Uplink/Downlink Decoupled Energy Efficient User Association in Heterogeneous Cloud Radio Access Networks*

Merve Saimler[a,*], Sinem Coleri Ergen[b]

[a] *Rumelifeneri Yolu 34450, Koc University, Istanbul, TURKEY*
[b] *Rumelifeneri Yolu 34450, Koc University, Istanbul, TURKEY*

Abstract

Heterogeneous Cloud Radio Access Networks ($H-CRAN$) is a network architecture that combines Macro Base Stations (MBS)s and Small Base Stations (SBS)s with cloud infrastructures. The dense deployment of SBSs in H-CRAN is needed to provide high data rates to User Equipments (UEs) but causes high energy consumption. Unrealistic power models lead to inefficient UE association schemes in terms of energy. In this paper, we study the joint optimization of Uplink (UL) and Downlink (DL) decoupled UE association and switching on/off the SBSs in H-CRAN by incorporating a realistic power model with the objective of minimizing the power consumption in H-CRAN. The power model encompasses static and the dynamic power consumption of MBS, the static power consumption of SBS, the power consumption of transmission links to cloud infrastructure and the power consumption of UEs. The problem is transformed into Single Source Capacitated Facility Location Problem ($SSCFLP$) which is NP-Hard. We then propose a heuristic algorithm based on the use of LP relaxation and solving many-to-one assignment problem with generalized assignment problem heuristics. Extensive simulations demonstrate that the proposed heuristic algorithm performs very close to optimal and achieves significant improvements in minimizing total power consumption compared to coupled UE association algorithm and algorithms utilizing the power consumption models that do not encompass MBS dynamic power consumption and the power consumption of transmission links to cloud infrastructure for various scenarios.

Keywords: Energy Efficiency, User Association, Uplink/Downlink Decoupling

1. Introduction

5G wireless networks are expected to provide 1000-fold data rate improvement and save up to 90 percent of energy consumption compared to the present systems [1]-[3]. Heterogeneous Cloud Radio

*This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

*Corresponding author
Email addresses: msaimler13@ku.edu.tr (Merve Saimler), sergen@ku.edu.tr (Sinem Coleri Ergen)

Access Networks ($H-CRAN$) has been proposed to provide energy saving and rate improvement in 5G through the integration of Heterogeneous Networks ($HetNet$)s and Cloud Radio Access Networks ($CRAN$)s. From HetNets, H-CRANs inherit tiered deployment of MBSs and SBSs. In HetNets, MBSs provide coverage and SBSs overlaid on MBSs serve to boost capacity, accomplish desired data demand and improve energy efficiency by offloading UEs from MBSs to SBSs with the aim of reducing dynamic power consumption of MBS. However, in case of expansion of the network size with dense deployment of BSs, the power consumption of the network increases and association of UEs to BSs of different tiers results in high inter-tier interference between MBS and SBS tiers [4]. On the other hand, CRAN introduces a centralized processing entity, BBU Cloud, which computes workloads, accounts for QoS and enables optimized orchestration of the number of UEs per BS and energy consumption. In CRAN, baseband units (BBU)s are decoupled from the remote radio heads (RRH)s, BBUs are gathered into the BBU Cloud and RRHs and BBUs are connected to each other through Fronthaul (FH) links. Nonetheless, the non-ideal FH links with limited capacity and long time delay degrade performances of CRANs [5]. To better manage aforementioned challenges and to ensure a better designed network aiming to support challenging 5G technology requirements [6], H-CRAN has been proposed [3, 7]. In H-CRAN, CRAN operation mechanism is incorporated into HetNets and BSs from different tiers are connected to the BBU Cloud via FH links whereas the BBU Cloud is connected to Internet through Backhaul (BH) links. Additionally, most of the control plane signaling takes place in the BBU Cloud for coordinating the inter-tier interferences between SBSs and MBSs. The control and broadcast functionalities are shifted from SBSs to MBSs, which alleviates capacity and time delay constraints on FH links and supports the burst traffic efficiently. This configuration allows UEs to communicate via several possible sets of BSs. Such a change in the architecture in the long run may lead to deploying more BSs than the UEs served, which causes high power consumption.

Power saving strategies proposed for HetNets include smart topological designs with the goal of providing better coverage to UEs, improvement of BS hardware design, introduction of green metrics and introduction of dynamic on/off schemes according to UE distribution over time. First, in the context of smart topological designs, utilization of relays between BSs and UEs is proposed with the aim of bringing network closer to UEs [8]. In addition to relays, cell-zooming technique [9], which adjusts the cell size by adjusting the TX power according to UE traffic in the cell, is proposed to provide minimum required Quality of Service (QoS) for all UEs. Second, the BS hardware is improved by adopting energy efficient power amplifiers [10], or by replacing existing power amplifiers with new efficient devices using digital pre-distortion or envelope tracking for wide band signals [11]. Then, green metrics consist of equipment level metrics and system level metrics that evaluate the energy efficiency of a UE or BS on the basis of its energy consumption at idle or maximum load and measure the degree of proportionality between energy consumption and full/half load or idle periods of the network, respectively [12].

With controlled planning of turning on/off schemes of SBSs, the total power consumption is minimized while satisfying the QoS constraints of UEs. QoS constraints of UEs include coverage probability [13], that is defined as the probability that a user is in coverage, and area spectral efficiency[14], which is defined as the sum of the maximum average rates of UEs per unit bandwidth per unit area. Other QoS constraints are outage probability [15], bandwidth limitation of BSs [16] and minimum rate requirement [17]. Nonetheless, the MBSs are assumed to be always active to ensure the coverage and SBSs are turned on/off based on the system load. A potential deficiency of existing strategies is that overall power models for HetNets do not include capacity limit and the dynamic power consumption of MBS. Besides ensuring coverage, the UE traffic can also be accommodated by the MBS as well as the rest of on SBSs, when any of the SBSs is turned off. Therefore, load dependency and hardware dependency of MBS dynamic power consumption have remarkable impacts on the total power consumption.

