the same information is employed for KPI and QoS analysis of other nominated RRH-sector allocations. This information involves Users served per RRH, RRH to RRH separations/distances, and initial RRH-sector configuration. The SOCRAN algorithm adjusts the RRH-sector configuration at the end of optimisation procedure by comparing the QoS values of both initial and optimised RRH-sector configuration. The KPIs are maximised by SOCRAN so as to improve the QoS of the network by utilising evolutionary algorithms.

This paper examines Genetic Algorithm (GA) and Discrete Particle Swarm Optimisation (DPSO) as evolutionary algorithms to solve the RRH-sector allocation optimisation problem. Both GA [48] and PSO [49] are population-based search algorithms, where population means a collection of candidate solutions. Chromosomes in GA and particles in PSO produces solution strings which collectively forms a population. The fitness value of each candidate solution indicates the measure of solution quality in problem-solving. In a GA, a group of random candidate solutions (or chromosomes) are created and represented individually in a population. The individual solutions are then evaluated based on how well they perform at a given problem function. The individuals with higher fitness levels are then selected for a technique inspired by natural evolution to produce new candidate solutions/chromosomes, such as mutation and crossover. The process continues until an optimal/near-optimal solution is achieved or a certain stopping criterion is satisfied, i.e., a predefined number of generations have passed. Unlike GA, PSO utilises a Swarm (or population) of particles where each particle represents a candidate solution. These particles probe the solution space (or search space) randomly with different velocities. To direct the particles to their best fitness values, the velocity of an individual particle is changed stochastically at each iteration (i.e., generating new particles). The velocity update of each particle depends on the historical best position experience (pbest) of the particle itself and the best position experience of neighbouring particles, i.e., the global best position (gbest). Since the solution vector (i.e., the RRH-sector allocation vector $\mathbf{X}^T$) is real-valued, the standard PSO algorithm can not be applied to solve this discrete optimisation problem. In this paper, a Discrete Particle Swarm Optimisation (DPSO) is used to solve the QoS maximisation problem defined in (14). In general, the parameters and notations used to determine both GA and DPSO are given in Table I

The GA and DPSO are explained in the following steps, and Fig. 5 represents the SOCRAN algorithm

Step 1: Generate the initial population R^0 with $|\Delta|$, M-bit chromosomes/particles (RRH-sector allocations). M is taken according to the number of RRHs in the network. For DPSO, initialise the best position for each particle $\text{pbest}_j^0 = \text{r}_j^0, 1 \leq j \leq |\Delta|$ and assign random velocity v_j^I to each particle.

Step2: Calculate the fitness value of each chromosome/particle (RRH-sector allocation) in current population using the fitness function F (i.e., QoS in 14). For DPSO, initialise global best position as, $\text{gbest}^0 = \underset{1 \leq j \leq |\Delta|}{\text{argmax}} \, \text{F}(\text{pbest}_j^I)$.

Step 3: For GA, if the convergence criterion is qualified by the best candidate RRH-sector solution (chromosome) or the maximum number of generations have passed, then end, else proceed to step 4. For DPSO, Update particle j position by updating its velocity. The velocity update equation is given as

$$\text{Max QoS} = w_1\left[\sum_s \max\left[\left(\left(\sum_i \text{U}_i\text{R}_{is}^{t+1}\right) - \text{HC}_s\right), 0\right]\right]^{-1} + w_2\left[\sum_i \sum_{j \neq i} \frac{\text{U}_i}{\text{D}_{ij}}\left(1 - \left(\sum_n \sum_{s \in \text{SOS}_n}\left(\text{R}_{is}^{t+1}.\ \text{R}_{js}^{t+1}\right)\right)\right)\right]^{-1}$$
$$+ w_3\left[\sum_i \sum_{j \neq i} \frac{\text{U}_i}{\text{D}_{ij}}\left(\sum_n \sum_{s \in \text{SOS}}\left(\text{R}_{is}^{t+1}.\ \text{R}_{js}^{t+1}\right) - \sum_s\left(\text{R}_{is}^{t+1}.\ \text{R}_{js}^{t+1}\right)\right)\right]^{-1} + w_4\left[\sum_s \sum_i \left(\text{R}_{is}^t + \text{R}_{is}^{t+1}\right)\text{U}_i\right]^{-1} \tag{15}$$