Unlike HetNets that consider only dynamic on/off scheme of SBSs, in CRAN, capacity and the power consumption modes of FH links are considered in addition to RRH power consumption. The power consumption for energy efficient UE association in CRAN is modeled by using the power consumption of RRHs and FH links for different modes considering the capacity of FH links [18]-[21]. RRH power consumption and FH link power consumption are assumed to be constant in [18],[20] or stated as a function of data rate and transmit power of RRHs in [21] and [22]. Additionally, static power consumption of FH links is considered in [22], sleep mode power consumption of RRHs and FH links are included in [18]-[21], while being ignored in [19]. On the other hand, capacity of FH links is the energy bottleneck of CRANs since the number of on RRHs and FH links is determined according to the capacity of FH links and the UE traffic being carried. FH links are assumed to have sufficiently high capacity in [18], [20] and [22], constant capacity in [19] or stated as the maximum number of UEs that can be supported on each FH link in [21]. However, none of the studies consider the joint effect of capacity limitation in terms of the maximum traffic load that can be carried by the switch in FH link associated with RRH and the power consumption as a function of data rate and power consumed per bit/s by a FH link.

In H-CRAN, the challenges arisen out of the integration of HetNets and CRAN are offset by the dynamic resource and power allocation with the aim of maximizing energy efficiency in DL or in UL [4, 23] and investigating the effect of UL and DL power consumption on the capacity [5, 24, 25]. DL H-CRAN power consumption model includes the power consumption of SBSs, MBS and FH links in [24] or RRHs and FH links without incorporating the power consumption of MBSs [5] with aim of maximizing energy efficiency subject to rate [24] or inter-tier interference constraints[5]. On the other hand, UL H-CRAN power consumption model in [25] encompasses the UE power consumption aiming to optimize relay selection, network selection and power allocation jointly with the objective of maximizing energy efficiency subject to rate constraints. Nonetheless, none of the studies for energy efficient UE association in H-CRAN consider dynamic power consumption of MBS, the power consumption of FH link and capacity of FH link jointly. Also, none of them consider to optimize UL and DL EE performances, jointly.

Since UEs can have different SINR values in UL and DL, UE association to the same BS both in UL and DL may not be energy efficient [26]-[29]. Therefore, load balancing with UL and DL decoupling paves the way for energy efficient UE association by considering UL SIR for improved UE association in UL to SBSs with reduced transmit power in [30], assigning UEs to the UL and DL of different BSs independently based on UL and DL SIR values in [31] or by exploiting unused spectrum to assign a UE to a BS and allocating more than the minimum bandwidth to reduce transmit power of the UE [32]. On the other hand, SIR is not sufficient to decide on the load balancing for decoupled UL/DL scenarios, so UE association should also consider overall load of the cells and cell BH capacity [33], whereas UL and DL UE association should also consider beam-forming design and transmit power of BSs and UEs [34]. However, none of the studies above aim to improve achievable rate with the help of improved UL and DL SNR, considering FH link capacity and reduced transmit power by enabling UEs to connect to different BSs both in UL and DL.

The goal of this paper is to formulate and propose efficient solution methodology for the optimization problem of the UL/DL decoupled energy efficient UE association with the goal of minimizing UL and DL power consumption in H-CRAN while considering the accurate power consumption models, rate, capacity and effect of FH links. The main contributions of this paper are summarized as follows:

- We provide a holistic framework for energy efficient UE association incorporating UL/DL decoupling and dynamic on/off scheme for SBSs and realistic power consumption models, including the static and the dynamic power consumption of MBS, the static power consumption of SBS, the power consumption of FH links as a function of data rate and the power consumption of UEs in UL for the first time in the literature. We adopt mmWave physical layer characteristics in which we restrict the number of simultaneous transmissions by the number of RF chains [35].
- We formulate the optimization problem as a Single Source Capacitated Facility Location Problem (SSCFLP). Since SSCFLP is NP Complete, it cannot be solved to optimality in polynomial time. Therefore, we propose a heuristic algorithm denoted as Selection and Repetition based UE Association Algorithm ($SRAA$) for the formulated optimization problem. Rather than dealing with all UEs and all BSs in each run as in repeated matching algorithm (RMA) [36] and perfect

matching algorithm (PMA) [37], SRAA forms feasible UE-BS pairs with LP relaxation method and solves matching problems for each feasible UE-BS pairs repeatedly to find minimum power consuming UE-BS pairs.

- We illustrate the superiority of the SRAA to previously proposed solution methods in terms of closeness to the optimality for different networks sizes and various data rate requirements via extensive simulations.

The rest of the paper is organized as follows. Section 1 describes the system model and the assumptions used throughout the paper. The joint optimization of UE association and switching on/off the SBSs with the objective of minimizing the system power consumption has been formulated in Section 3. Section 3 proves the NP-Hardness of the optimization problem. Section 4 presents the solution methodology for the optimization problem. Simulations and performance evaluation are presented in Section 5. Finally concluding remarks are given in Section 6.

2. System Model

2.1. HCRAN Architecture

The architecture of H-CRAN is shown in Figure 1. In our model, MBS provides C-Plane, carrying control signaling, cell specific reference signals, resource allocation and traffic processing, for the whole architecture. MBS operates at LTE cellular frequencies and is equipped with $N_{RF,M}$ RF chains. Within the coverage of MBS, there are K SBSs containing $N_{RF,S}$ RF chains and operating at mmWave frequencies. SBSs are mainly used to provide high speed data transmission with D-Plane, which carries actual user traffic and executes traffic processing to satisfy QoS (without C-Plane), in hot spots. Additionally, there are N_b UEs that request connection in the coverage of MBS and are equipped with a single antenna. Each UE can be served by exactly one BS both in UL and DL, therefore UE demand is not splittable. Lastly, dynamic on/off transition of SBSs is analyzed at a certain time instant.

BSs and BBU Cloud are connected through FH links whereas BBU Cloud and Internet is connected through low-latency with sufficiently large capacity BH links. FH links are assumed to be constrained by capacity denoted by C_{fh}.

2.2. HCRAN Channel Characteristics

The channel estimation is done based on Channel State Information Reference Signal transmitted from MBS and SBSs [38]. The attenuation levels from MBS and SBS S to UE j are denoted by α_{Mj} and α_{Sj}, respectively. The attenuation levels from UE j to MBS and SBS S in UL are θ_{Mj} and θ_{Sj}, respectively. With the information gathered about attenuation levels, BBU Cloud decides on the least

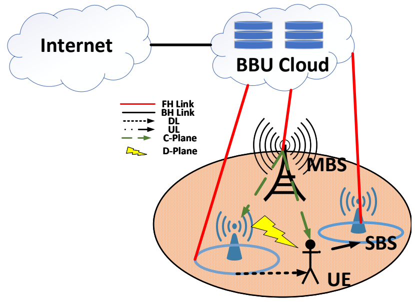


Figure 1: H-CRAN Architecture

power consuming UE-BS pair both in UL and DL for each BS and UE and establishes a connection [39].