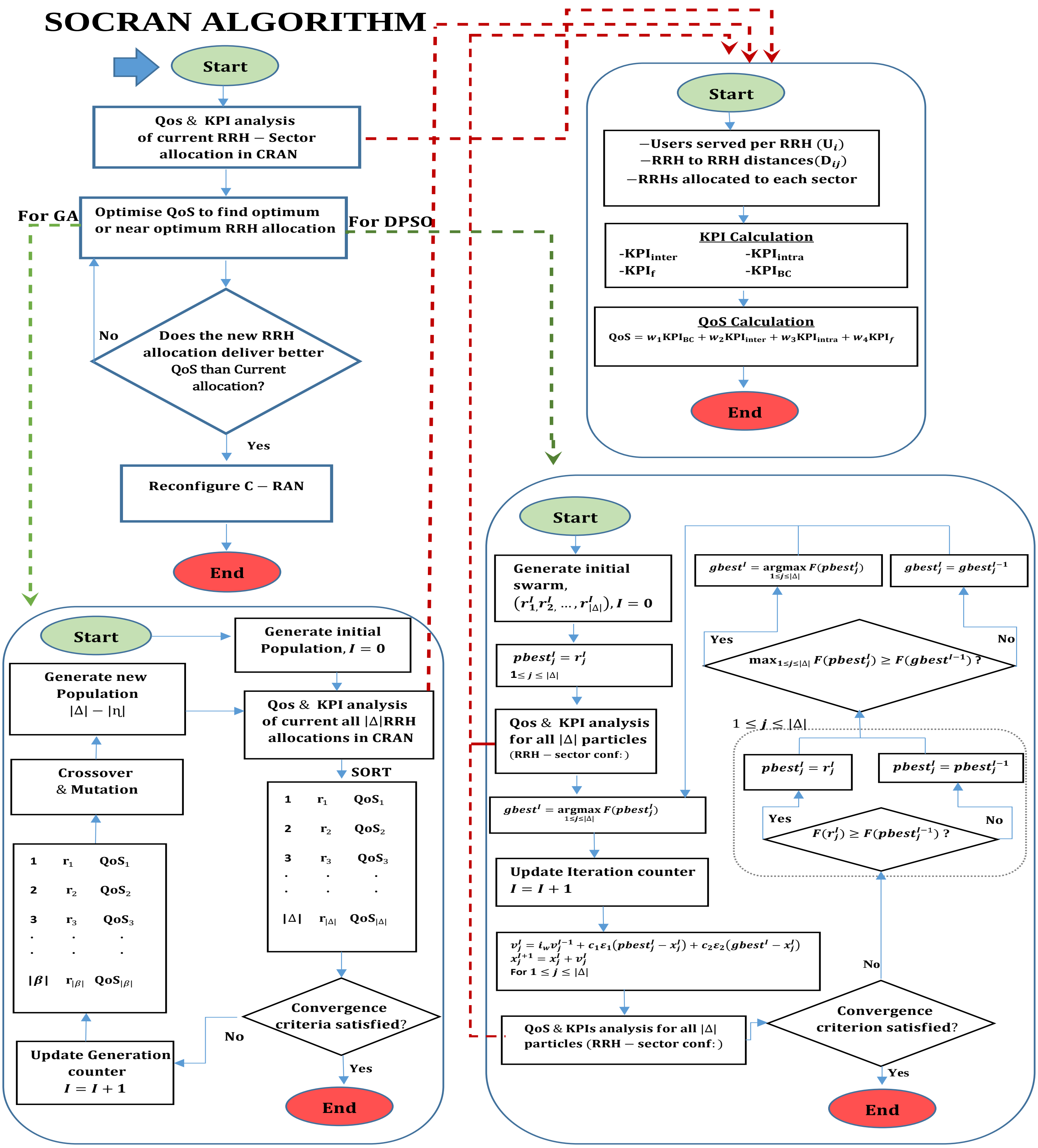

Fig. 5. SOCRAN Algorithm block diagram

$$v_j^I = i_w v_j^{I-1} + c_1\varepsilon_1\left(\text{pbest}_j^I - x_j^I\right) + c_2\varepsilon_2(\text{gbest}_j^I - x_j^I)$$
$$1 \le j \le |\Delta| \tag{16}$$

where x_j^I is the current position of particle j in iteration I and $\varepsilon_1, \varepsilon_2$ are random numbers between 0 and 1. Both c_1 and c_2 are acceleration constants that pulls the particle towards best position. Values in the range 0-5 are chosen for c_1 and c_2. The inertial weight i_w represents the effect of preceding velocity on the updated velocity. Larger and smaller value of i_w are

TABLE I
NOTATIONS DEFINITION FOR GA AND DPSO

$\|\Delta\|$	Population/Swarm size
$I_{\max}$	Maximum iterations/generations
R^I	Population/Swarm at I^{th} iteration/generation $\mathrm{R}^I = \left[\mathrm{r}_1^I; \mathrm{r}_2^I; ...; \mathrm{r}_{\|\Delta\|}^I\right]$
r_j^I	Particle/chromosome positioned at index j representing RRH-sector allocation at I^{th} iteration/generation
$\mathrm{r}_{j,m}^I$	RRH_m in a Particle/Chromosome solution which is indexed at position j in the I^{th} iteration/generation
β^I	Set of best RRH-sector allocations chosen from Population R^I at I^{th} generation
P_c	Cross-over probability
P_s	Selection probability
P_m	Mutation probability
pbest_j^I	Best position of particle j up to I^{th} iteration. $\mathrm{pbest}_j^I \equiv \underset{1\le s\le I}{\operatorname{argmax}} \mathrm{F}(\mathrm{r}_j^s)$
gbest^I	Best position experienced by any particle up to I^{th} iteration. $\mathrm{gbest}^I \equiv \underset{1\le j\le \|\Delta\|, 1\le s\le I}{\operatorname{argmax}} \mathrm{F}(\mathrm{pbest}_j^s)$
F	Objective/Fitness function defined in (10)

used for global exploration and local search expedition in the search-space, respectively. However, choosing an optimum value for i_w can assist a balanced proportion between global and local exploration of the search space. Usually values between 0-1 are selected for i_w. A value of 0.9 for i_w is selected in this paper. The new position of particle j for the next iteration $I+1$ will be:

$$\mathrm{x}_j^{I+1} = \mathrm{x}_j^I + \mathrm{v}_j^I \tag{17}$$

Step 4: For GA, create a set of $|\beta|$ best chromosomes (RRH-sector allocations) from the currently sorted population R^I. The selection probability P_s is used to select the best RRH-sector allocations to form set β (i.e., $\beta = P_s|\Delta|$). For DPSO, update iteration counter (*I=I+1*).

Step 5: for GA, generate new chromosomes η ($|\eta| = |\mathrm{R}^I| - |\beta|$) by performing crossover and mutation operations on set β. The newly generated RRH-sector solutions η then replaces the infeasible solutions (R^I-β) of the current population R^I in order to generate a new population. For DPSO, end if convergence criteria are satisfied, else go to step 6.

Step 6: For GA, go to step 2 and repeat all steps. For DPSO, update particle j's personal best position as:

$$\mathrm{pbest}_j^I = \begin{cases} \mathrm{pbest}_j^{I-1} & \text{if} \quad \mathrm{F}(\mathrm{r}_j^I) \le \mathrm{F}(\mathrm{pbest}_j^{I-1}) \\ \mathrm{r}^{I-1} & \text{if} \quad \mathrm{F}(\mathrm{r}_j^I) > \mathrm{F}(\mathrm{pbest}_j^{I-1}) \end{cases} \tag{18}$$

Step 7: Update global best position achieved

$$\mathrm{gbest}^I = \begin{cases} \underset{1\le j\le|\Delta|}{\operatorname{argmax}} \mathrm{F}(\mathrm{pbest}_j^I) & \text{if}\, \mathrm{F}(\mathrm{pbest}_j^I) > \mathrm{F}(\mathrm{gbest}^{I-1}) \\ \mathrm{gbest}^{I-1} & \text{otherwise} \end{cases} \tag{19}$$

Step 8: Repeat all steps starting from step 1 for DPSO.

VII. COMPUTATIONAL RESULTS AND COMPLEXITY

The priority levels or weights selected for (14) is based on Rank Order Centroid (ROC) method [50]. The ROC is a simple method of assigning weights to some functions, ranked according to their priority or importance. The priority of each function is taken as an input and converted into weight. The following formula does the conversion:

$$w_i = \left(\frac{1}{\mathrm{F}}\right) \sum_{n=i}^{\mathrm{F}} \frac{1}{n} \tag{20}$$

where F is the number of functions (KPIs) and w_i is the weight of the i^{th} function. $\mathrm{KPI_{BC}}$ ranked first is weighted as $\left(1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}\right)/4 = 0.52$, $\mathrm{KPI_{inter}}$ ranked second is weighted as $\left(\frac{1}{2}+\frac{1}{3}+\frac{1}{4}\right)/4 = 0.27$, $\mathrm{KPI_{intra}}$ ranked third is weighted as $\left(\frac{1}{3}+\frac{1}{4}\right)/4 = 0.15$, KPI_f ranked fourth is weighted as $\left(\frac{1}{4}\right)/4 = 0.06$. Fig. 4 shows weights assigned for each KPI during QoS calculations.

A higher weight is given to the inter-BBU handovers due to the signalling overhead involved with this type of handovers. The aim is to achieve a suitable RRH-sector configuration with minimum inter-BBU handovers. Inter-BBU handovers require signalling among the BBUs participating in the transfer as well as the S-GW and MME. Network performance is degraded with increased amount of inter-BBU handovers, making the new RRH-sector transition. A lower priority level is selected for intra-BBU handovers compared to inter-BBU handovers, considering that the BBU itself manages the entire handover process without involving the MME and S-GW. The KPI for the number of blocked users has the highest priority and therefore given the highest weight.

Both in GA and DPSO, the initial population/swarm is randomly chosen with uniform distribution. Note that, If the population/swarm size is selected to be smaller compared to the search space size, an inappropriate or significantly small number of appropriate solutions (RRH-sector allocations) are achieved. The problem is tackled by considering a percentage of initial population/swarm satisfying RRH proximity constraints. An initial population/swarm with 30% of random particles or chromosomes, fulfilling the RRH proximity, are selected for both GA and DPSO algorithms.

This paper presents three benchmark problems P_1, P_2, and P_3 to analyse and verify the performance of the SOCRAN algorithm. The spatial distribution of RRHs in each benchmark problem follows a homogeneous Poisson Point Process (PPP), with coverage areas corresponding to a Voronoi tessellation as shown in Figs 6, 7, 8, 9, 10 and 11. The RRHs are distributed in the test area with a non-negative intensity $\lambda_{P_i}|\mathrm{A}|$, where $|A|$ is the area of test region such that $\mathrm{A}{=}\pi\mathrm{R}^2$ with Radius R=5 km and $\lambda_{P_i} = 0.45$ where i=1,2,3. Note that, each point in the PPP is stochastically independent of all the other points in the process. This allows RRHs to be located very close to each other but with significant coverage area. However, the natural inclusion of different cell sizes and shapes and the indefinite network extension in all directions makes it a more realistic approach. In P_1, 19 RRHs are divided into six sectors and served by two BBUs. In P_2, 37 RRHs fall into nine sectors and are served by three BBUs. Whereas in P_3, 61 RRHs are divided into twelve sectors and served by four BBUs as given in Table II. Each BBU is fixed to manage three sectors. The hard-capacity of each sector is considered to be 200, i.e., a maximum of 200 users can be served within a sector. The convergence rate in this paper is defined as the number of best RRH-sector allocations found in the total number of generations or iterations performed. Fig. 4-6 demonstrates P_1, P_2, and P_3 at time t and $t+1$ where the number inside each cell represents the number of connected users in that cell. To test the performance of the SOCRAN algorithm, an exhaustive search for optimal RRH-sector allocation is performed. The Exhaustive Search algorithm (ES) is a useful and efficient way to find all possible RRH-sector configurations. ES finds S^{M} solutions for P_1, P_2, and P_3, where S is the number of Sectors and M is the number of RRHs considered in each problem (i.e., 6^{19}, 9^{37}, and 12^{61} solutions for P_1, P_2, and P_3, respectively).

The SOCRAN algorithm is tested 30 times over 30 different initial RRH-sector configurations for each benchmark problem (P_1, P_2, and P_3) and the average of results obtained are considered for Monte Carlo Analysis. Note that, all 30 initial configurations are improper RRH-sector allocations and each improper configuration generates 80, 177, and 128 blocked users for P_1, P_2, and P_3, respectively. The performances of both GA and DPSO in the SOCRAN algorithm are compared to ES and K-mean clustering algorithm. K-mean is a known clustering approach considered for LTE and one of the most used clustering algorithms in wireless sensor networks (WSN) [51]. Since the main problem is to cluster RRHs into different sectors, the K-mean clustering algorithm is a suitable technique for such problems. The QoS figures for P_1, P_2, and P_3 are represented separately for simplicity. Figs. 7-9 shows the average QoS (fitness function) for P_1, P_2, and P_3, respectively. Whereas, Fig. 10 and Fig. 11 shows the average number of blocked users and the average number of actual handovers over 200 generations for all benchmark problems. ES algorithm finds the optimum values shown in figures after searching for all possible RRH-sector allocations (S^{M}). Notice that, the ES algorithm is independent of the number of generations or iterations. The optimum values obtained from ES algorithm are used to demonstrate the improvement at each generation/iteration of GA and DPSO. The final results from GA and DPSO (in the SOCRAN Algorithm) are shown in Table III and are compared with the result of ES and K-mean clustering algorithm.

TABLE II
SPECIFICATION OF THREE BENCHMARK PROBLEMS

Problem	# of RRH/micro-cells	# of BBUs	# of Sectors
P_1	19	2	6
P_2	37	3	9
P_3	61	4	12

For P_1 (19 RRHs and 2 BBUs), both GA and DPSO converges to the optimum RRH-sector configuration with an average QoS evaluation value of 0.5231. However, the average QoS value for initial improper RRH configuration is 0.00404 as shown in Table III. The QoS values for GA, DPSO, and ES are same. 188 optimal RRH-sector allocations are achieved by DPSO over 200 iterations with a convergence rate of 0.94 and 0.079 CPU seconds. However, the optimum RRH-sector allocations by GA are 184 over 200 generations with