Though directional transmission is utilized, each UE j is exposed to interference power by other BSs, that do not serve UE j. The interference power at UE j in DL, $P_{j,Idl}$, is calculated as in [40]:

$$P_{j,Idl} = \sum_{i'=1, i \neq i'}^{K+1} P_{t,i'} * \alpha_{i'j} \tag{1}$$

where i is M or S, $P_{t,M'}$ and $P_{t,S'}$ indicate the transmission power of interfering MBS and SBS S, respectively and $\alpha_{i'j}$ is the DL path loss measured at the UE j based on the DL transmit power of the interfering BSs.

Due to different transmission bandwidths and transmit powers, BSs provide different achievable rates in DL from MBS and SBS S to UE j, denoted by $R_{d,Mj}$ and $R_{d,Sj}$, respectively. The achievable rates in DL are formulated as [41]:

$$R_{d,ij} = B_i * \log\left(1 + \frac{P_{t,i} * \alpha_{ij}}{|N_{RF,i}| * (N_0 * B_i + P_{j,Idl})}\right). \tag{2}$$

where i is M or S, B_M and B_S represent the transmission bandwidth for MBS and SBS S, respectively. $P_{t,M}$ and $P_{t,S}$ indicate the transmission power of MBS and SBS S, respectively, N_0 is the power spectral density of the noise and α_{ij} is the DL path loss measured at the UE based on the DL transmit power of the reference symbols. Maximum number of UEs that can be simultaneously served by SBS S and MBS is less than or equal to the number of RF chains $N_{RF,S}$ and $N_{RF,M}$, respectively, which is motivated by the spatial multiplexing gain of the described multi-user hybrid precoding system in [35].

The power assignment for the transmission at the UE is performed to overcome γ fraction of the path loss with respect to a reference received power per resource block at 0 dB path loss, P_0, below P_{max} [42], and equal to:

$$P_{t,UE} = \min\{P_{max}, P_0 + \gamma * \alpha_{ij}\} \tag{3}$$

where i is M or S. Though directional transmission is done, each UE j is exposed to interference power. The interference power at UE j in UL, $P_{j,Iul}$, is calculated as in [40]:

$$P_{j,Iul} = \sum_{i'=1, i \neq i'}^{N_b} P_{t,UE} * \theta_{i'j} \tag{4}$$

where i is M or S and $\theta_{M'j}$ and $\theta_{S'j}$ is the UL path loss measured at the UE j based on the UL transmit power of the reference symbols for interfering MBS and SBS S, respectively.

The achievable rates in UL $R_{u,Mj}$ and $R_{u,jS}$ are formulated as in [41]:

$$R_{u,ji} = B_i * \log\left(1 + \frac{P_{t,UE} * \theta_{ij}}{N_{0i} * B_i + P_{j,Iul}}\right) \tag{5}$$

where i is M or S and, θ_{Mj} and θ_{Sj} is the UL path loss measured at the UE based on the UL transmit power of the reference symbols for MBS and SBS S, respectively. B_M and B_S represent the transmission bandwidth for MBS and SBS S, respectively. N_{0M} and N_{0S} is the power spectral density of the noise at each SBS S and MBS, respectively.

$Q_{d,j}$ and $Q_{u,j}$ are demanded data rate of UE j in UL and DL, respectively.

The channel utilization is defined as the ratio of the demanded data rate to the data rate that can be supported over the channel. The channel utilization from MBS and SBS S to UE j in DL are denoted by $\beta_{d,Mj}$ and $\beta_{d,Sj}$, respectively, and formulated as $\beta_{d,ij} = \frac{Q_{d,j}}{R_{d,ij}}$, where i is M or S as in [43]. On the other hand, the channel utilization from MBS and SBS S to UE j in UL are denoted by $\beta_{u,jM}$ and $\beta_{u,jS}$, respectively, and formulated as $\beta_{u,ji} = \frac{Q_{u,j}}{R_{u,ji}}$ where i is M or S as in [43].

2.3. Power Consumption Models of Network Components

MBS power consumption, P_M, includes both the static power consumption P_{Static} and the dynamic power consumption $P_{Dyn,M}$ [44]. P_{Static} consists of the power consumption of rectifier, air conditioner and BH link [45]. $P_{Dyn,M}$ is formulated as $F * P_l$, where $F = \frac{\sum_{j=1}^{N_b} \beta_{d,Mj} + \beta_{u,jM}}{N_{rfM}}$ is the load factor that represents the ratio of the sum of channel utilization of the active users both in UL and DL to the total capacity of MBS in terms of RF chains, and P_l consists of the power consumption of power amplifier, the transceiver, the digital signal processing unit.

Each SBS S with coverage range less than 30 meters and output power around 20 dBm is assumed to consume constant power, P_S, across all traffic load since SBS S has a limited service capacity (4 to 16 UEs) and can easily get overloaded. P_S includes power consumed in microprocessor P_{mp}, in FPGA P_{FPGA}, in transceiver $P_{tx,S}$ and in power amplifier $P_{amp,S}$ as in [44]:

$$P_S = P_{mp} + P_{FPGA} + P_{tx,S} + P_{amp,S} \tag{6}$$

Due to decoupled UL/DL transmission, the power consumed on FH link in DL is denoted as $P_{d,fhl}$ and is a linear function of the demanded data rate of UE j in DL, $Q_{d,j}$ [46], whereas the power consumed on FH link in UL is denoted as $P_{u,fhl}$ and is a linear function of the demanded data rate of UE j in UL, $Q_{u,j}$.

The power consumption of a UE, $P_{u,j}$, consists of circuit power consumption $P_{circuit}$ and power consumed in the power amplifier P_{amp}, which is formulated as $\frac{P_{t,UE}}{\rho}$ in [47], where ρ is power amplifier efficiency:

$$P_{u,j} \quad = \quad P_{circuit} + P_{amp} \tag{7}$$

The power consumed on FH link is denoted as P_{fhl} and it is a linear function of the demanded data rate of UE j, Q_j [46].