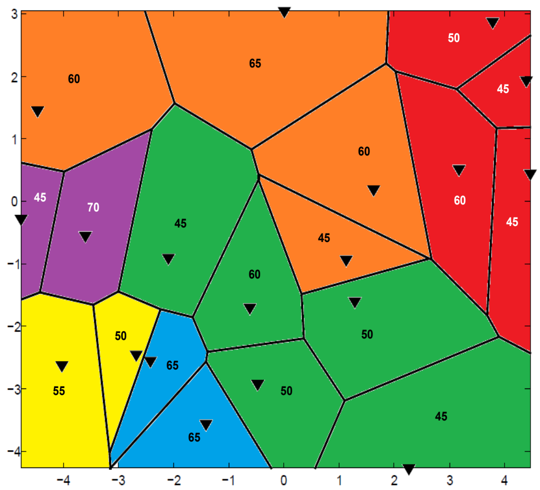

Fig. 6. Problem P_1 with improper RRH-sector allocation at time t

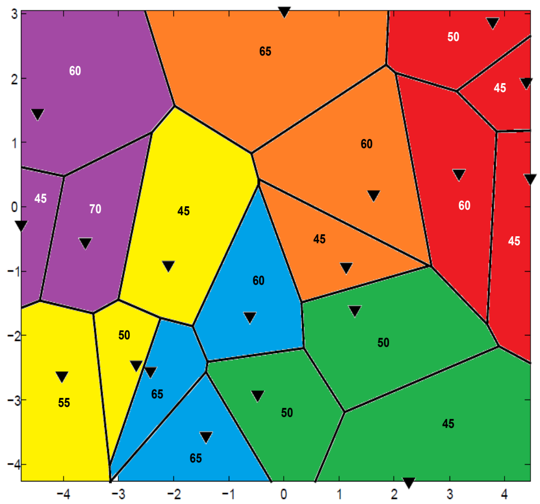

Fig. 7. Problem P_1 with proper RRH-sector allocation at time $t+1$

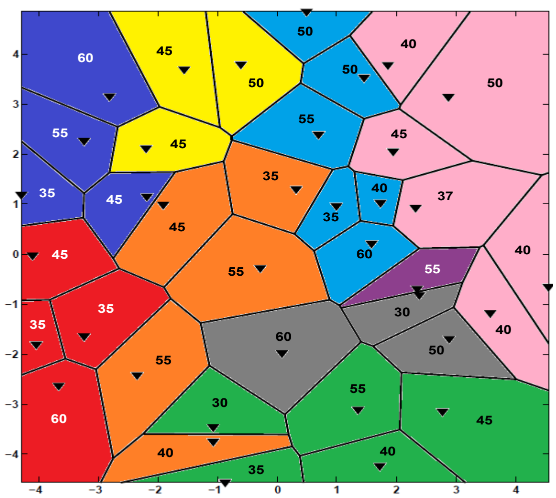

Fig. 8. Problem P_2 with improper RRH-sector allocation at time t

TABLE III
COMPUTATIONAL RESULTS

		P_1 (19 RRH)	P_2 (37 RRH)	P_3 (61 RRH)
Quality of Service	Improper RRH-sector allocation	0.00404	0.002952	0.00407
	Genetic Algorithm	0.5231	0.07211	0.02531
	Discrete Particle Swarm Optimisation	0.5231	0.0782	0.05214
	K-mean clustering	0.5231	0.0643	0.0041
	Exhaustive Search	0.5231	0.07989	0.05257
Blocked Users (%)	Improper RRH-sector allocation	7.766%	10.523%	5.394%
	Genetic Algorithm	0%	0.421%	0.611%
	Discrete Particle Swarm Optimisation	0%	0.409%	0.56%
	K-mean clustering	0%	0.765%	0.923%
	Exhaustive Search	0%	0.297%	0.252%
Forced Hand-Overs	Improper RRH-sector allocation	0	0	0
	Genetic Algorithm	152	173.5	115.2
	Discrete Particle Swarm Optimisation	152	172.1	120.4
	K-mean clustering	152	208.3	195.7
	Exhaustive Search	152	153	102

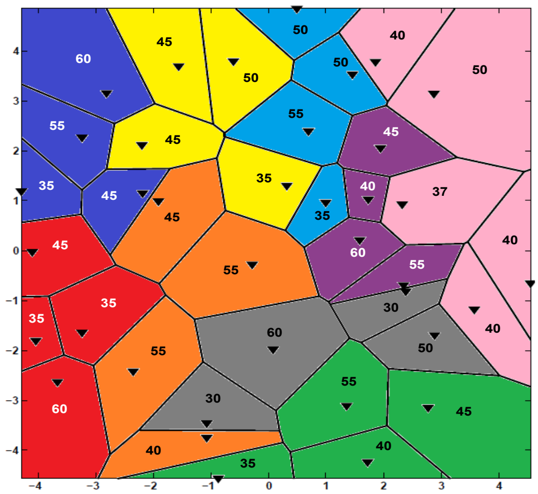

Fig. 9. Problem P_2 with proper RRH-sector allocation at time $t+1$

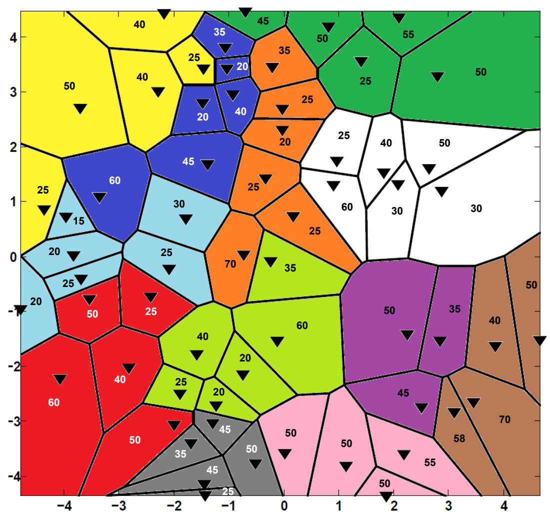

Fig. 10. Problem P_3 with improper RRH-sector allocation at time t

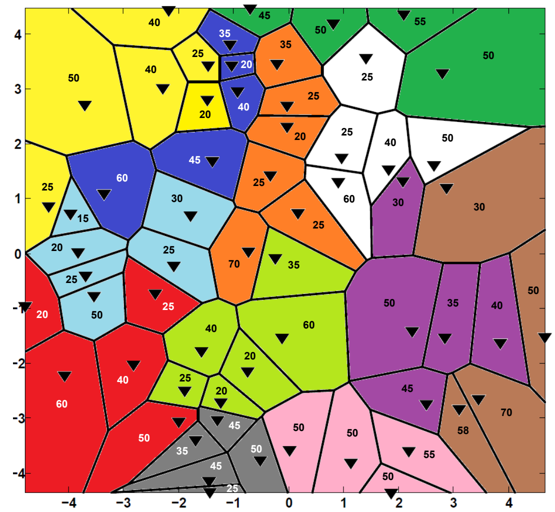

Fig. 11. Problem P_3 with proper RRH-sector allocation at time $t+1$

a convergence rate of 0.92 and 0.095 CPU seconds. The DPSO converges to the optimum RRH-sector allocation faster compared to GA (Fig. 12 and Table IV).The optimum QoS value is produced after 17×500 ($17 \times |\Delta|$) and 11×500 ($11 \times |\Delta|$) fitness evaluations by GA and DPSO, respectively. However, an exhaustive search of 6^{19} possible solutions is performed by ES to generate the optimal QoS value, which is too large. Note that, bot GA and DPSO deliver the same QoS value as the K-mean clustering algorithm for smaller networks (such as P_1 with 19 RRHs). Compared to GA, the DPSO converges to the optimum RRH-sector allocation configuration much faster as shown in Fig. 12 and Table III.

The QoS values evaluated for optimum and improper RRH-sector configuration for P_2 (37 RRHs) are 0.07989 and 0.002952, respectively, whereas 0.05257 and 0.00407, respectively, for P_3 (61 RRHs). The optimum QoS values are achieved by ES method. Since the number of RRH-sector allocations is too large in P_1 and P_2, both GA and DPSO fails to deliver optimum solution due to a considerable number of solutions with particularly limited generations/iterations. However, close optimum solutions are offered by both GA

TABLE IV
GA,DPSO AND K-MEAN CLUSTERING COMPARISON RESULTS

	QoS %			Convergence Rate		Number of Iterations or Generations		CPU time		
	GA	DPSO	K-mean	GA	DPSO	GA	DPSO	GA	DPSO	K-mean
19 RRH	128.48	128.48	128.48	0.92	0.94	16	12	0.095	0.079	0.21
37 RRH	23.42	25.49	20.78	0.52	0.84	97	32	0.19	0.16	0.54
61RRH	5.21	11.81	0.0073	0.135	0.655	173	69	0.55	0.35	0.67

and DPSO for both problems as shown in Table III and Fig. 13. The convergence rate of DPSO in P_2 (37 RRHs) is 0.84 with 0.16 CPU seconds, where 168 best RRH-sector allocations are found over the entire number of iterations. However, the convergence rate of GA is 0.52 with 0.19 CPU seconds, and 104 best RRH-sector allocations found over 200 generations as shown in Table IV. In P_3 (61 RRHs), the convergence rates of DPSO and GA are 0.655 and 0.135, respectively, with a CPU time of 0.35 seconds for DPSO and 0.55 seconds for GA. The DPSO delivers 131 best RRH-sector allocations over 200 iterations, but the GA provides 27 best RRH allocation solutions over 200 generations (Fig. 14 and Table IV). Even though both DPSO and GA can not find the optimum solution in P_1 and P_2, however, both the algorithms improves the network QoS by finding the best RRH-sector allocation compared to the improper RRH-sector allocation. Note that, the K-mean clustering algorithm takes longer times than GA and DPSO to find a proper RRH-sector allocation in all Benchmark problems as shown in Table IV.

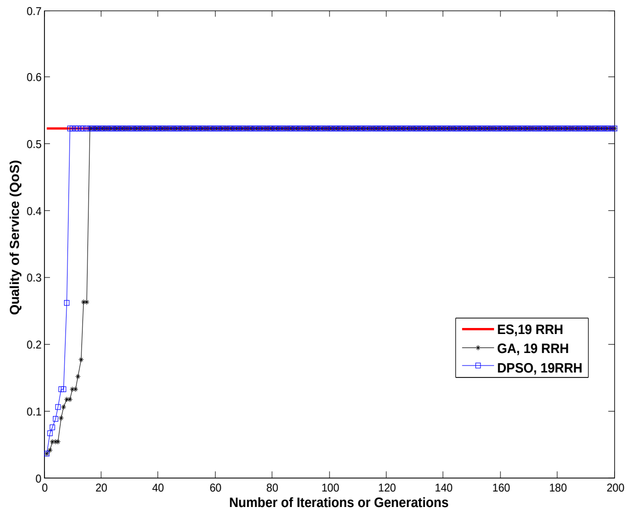

Fig. 12. Average Quality of Service (QoS) values for ES, GA, and DPSO in benchmark problem P_1

Fig. 15, Fig. 16, and Table III show the number of blocked users and handovers for both GA and DPSO. Both GA and DPSO minimise the number of blocked users in all benchmark problems. In P_2, although both algorithms do not achieve the optimum value, however, $\approx$ 63% and 68% of the optimum value is obtained by GA and DPSO, respectively. Similarly, in P_3, GA minimises the blocked users by $\approx$ 70% of the optimum value, whereas, $\approx$ 89% by DPSO. Figs 12-13 proves that the DPSO dominates GA regarding convergence rate. Moreover,

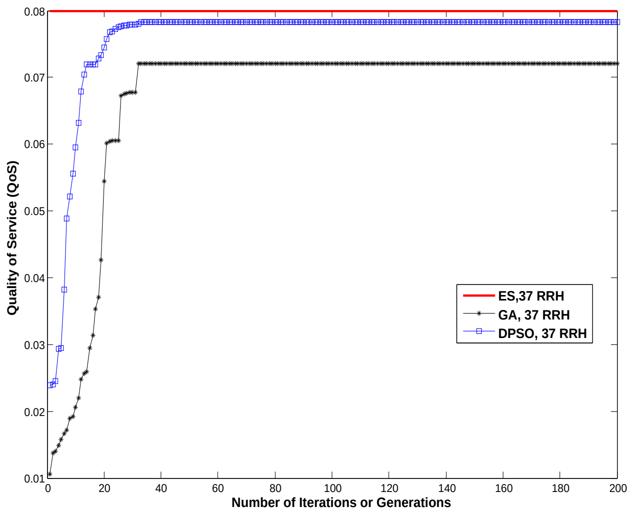

Fig. 13. Average Quality of Service (QoS) values for ES, GA, and DPSO in benchmark problem P_2

the RRH-sector allocation produced by DPSO are much closer to the optimum RRH-sector allocation provided by ES. The reason why DPSO outperforms GA in problem-solving is that in DPSO, the particles (RRH-sector allocations) acts as semi-autonomous agents that are aware of each others position status and decides to change their states (at each iteration) with respect to the best-observed particle position in the population. However, the chromosomes (RRH-sector allocations) in GA are not agent-like and lacks the ability to sense the neighbouring environment. GA relies on operations like crossover and mutation instead, to generate new population for the next generations. Crossover and mutation operations in GA disturbs better solutions and may converge into local optimal instead of an optimal global solution.