3. Problem Formulation

The joint optimization of UE association and switching on/off the SBSs with the objective of minimizing the system power consumption by incorporating the dynamic power consumption of MBS, FH link power consumption and the UE power consumption given the capacity and rate constraints is formulated as follows:

$$\begin{aligned}
\min \quad & P_{static} + \sum_{j=1}^{N_b}(\beta_{d,Mj} * P_l * x_{Mj}) + P_{fhl} * \sum_{j=1}^{N_b} Q_{d,j} * x_{Mj} + \sum_{S=1}^{K} P_S * T_S + \sum_{S=1}^{K} P_{fhl} * \sum_{j=1}^{N_b} Q_{d,j} * x_{Sj} \\
& + w * \left(P_{fhl} * \sum_{j=1}^{N_b} Q_{u,j} * y_{Mj} \right) + w * \left(P_{fhl} * \Big(\sum_{S=1}^{K} \sum_{j=1}^{N_b} Q_{u,j} * y_{Sj} \Big) \right) + w * \left(\sum_{j=1}^{N_b} (\beta_{u,jM} * P_l * y_{Mj}) \right) + \\
& w * \left(\sum_{j=1}^{N_b} (P_{u,j} * \beta_{u,jM} * y_{Mj}) \right) + w * \left(\sum_{j=1}^{N_b} P_{u,j} \sum_{s=1}^{K} \beta_{u,jS} * y_{Sj} \right) && \text{(1a)} \\
\text{s.t.} \quad & Q_{d,j} * x_{Sj} \le R_{d,Sj}, \ j \in \{1, N_b\}, \ S \in \{1, K\} && \text{(1b)} \\
& Q_{d,j} * x_{Mj} \le R_{d,Mj}, \ j \in \{1, N_b\} && \text{(1c)} \\
& Q_{u,j} * y_{Sj} \le R_{u,jS}, \ j \in \{1, N_b\}, \ S \in \{1, K\} && \text{(1d)} \\
& Q_{u,j} * y_{Mj} \le R_{u,jM}, \ j \in \{1, N_b\} && \text{(1e)} \\
& x_{Mj} + \sum_{S=1}^{K} x_{Sj} = 1, \ j \in \{1, N_b\} && \text{(1f)} \\
& y_{Mj} + \sum_{S=1}^{K} y_{Sj} = 1, \ j \in \{1, N_b\} && \text{(1g)} \\
& \sum_{j=1}^{N_b} \beta_{d,Sj} * x_{Sj} + \sum_{j=1}^{N_b} \beta_{u,jS} * y_{Sj} \le N_{RF,S}, \ S \in \{1, K\} && \text{(1h)} \\
& \sum_{j=1}^{N_b} \beta_{d,Mj} * x_{Mj} + \sum_{j=1}^{N_b} \beta_{u,jM} * y_{Mj} \le N_{RF,M} && \text{(1i)} \\
& \sum_{j=1}^{N_b} Q_{d,j} * x_{Sj} + \sum_{j=1}^{N_b} Q_{u,j} * y_{Sj} \le C_{fh}, \ S \in \{1, K\} && \text{(1j)} \\
& \sum_{j=1}^{N_b} Q_{d,j} * x_{Mj} + \sum_{j=1}^{N_b} Q_{u,j} * y_{Mj} \le C_{fh} && \text{(1k)} \\
& x_{Sj} \le T_S, \ j \in \{1, N_b\}, \ S \in \{1, K\} && \text{(1l)} \\
& y_{Sj} \le T_S, \ j \in \{1, N_b\}, \ S \in \{1, K\} && \text{(1m)} \\
\text{Vs} \quad & T_S, x_{Mj}, x_{Sj}, y_{Mj}, y_{Sj} \in \{0, 1\}, \ \ j \in \{1, N_b\}, \ S \in \{1, K\}
\end{aligned}$$

The decision variables of the problem are T_S, which is a binary variable taking value 1 when SBS S turns on and 0 otherwise; x_{Mj} and x_{Sj}, which are binary variables taking values 1 in case of association of UE j to MBS and SBS S in DL, respectively and 0 otherwise, y_{Mj} and y_{Sj}, which are binary variables taking values 1 in case of association of UE j to MBS and SBS S in UL, respectively and 0 otherwise.

Objective function (1a) of the optimization problem is minimization of the total power consumption. It includes the static power consumption of MBS, the dynamic power consumption of MBS in DL, the opening cost of SBS S, FH link power consumption in UL and DL, the dynamic power consumption of MBS in UL, the UE power consumption in case of association with MBS in UL and the UE power consumption in case of association to SBS S. Weight (w) emphasizes the contribution of the power consumption in UL to the total power consumption. Constraints (1b) and (1c) ensure that the data rate demand of UE j in DL should be less than or equal to the achievable rate from SBS S and MBS to UE j, respectively. Constraints (1d) and (1e) ensure that the data rate demand of UE j in UL should be less than or equal to the achievable rate at SBS S and MBS from UE j, respectively. Constraints (1f) and (1g) assure that UE j can only be assigned to MBS or SBS S in UL and DL, respectively. Moreover, constraints (1h) and (1i) indicate that maximum number of UEs that can be simultaneously served by a BS in UL and DL is less than or equal to the number of RF chains that belongs to SBS, ($N_{RF,S}$), or to MBS, ($N_{RF,M}$), respectively. Constraints (1j) and (1k) imply that the total data being forwarded on the FH link by the assigned UEs both in UL and DL can not exceed the total FH link capacity. Constraints (1l) and (1m) show that SBS S is turned on if there exists an assigned UE in DL or UL.

Theorem 1. *The minimization problem formulated in Section 3 is an NP-Complete SSCFLP [36], which is a special case of Capacitated Facility Location Problem ($CFLP$).*

Proof. We reduce NP-Complete SSCFLP [36] proposed in Section 3 to set cover problem [48]. Given a set of j UEs, i BSs, the set cover problem aims to minimize the cost weighted average distance between UE and the nearest of the selected BSs.

First, we assume that the UEs can be served by more than one BS in UL and DL or, which relaxes variables x_{Mj}, y_{Mj}, x_{Sj} y_{Sj} and T_S to be continuous. In this case, our optimization problem is reduced to CFLP [49].

The CFLP is NP-Complete as a generalization of the NP-complete set cover problem [48], where $N_{rf,M} = +\infty$, $N_{rf,S} = +\infty$ and $C_{fh} = +\infty$. Therefore, the SSCFLP studied in this paper is NP-Complete. □

4. Solution Methodology

We propose a polynomial time heuristic algorithm to solve $SSCFLP$. The objective of SRAA is to minimize the total power consumption by improving the output of the primal heuristic composed of LP-Relaxation and rounding process with the help of repeated use of two heuristic algorithms, heuristic for modified matching and heuristic for re-allocation of UEs. In previously proposed heuristic algorithms such as RMA [36] and PMA [37], all UEs as unassigned UEs and all BSs as off BSs are taken as input to heuristic algorithms. In contrast, in SRAA, as a result of LP relaxation and rounding, we obtain a set of assigned UE-BS pairs, unassigned UEs and off BSs as the input to the proceeding heuristic algorithms. This reduces the complexity of the algorithm, enables comprehensive and fast search for the minimum and allows the algorithm perform close to optimal.