A QoS percentage (QoS%) is defined in this paper in order to present the progress level of both GA and DPSO, such that $\text{QoS}\% = \left(\frac{|\text{QoS}_i - \text{QoS}_b|}{\text{QoS}_i}\right)$. Where QoS_i is the QoS evaluation value for improper RRH-sector allocation, and QoS_b is the best QoS evaluation value at the last generation or iteration for GA and DPSO. In P_1, the QoS% for GA, DPSO and K-mean is 12.8. In P_2, the $\text{QoS}\%s$ for GA, DPSO and K-mean are 23.42, 25.49, and 20.78, respectively. The $\text{QoS}\%s$ for P_3 are 5.21, 11.81, and 0.0073 for GA, DPSO, and K-mean algorithm, respectively. $\text{QoS}\%s$ in all benchmark problems for GA, DPSO, and K-mean algorithm are shown in Table IV.

Both GA and DPSO can deliver improved RRH-sector allocations provided that the parameters selected for a given

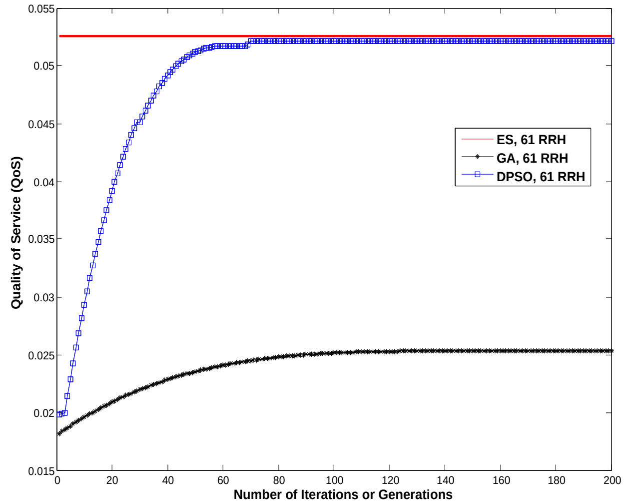

Fig. 14. Average Quality of Service (QoS) values for ES, GA, and DPSO in benchmark problem P_3

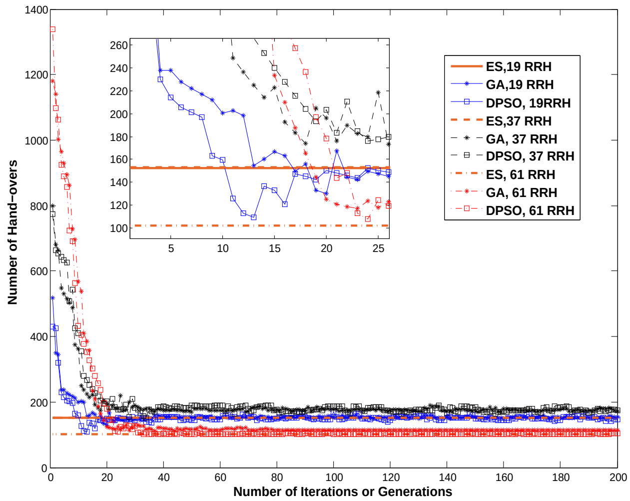

Fig. 16. Number of Handovers in P_1, P_2, and P_3 for GA and DPSO

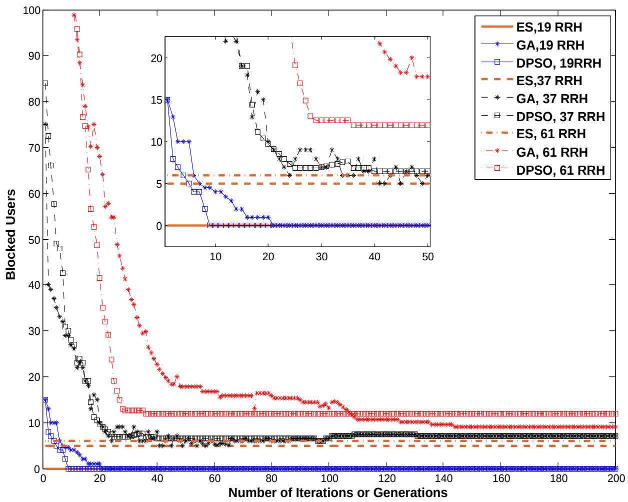

Fig. 15. Number of Blocked Users in P_1, P_2, and P_3 for GA and DPSO

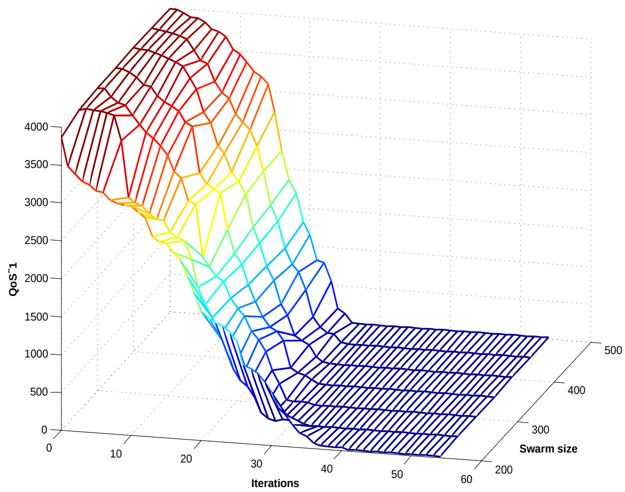

Fig. 17. Number of particles (Swarm size) Vs Number of iterations trade-off for DPSO Benchmark problem P_3

problem are tuned correctly. In the case of a vast network scenario like P_3, both algorithms can deliver more appropriate RRH-sector configurations by utilising a higher number of generations/iterations. Another approach is to increase the population size $|\Delta|$ while keeping the number of generations/iterations fixed. Depending on the available system resources, both algorithms can be adjusted accordingly, e.g., increasing the population size requires a system to have a large memory size. Moreover, with increased number of generations/iterations and limited system hardware resources, the execution time also increases due to increased evaluations at each generation/iteration. A trade-off between the number of generations/iterations and the population/swarm size is recommended to produce appropriated performance. Fig. 17 presents a trade-off between swarm size and a number of iterations for Benchmark problem P_3 (i.e., 61 RRHs) provided that the distance between RRHs is fixed as given in Fig. 4. This is resolved using DPSO based on $[\text{QoS}]^{-1}$. The DPSO and GA algorithms can be configured to the hardware available based on the trade-off. In Fig. 17, it is observed that a larger swarm size tends to achieve the optimal solution faster than a smaller swarm size, however, with limited hardware resources, larger swarm sizes often leads to more evaluation time to achieve the optimal solution.

VIII. Conclusion

To conclude all this, the dynamic RRH-sector allocations in C-RAN are examined, with an aim to improve the QoS. Proper RRH-sector mappings achieve a well-balanced traffic in the network. A self-optimised C-RAN algorithm is proposed which utilises the network resources efficiently. RRH-sector mapping is formulated as an optimisation problem, which is used for maximising the QoS of C-RAN, minimising the number of blocked users, and reducing the handovers required to make a new RRH-sector mapping transition. Two evolutionary algorithms, i.e., the GA and DPSO are utilised in the SOCRAN algorithm to solve the RRH-sector allocation problem. The performances of both GA and DPSO are compared using three benchmark problems. The DPSO delivered noticeable faster and better convergence compared to GA. Both GA and DPSO provided a near optimum solution for larger networks. However, the DPSO outperforms GA in all network scenarios. The SOCRAN architecture contributes to the development of SON by providing high levels of QoS in a time-varying traffic environment and enabling dynamic inter-cell optimisation, which is one of the important issues in SON.

REFERENCES

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

[2] E. U. T. R. A. Network, "Self-configuring and self-optimizing network (SON) use cases and solutions," *Third Generation Partnership Project (3GPP) specification TR 36.902 V9.3.1*, vol. 36, 2012.

[3] F. Ahmed, A. A. Dowhuszko, and O. Tirkkonen, "Network optimization methods for self-organization of future cellular networks: Models and algorithms," *Self-Organized Mobile Communication Technologies and Techniques for Network Optimization*, pp. 35–65, 2016.

[4] 3GPP, "Universal Mobile Telecommunication system (UMTS); LTE; Telecommunication Management; Self-Organizing Networks (SON); Concepts and requirements (Release 14) ," 3rd Generation Partnership Project (3GPP), TS 32.500 v14.0.0 , Apr. 2017.

[5] N. Dandanov, H. Al-Shatri, A. Klein, and V. Poulkov, "Dynamic Self-Optimization of the Antenna Tilt for Best Trade-off Between Coverage and Capacity in Mobile Networks," *Wireless Personal Communications*, vol. 92, no. 1, pp. 251–278, 2017.

[6] P. Muoz, R. Barco, and I. de la Bandera, "On the Potential of Handover Parameter Optimization for Self-Organizing Networks," *IEEE Transactions on Vehicular Technology*, vol. 62, no. 5, pp. 1895–1905, Jun 2013.

[7] J. M. Ruiz-Avils, M. Toril, S. Luna-Ramrez, V. Buenestado, and M. A. Regueira, "Analysis of Limitations of Mobility Load Balancing in a Live LTE System," *IEEE Wireless Communications Letters*, vol. 4, no. 4, pp. 417–420, Aug 2015.

[8] M. Huang and J. Chen, "A conflict avoidance scheme between mobility load balancing and mobility robustness optimization in self-organizing networks," *Wireless Networks*, pp. 1–11, 2016. [Online]. Available: http://dx.doi.org/10.1007/s11276-016-1331-y

[9] Y. Dhungana and C. Tellambura, "Multichannel Analysis of Cell Range Expansion and Resource Partitioning in Two-Tier Heterogeneous Cellular Networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 3, pp. 2394–2406, March 2016.

[10] O. G. Aliu, A. Imran, M. A. Imran, and B. Evans, "A survey of self organisation in future cellular networks," *IEEE Communications Surveys & Tutorials*, vol. 15, no. 1, pp. 336–361, 2013.

[11] O. Østerbø and O. Grøndalen, "Benefits of Self-Organizing Networks (SON) for mobile operators," *Journal of Computer Networks and Communications*, vol. 2012, 2012.

[12] L. Jorguseski, A. Pais, F. Gunnarsson, A. Centonza, and C. Willcock, "Self-organizing networks in 3GPP: standardization and future trends," *IEEE Communications Magazine*, vol. 52, no. 12, pp. 28–34, December 2014.

[13] C. Liu, L. Zhang, M. Zhu, J. Wang, L. Cheng, and G.-K. Chang, "A Novel Multi-Service Small-Cell Cloud Radio Access Network for Mobile Backhaul and Computing Based on Radio-Over-Fiber Technologies," *Lightwave Technology, Journal of*, vol. 31, no. 17, pp. 2869–2875, Sept 2013.

[14] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. S. Berger, and L. Dittmann, "Cloud RAN for Mobile Networks A Technology Overview," *IEEE Communications Surveys Tutorials*, vol. 17, no. 1, pp. 405–426, Firstquarter 2015.

[15] C. Chen, J. Huang, W. Jueping, Y. Wu, and G. Li, "Suggestions on Potential Solutions to C-RAN," *NGMN Alliance, Version 4*, pp. 1–41, 2013.

[16] Next Generation of Mobile networks. (2013, Feb.) Project Centralized Processing, Collaborative Radio, Real-Time Computing, Clean RAN System (P-CRAN). [Online]. Available: https://www.ngmn.org/publications/technical/browse/2.html

[17] E. Commission. (2013, Apr) Mobile Cloud Networking (MCN) Project. [Online]. Available: http://www.mobile-cloud-networking.eu/site

[18] ——. (2014, Feb) FP7 Project High Capacity Network Architecture With Remote Radio Heads and Parasitic Antenna Arrays (HARP). [Online]. Available: http://www.fp7-harp.eu

[19] ——. (2013, Sep) iJOIN: Interworking and Joint Design of an Open Access and Backhaul Network Architecture for Small Cells Based on Cloud Networks. [Online]. Available: http://www.ict-ijoin.eu/

[20] ——. (2014, Feb) Connectivity Management for Energy Optimised Wireless Dense networks (CROWD). [Online]. Available: http://www.ict-crowd.eu/publications.html

[21] A. Checko, "Cloud Radio Access Network architecture. Towards 5G mobile networks," Ph.D. dissertation, 2016.

[22] J. Wu, Z. Zhang, Y. Hong, and Y. Wen, "Cloud radio access network (C-RAN): a primer," *Network, IEEE*, vol. 29, no. 1, pp. 35–41, Jan 2015.

[23] M. Peng, Y. Sun, X. Li, Z. Mao, and C. Wang, "Recent advances in cloud radio access networks: System architectures, key techniques, and open issues," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 3, pp. 2282–2308, 2016.

[24] T. Q. Quek, M. Peng, O. Simeone, and W. Yu, *Cloud radio access networks: Principles, technologies, and applications*. Cambridge University Press, 2017.

[25] H. Venkataraman and R. Trestian, *5G Radio Access Networks: centralized RAN, cloud-RAN and virtualization of small cells*. CRC Press, 2017.