4.1. Selection and Repetition based UE Association Algorithm

The flowchart of the SRAA is given in Figure 2. The algorithm starts with the initialization of Split Number (SN), S_{max}, the Local Value (LV), the Main Value (MV) and the Reference Value (RV). SN counts how many times a UE-BS pair is split into unassigned UEs and turned off BS, S_{max} is the maximum number of allowed split process, LV is the power consumption value obtained from the heuristic algorithm for modified matching or re-allocation, MV is the power consumption value of the SRAA and RV is the sum of the power consumption value obtained from LP relaxation and rounding and the power consumption value of assigning unassigned UEs to random on BSs.

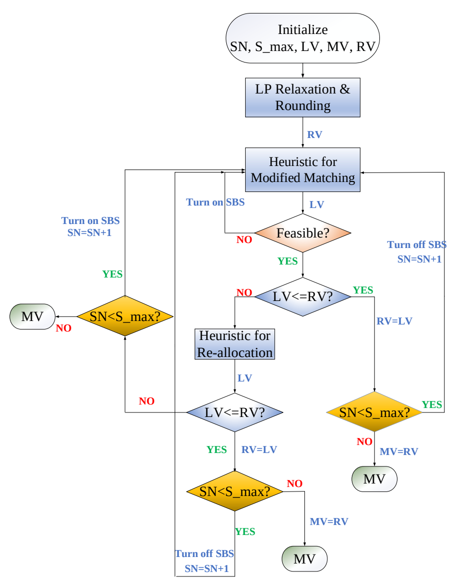


Figure 2: General Structure of the SRAA

First, LP Relaxation of the binary integer problem in Equation (1a) is solved, and RV is set to value of LP relaxation and rounding. After rounding process, we obtain a set of assigned UE-BS pairs, unassigned UEs and off BSs. Then, the set is the input to the heuristic for modified matching, where a heuristic for Generalized Assignment Problem (GAP) is solved to assign unassigned UEs to on BSs. If the on BSs are not able to provide service to unassigned UEs, then the solution of the heuristic is infeasible, the BS on the top of off BS list, where off BSs are listed in capacity increasing order according to sum of β values in UL and DL of unassigned UEs, where each sum is calculated for each BS and all unassigned UEs present in the network, is turned on and heuristic for modified matching is run again. Otherwise, LV is set to the power consumption of the output of the heuristic for modified matching. If the power consumption value obtained from the heuristic for modified matching is less than RV, then SRAA turns off the BS on top of on BS list, where on BSs are listed in capacity increasing order according to their residual capacities, which are obtained by subtracting the sum of β values of assigned UEs in UL and DL from the total capacity values. The UEs assigned to the turned off BS become unassigned and heuristic for modified matching is run again to check if the power consumption can be further decreased by turning off an on BS as long as SN is less than S_{max}.

If the heuristic for modified matching does not provide decrease in the power consumption, SRAA checks if the UEs of the assigned UE-BS pairs can be re-allocated to further minimize total power consumption by using the heuristic for re-allocation of UEs. If UE re-allocation provides further decrease in total the power consumption, the BS on top of on BS list is turned off, the UEs assigned to the turned off BS become unassigned and heuristic for modified matching is run again to check if the power consumption can be further decreased as long as SN is less than S_{max}. Otherwise, the BS on top of off BS list is turned on and heuristic for modified matching is run again to check if the power consumption can be further decreased as long as SN is less than S_{max}.

The selection process to find the least power consuming BSs out of n BSs consumes $\mathcal{O}(\log n)$ time, since we divide n BSs into two sets, which contain on BSs and off SBSs and each set is sorted according to capacity values. Instead of checking every BS to find the least power consuming BSs, we simply divide-and-conquer by looking based on where each BS stands in the list with respect to their capacity value and we select the least power consuming out of on BS list. We associate each of m UE with the on BSs, where the association of each UE consumes $\mathcal{O}(m)$ time. Therefore, the complexity of the algorithm is $\mathcal{O}(m \log n)$.

The detailed description of the main functions of the SRAA algorithm is stated as follows:

4.1.1. LP Relaxation & Rounding

LP Relaxation of the binary integer optimization problem in Equation (1a), is solved, rounded and set of assigned UE-BS pairs, unassigned UEs and off BSs is obtained as an output. However, we encounter infeasibility issues related to violation of some constraints of the optimization problem in Equation (1a). Detailed explanations on how we overcome infeasibility problems and obtain this set are as follows:

- Values of all decision variables are equal to 0.5. With rounding process, all values are rounded up to 1 revealing each UE is assigned to two BSs. In this case, the constraints (1f) and (1g) are violated. To prevent one UE-multiple BS association, we calculate the power consumption value of association of UE with each BS by setting the corresponding association values to 1 and evaluating the additional contribution in the objective function of the optimization problem (1). The BS providing a lower power consumption value is set to 1 while the other is set to 0. However, the input to the following heuristic for modified matching should contain at least one unassigned UE, one on BS and one UE-BS pair. Therefore, the BS on top of on BS list is turned off and the UEs associated with it become unassigned. We gather unassigned UEs and UE-BS pairs as input to heuristic for modified matching.

- Values of all decision variables are less than 0.5 and rounded down to 0 indicating all UEs are unassigned and all BSs are off. However, the input to the following heuristic for modified matching should contain at least one unassigned UE, one on BS and one UE-BS pair. Therefore, the BS on top of off BS list is turned on. We obtain unassigned UEs and an on BS as input to heuristic for modified matching.

- Values of decision variables have such values that after rounding the UE-BS associations are feasible. In this case, the BS on top of on BS list is turned off and the UEs associated with it become unassigned. We get unassigned UEs and UE-BS pairs as input to heuristic for modified matching.