[26] C. PAN, H. Zhu, N. Gomes, and J. Wang, "Joint Precoding and RRH selection for User-centric Green MIMO C-RAN," *IEEE Transactions on Wireless Communications*, vol. PP, no. 99, pp. 1–1, 2017.

[27] K. Wang, W. Zhou, and S. Mao, "On Joint BBU/RRH Resource Allocation in Heterogeneous Cloud-RANs," *IEEE Internet of Things Journal*, vol. PP, no. 99, pp. 1–1, 2017.

[28] L. Liu and W. Yu, "Cross-Layer Design for Downlink Multihop Cloud Radio Access Networks With Network Coding," *IEEE Transactions on Signal Processing*, vol. 65, no. 7, pp. 1728–1740, April 2017.

[29] Y. S. Chen, W. L. Chiang, and M. C. Shih, "A Dynamic BBU-RRH Mapping Scheme Using Borrow-and-Lend Approach in Cloud Radio Access Networks," *IEEE Systems Journal*, vol. PP, no. 99, pp. 1–12, 2017.

[30] S. Namba, T. Warabino, and S. Kaneko, "BBU-RRH switching schemes for centralized RAN," in *Communications and Networking in China (CHINACOM), 2012 7th International ICST Conference on*. IEEE, 2012, pp. 762–766.

[31] K. Sundaresan, M. Y. Arslan, S. Singh, S. Rangarajan, and S. V. Krishnamurthy, "Fluidnet: a flexible cloud-based radio access network for small cells," *IEEE/ACM Transactions on Networking*, vol. 24, no. 2, pp. 915–928, 2016.

[32] Z. Yu, K. Wang, H. Ji, X. Li, and H. Zhang, "Dynamic resource allocation in TDD-based heterogeneous cloud radio access networks," *China Communications*, vol. 13, no. 6, pp. 1–11, 2016.

[33] K. Lin, W. Wang, Y. Zhang, and L. Peng, "Green Spectrum Assignment in Secure Cloud Radio Network with Cluster Formation," *IEEE Transactions on Sustainable Computing*, vol. PP, no. 99, pp. 1–1, 2017.

[34] W. He, J. Gong, X. Su, J. Zeng, X. Xu, and L. Xiao, "SDN-Enabled C-RAN? An Intelligent Radio Access Network Architecture," in *New Advances in Information Systems and Technologies*. Springer, 2016, pp. 311–316.

[35] C. Yang, Z. Chen, B. Xia, and J. Wang, "When ICN meets C-RAN for HetNets: an SDN approach," *IEEE Communications Magazine*, vol. 53, no. 11, pp. 118–125, November 2015.

[36] O. Simeone, A. Maeder, M. Peng, O. Sahin, and W. Yu, "Cloud radio access network: Virtualizing wireless access for dense heterogeneous systems," *Journal of Communications and Networks*, vol. 18, no. 2, pp. 135–149, April 2016.

[37] J. G. Herrera and J. F. Botero, "Resource Allocation in NFV: A Comprehensive Survey," *IEEE Transactions on Network and Service Management*, vol. 13, no. 3, pp. 518–532, Sept 2016.

[38] R. Cziva, S. Jout, D. Stapleton, F. P. Tso, and D. P. Pezaros, "SDN-Based Virtual Machine Management for Cloud Data Centers," *IEEE Transactions on Network and Service Management*, vol. 13, no. 2, pp. 212–225, June 2016.

[39] O. K. Tonguz and E. Yanmaz, "The Mathematical Theory of Dynamic Load Balancing in Cellular Networks," *IEEE Transactions on Mobile Computing*, vol. 7, no. 12, pp. 1504–1518, Dec 2008.

[40] Q. Han, B. Yang, C. Chen, and X. Guan, "Energy-aware and QoS-aware load balancing for HetNets powered by renewable energy," *Computer Networks*, vol. 94, pp. 250–262, 2016.

[41] M. Peng, C. Wang, V. Lau, and H. V. Poor, "Fronthaul-constrained cloud radio access networks: Insights and challenges," *IEEE Wireless Communications*, vol. 22, no. 2, pp. 152–160, 2015.

[42] M. Y. Arslan, K. Sundaresan, and S. Rangarajan, “Software-defined networking in cellular radio access networks: potential and challenges,” *IEEE Communications Magazine*, vol. 53, no. 1, pp. 150–156, January 2015.

[43] K. Tsagkaris, G. Poulios, P. Demestichas, A. Tall, Z. Altman, and C. Destré, “An open framework for programmable, self-managed radio access networks,” *IEEE Communications Magazine*, vol. 53, no. 7, pp. 154–161, 2015.

[44] T. T. Vu, L. Decreusefond, and P. Martins, “An Analytical Model for Evaluating Outage and Handover Probability of Cellular Wireless Networks,” *Wireless Personal Communications*, vol. 74, no. 4, pp. 1117–1127, 2014. [Online]. Available: http://dx.doi.org/10.1007/s11277-013-1567-0

[45] X. Ge, J. Ye, Y. Yang, and Q. Li, “User Mobility Evaluation for 5G Small Cell Networks Based on Individual Mobility Model,” *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 3, pp. 528–541, March 2016.

[46] 3GPP, “General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access,” 3rd Generation Partnership Project (3GPP), TS 23.401, Sep. 2016.

[47] D. Pompili, A. Hajisami, and T. X. Tran, “Elastic resource utilization framework for high capacity and energy efficiency in cloud RAN,” *IEEE Communications Magazine*, vol. 54, no. 1, pp. 26–32, January 2016.

[48] M. Srinivas and L. M. Patnaik, “Genetic algorithms: a survey,” *Computer*, vol. 27, no. 6, pp. 17–26, June 1994.

[49] Y. Zhang, S. Wang, and G. Ji, “A comprehensive survey on particle swarm optimization algorithm and its applications,” *Mathematical Problems in Engineering*, vol. 2015, pp. 1–38, 2015.

[50] B. S. Ahn, “Compatible weighting method with rank order centroid: Maximum entropy ordered weighted averaging approach,” *European Journal of Operational Research*, vol. 212, no. 3, pp. 552–559, 2011.

[51] M. Wazid and A. K. Das, “An Efficient Hybrid Anomaly Detection Scheme Using K-Means Clustering for Wireless Sensor Networks,” *Wireless Personal Communications*, vol. 90, no. 4, pp. 1971–2000, 2016. [Online]. Available: http://dx.doi.org/10.1007/s11277-016-3433-3