4.1.2. Modified Matching Problem

The Modified Matching Problem aims to minimize the system power consumption by assigning unassigned UEs to UE-BS pairs or on BSs without turning on additional BSs. Unlike the optimization problem 1, which takes only unassigned UEs and off SBSs as input, Modified Matching Problem takes unassigned UEs, UE-BS pairs and on BSs as input. Moreover, the capacity constraints at SBS sb, MBS and FH link are updated as $N_{RF,sb} - C_{sb}$, $N_{RF,M} - C_M$, $C_{fh,sb} - N_{sb}$ and $C_{fh,M} - N_M$ respectively where C_{sb}, C_M, N_{sb} and N_M are the capacities allocated to assigned UEs at SBS sb, MBS, FH link of SBS and MBS, respectively. The optimization problem is formulated as:

$$
\begin{aligned}
\min \quad & \sum_{sb=1}^{R}\sum_{j=1}^{P}((Q^{t}_{d,j} * P_{fhl}) * z_{sb,j} + P_i) + \sum_{sb=1}^{R} P_S + \sum_{j=1}^{P}(\beta_{d,Mj} * P_l) * z_{Mj}) + P_{static} + \sum_{j=1}^{P}(P_{fhl} * Q_{d,j}) * z_{Mj} \\
& w * \sum_{sb=1}^{R}\sum_{j=1}^{P}(P_{u,j} \sum_{sb=1}^{R} \beta_{u(j,sb)} + (Q_{u,j} * P_{fhl})) * v_{sb,j} + w * \sum_{j=1}^{P} \Big((P_{u,j} + P_l) * \beta_{u,jM}\Big) * v_{Mj} \qquad \text{(2a)}
\end{aligned}
$$

$$
\begin{aligned}
\text{s.t.} \quad & \sum_{j=1}^{P} \beta_{d,(sb,j)} * z_{sb,j} + \sum_{j=1}^{P} \beta_{u(j,sb)} * v_{sb,j} \leq N_{RF,sb} - C_{sb}, \ sb \in \{1, R\} && \text{(2b)} \\
& \sum_{j=1}^{P} \beta_{d,Mj} * z_{Mj} + \sum_{j=1}^{P} \beta_{u,jM} * v_{Mj} \leq N_{RF,M} - C_M && \text{(2c)} \\
& \sum_{j=1}^{P} Q_{d,j} * z_{sb,j} + \sum_{j=1}^{P} Q_{u,j} * v_{sb,j} \leq C_{fh,sb} - N_{sb}, \ sb \in \{1, R\} && \text{(2d)} \\
& \sum_{j=1}^{P} Q_{d,j} * z_{Mj} + \sum_{j=1}^{P} Q_{u,j} * v_{Mj} \leq C_{fh,M} - N_M, && \text{(2e)} \\
& z_{Mj} + \sum_{sb=1}^{R} z_{sb,j} = 1, \ \ j \in \{1, P\} && \text{(2f)} \\
& v_{Mj} + \sum_{sb=1}^{R} v_{sb,j} = 1, \ \ j \in \{1, P\} && \text{(2g)} \\
\text{Vs} \quad & z_{sb,j}, v_{sb,j}, z_{Mj}, v_{Mj} \in \{0, 1\}, \ \ j \in \{1, P\}, \ sb \in \{1, R\}
\end{aligned}
$$

where R is the number of on SBSs and P is the number of unassigned UEs. The decision variables of the problem are $z_{sb,j}$ and $v_{sb,j}$, which are binary variables taking values 1, when an on SBS sb and an unassigned UE j gets associated with each other in UL and DL, respectively and 0 otherwise; z_{Mj} and v_{Mj}, which are binary variables taking values 1 when MBS and an unassigned UE j gets associated with each other in UL and DL, respectively and 0 otherwise.

Different from the binary integer optimization problem with the objective function, Equation (1a), objective function Equation (2a) aims to minimize the power consumption by including the power consumption of UE-BS pairs, on BSs and association of unassigned UEs to on BSs or UE-BS pairs. Additionally, unlike the capacity constraints in Equations (1h), (1i) and (1j), capacity constraints in Equations (2b), (2c), (2d) and (2e) provide less capacity due to capacity reserved for assigned UEs at SBS sb, MBS and FH link of SBS sb and MBS, respectively.

The optimization problem in Section 4.1.2 has the form GAP. GAP is NP-Hard and due to its NP-Hard nature, it cannot be solved to optimality in polynomial time [50]. We utilize the heuristic algorithm, which aims to find a subset of UEs that has a feasible association with a BS and assigns all UEs with BSs, such that the power consumption is minimized, to solve the problem [51]. First of all, sum of β_d and β_u values are calculated for each UE-SBS and UE-MBS pair, stored in a capacity matrix. Similarly, the power consumption values of UE-SBS and UE-MBS pairs are calculated and stored in a profit function matrix, where columns represent BSs and rows represent UEs. Then, a BS is turned on randomly and the UEs, whose sum of β_d and β_u values do not exceed the capacity limit of corresponding BS are selected to get associated with the on BS. Accordingly, profit function matrix is decomposed into two sub-matrices. In the first sub-matrix, the rows of the selected UEs are assigned to the power consumption value of the selected UE-BS pair, the unselected UE-on SBS row-column pairs save their values and unselected UEs-off BSs row-column pairs are assigned to zero. The second sub-matrix is formed by subtracting the first sub-matrix from the profit function matrix. The columns, which are all composed of zeros, are eliminated and the matrix becomes the new profit function matrix. This decomposition and elimination processes continue until there is one column, which is all composed of zeros. The UEs associated with BSs are recorded and not allowed to re-associate in ongoing decomposition and elimination processes. The solution is the sum of the minimum power consumption values of each pair and a valid association of UE-BS pair.

4.1.3. Heuristic for re-allocation of UE-BS pairs

Heuristic for re-allocation of UE-BS pairs takes the outputs of the modified matching heuristic, which are UE-BS pairs and the total power consumption value, as inputs and checks if re-allocation of UE-BS pairs further minimizes the total power consumption.

The heuristic algorithm proceeds as follows: β values of UEs for each on BS are calculated. First, the BS on the top of on BS list is selected and the UEs associated with the selected BS are exchanged with the UEs associated with the remaining BSs in the on BS list one by one. In each exchange process, the capacity requirement is checked and corresponding power consumption is saved. The power consumption value of the heuristic is updated with the minimum power consumption value obtained in whole exchange process for a UE and UE-BS allocation is updated with UE-BS association, which provides the minimum power consumption. Then, this exchange process is repeated for all on BSs in on BS list and their associated UEs without changing the association of previously exchanged UEs.

We present an example to clarify the process with Figure 3. First, MV, LV, RV and SN are set to zero and S_{max} is set to 2. DL transmission is represented with straight line whereas UL transmission is represented with dashed line. On BSs and off BSs are listed in capacity increasing order according to their residual capacity values. Then, by using the output of the LP relaxation and rounding in Section 4.1.1, network is formed as in Figure 3-1, with three UE-BS pairs (green circles-red triangles), two unassigned UEs (black triangles). RV is updated with the solution of the LP relaxation of the binary integer optimization problem and rounding.

We run heuristic for modified matching. We assume that capacity of on BSs in Figure 3-1 is not sufficient to provide connection to UEs. Therefore the association between unassigned UEs and UE-BS pairs is infeasible. In this case LV does not change, we turn on SBS 3, which is on top of off BS. Heuristic for modified matching is run again, unassigned UEs get associated with on SBS 3, the

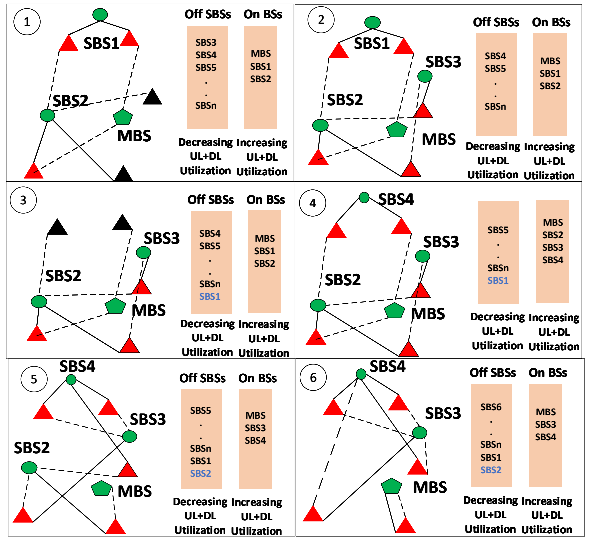


Figure 3: Heuristic Example. Green circle represents on SBSs and green pentagon represents on MBS. Red triangle is assigned UE, whereas black triangle is unassigned UE.

association is feasible and LV is set to the power consumption value of UE-BS association in Figure 3-2. Next, we assume that LV is less than RV. RV is set to LV, SN is less than 2, SBS 1 is turned off to check if less number of on BSs can provide service to UEs and appended to the end of off BS list. The UEs assigned with SBS 1 become unassigned and SN becomes 1 in Figure 3-3. Heuristic for modified matching is run again, the association is assumed to be infeasible, LV does not change and SBS 4 is turned on. The unassigned UEs get associated with SBS 4 and LV is set to the power consumption value of UE-BS association in Figure 3-4. We assume that LV is greater than RV, therefore RV does not change. Heuristic for re-allocation is run, LV is set to the power consumption value of UE-BS association in Figure 3-5. In the event, where we consider LV to be less than RV, RV is set to LV, SN is less than 2, SBS 2 is turned off and SN becomes 2 in Figure 3-6. Heuristic for modified matching is run again, LV is set to the power consumption value of UE-BS association in Figure 3-6, the output is feasible. LV is assumed to be less than RV, RV is set to LV but SN is equal to 2, which is S_{max}. Therefore, the algorithm stops and returns RV.

4.2. Algorithm Analysis

Energy efficient UE association is highly dependent on the power consumption and $Q_{d,j}$ and $Q_{u,j}$ threshold values are the most important decision factors which enable UEs to associate with BSs that minimizes the total power consumption. Therefore, in this section, we derive and compare the regions for $Q_{d,j}$ and $Q_{u,j}$ threshold values of UE association cases for the optimal algorithm, SRAA and RMA.

RMA is explained as follows:

- RMA uses the repeated matching approach to form new feasible UE-BS pairs from previously determined ones with the aim of minimizing the total power consumption and the power consumption of each UE-BS association is scaled by cost values, which are obtained by normalizing the distance between each UE-BS association with the minimum distance of UE-BS association attained in the network [36]. The process starts by taking any feasible UE-BS pair and then continues by forming sequence of feasible UE-BS pairs of decreasing cost. When repeated matching fails to yield any further reduction in the power consumption, the UE-BS pair is split so that UEs previously assigned to certain BSs become unassigned. Before the splitting, improved solution is found by applying a heuristic based on solving a GAP, which aims to find improved re-allocation of the UEs by considering all on and off BSs present in the network. After the splitting, a new sequence of feasible UE-BS pairs is generated to facilitate further progress. This process is then repeated until no further progress can be made in a fixed number of splits. The complexity of the algorithm is $\mathcal{O}(mn)$, where n is the number of BSs and m is the number of UEs.

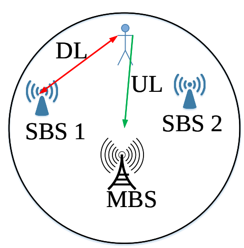


Figure 4: Network Topology

We consider a network, where the UE is located at an equal distance to SBSs and its distance to MBS is larger than its distance to SBSs as depicted in Figure 4. According to the network proposed in Figure 4, there exists nine different association cases for a UE to associate with any BS in UL and DL as follows:

1. Association to SBS 1 in DL and MBS in UL
2. Association to SBS 1 in DL and SBS 2 in UL
3. Association to SBS 1 in DL and SBS 1 in UL
4. Association to SBS 2 in DL and MBS in UL
5. Association to SBS 2 in DL and SBS 1 in UL

6. Association to SBS 2 in DL and SBS 2 in UL
7. Association to MBS in DL and SBS 1 in UL
8. Association to MBS in DL and SBS 2 in UL
9. Association to SBS 2 in DL and SBS 2 in UL

Energy efficient UE association for the optimal algorithm is determined based on the comparison of the power consumption values for UE-MBS and UE-SBS association for both UL and DL in objective function (1a). Then, UE associates with the BS which causes less power consumption by using the formulation of the minimum power consuming association in objective function (1a), the specific ranges are derived for $Q_{d,j}$ and $Q_{u,j}$.

To exemplify the derivation, we assume that UE gets associated with the SBS S in DL and MBS in UL. Detailed explanations on how we derive these values are as follows:

- Since we assume that UE gets associated with the SBS S in DL, the energy consumption corresponding to UE-SBS association is smaller than that of the UE-MBS association in Equation (1a). The inequality proving the statement is then stated as follows:

$$P_S + P_{fhl} * Q_{d,1} < P_l * \beta_{d,M1} + P_{fhl} * Q_{d,1}. \tag{3}$$

- $Q_{d,1}$ value should satisfy the inequality (3) as follows:

$$Q_{d,1} > R_{d,M1} * \left(\frac{P_S}{P_l} \right) \tag{4}$$

- Since there exists no additional UE associated with the SBS S in DL, by taking capacity constraint (1h) into consideration $Q_{d,1}$ takes the maximum value that SBS S can provide to a UE as follows:

$$Q_{d,1} \leq N_{rf,S} * R_{d,S1} \tag{5}$$

- To sum up, $Q_{d,1}$ value for the optimal algorithm should be between the $Q_{d,1}$ values stated in Equations (4) and (5).

- If there exists more than one SBS in the network, then in order to decide on SBS 1, $\beta_{d,11}$ should be less than the rest $\beta_{d,S1}$ values, which guarantees service to more UEs and forces $R_{d,11}$ to be greater than rest $R_{d,S1}$ values.

$$\beta_{d,11} < \beta_{d,S1}, \ \forall S \in \{2, K\} \tag{6}$$

- In order for a UE to get associated with MBS in UL, the power consumed in UL due to MBS transmission should be less than the power consumed in UL due to SBS transmission. By using the objective function (1a), the inequality proving the statement is stated as follows:

$$P_S + P_{fhl} * Q_{u,1} + \beta_{u,1S} * P_{u,1} > \beta_{u,1M} * \Big(P_l + P_{u,1}\Big) + P_{fhl} * Q_{u,1} \tag{7}$$

- $Q_{u,1}$ value supporting the inequality (7), which is based on the power consumption comparison, is derived as follows:

$$Q_{u,1} < P_S * \left(\frac{P_l + P_{u,1}}{R_{u,1M}} - \frac{P_{u,1}}{R_{u,1S}}\right)^{-1} \tag{8}$$

- Since there exists capacity constraints, comparison of the power consumption values is not sufficient to decide on the appropriate $Q_{u,1}$ value. Therefore, capacity constraints (1h) and (1i) are taken into consideration. There exists no additional UE associated with the MBS in DL and $Q_{u,1}$, which is based on capacity requirements of MBS takes the maximum value that MBS can provide to a UE and it is stated as follows:

$$Q_{u,1} \leq N_{rf,M} * R_{u,1M} \tag{9}$$

  However, UE is associated with SBS S in DL, therefore the maximum data rate SBS S can provide to UE is proportional to the residual capacity at SBS S, which is

$$Q_{u,1} \leq (N_{rf,S} - C_1) * R_{u,1S} \tag{10}$$

- To sum up, $Q_{u,1}$ value for the optimal algorithm should be equal to the minimum of the $Q_{u,1}$ values stated in Equations (8), (9) and (10).

$Q_{d,j}$ and $Q_{u,j}$ threshold values for UE association in SRAA are also derived based on the power consumption values in UL and DL in objective function 2a, respectively. Since there exists one UE in the network and it is assumed to be associated with SBS 1 in DL and MBS in UL, capacity reserved for UE 1 at MBS, C_{M1}, and at the rest of SBSs, C_{S1}, are zero. Therefore, $Q_{d,j}$ and $Q_{u,j}$ threshold values of SRAA are equal to data rate threshold values for the optimal algorithm for $C_{M1} = 0$ and $C_{S1} = 0$.

$Q_{d,j}$ and $Q_{u,j}$ threshold values for UE association in RMA are derived based on the power consumption values in UL and DL by a small difference from SRAA. Unlike SRAA, in RMA, the power consumption of each UE-BS association is scaled by cost values, c_{Sj} or c_{Mj}, which are formulated as $c_{Sj} = \frac{h_{Sj}}{h_{min}}$ and $c_{Mj} = \frac{h_{Mj}}{h_{min}}$, where h_{Mj} corresponds to the distance between MBS and UE j, h_{Sj} is the distance between SBS S and UE j and h_{min} is the minimum distance in the network.

- Since we assume that UE gets associated with the SBS S in DL, the energy consumption corresponding to UE-SBS association is smaller than that of the UE-MBS association in the modified version of objective function (1a), which is modified with c_{Mj} and c_{Sj} scaled the power consumption values of UE-BS association. The inequality proving the statement is stated as follows:

$$P_S + P_{fhl} * c_{S1} * Q_{d,1} < P_l * c_{M1} * \beta_{d,M1} + P_{fhl} * c_{M1} * Q_{d,1} \tag{11}$$

- $Q_{d,1}$ value in RMA should satisfy the inequality (11) as follows:

$$Q_{d,1} > P_S * \left(c_{M1} * \Big(\frac{P_l}{R_{d,M1}} + P_{fhl} \Big) - c_{S1} * P_{fhl} \right)^{-1} \tag{12}$$

- To sum up, $Q_{d,1}$ value for RMA should be between $Q_{d,1}$ values stated in Equations (11) and (12).

- Since there exists no additional UE associated with the SBS S in DL, by taking capacity constraint (1h) into consideration $Q_{d,1}$ takes the maximum value that SBS S can provide to a UE as follows:

$$Q_{d,1} \leq N_{rf,S} * R_{d,S1} \tag{13}$$

- If there exists more than one SBS in the network, then in order to decide on SBS 1 in RMA, $\beta_{d,11}$ scaled with c_{11} should be less than the rest $\beta_{d,S1}$ scaled with c_{S1} values, which guarantees service to more UEs and forces $R_{d,11}$ to be greater than rest $R_{d,S1}$ values.

$$c_{11} * \beta_{d,11} < c_{S1} * \beta_{d,S1}, \; \forall S \in \{2, K\} \tag{14}$$

- In order for a UE to get associated with MBS in UL, the power consumed in UL due to MBS transmission should be less than the power consumed in UL due to SBS transmission. By taking the scaled version of the objective function, Equation (1a) into consideration, the inequality proving the statement is stated as follows:

$$P_S + P_{fhl} * Q_{u,1} * c_{S1} + \beta_{u,1S} * P_{u,1} * c_{S1} > \beta_{u,1M} * c_{M1} \Big(P_l + P_{u,1} \Big) \tag{15}$$

- $Q_{u,1}$ value supporting the inequality (15), which is based on the power consumption comparison, is derived as follows:

$$Q_{u,1} < P_S * \left(\Big(\frac{P_l + P_{u,1}}{R_{u,1M}} + P_{fhl} \Big) * c_{M1} - P_{fhl} * c_{S1} - \frac{P_{u,1} * c_{S1}}{R_{u,1S}} \right)^{-1} \tag{16}$$

- Since there exists capacity constraints, comparison of the power consumption values is not sufficient to decide on the appropriate $Q_{u,1}$ value. Therefore, capacity constraints (1h) and (1i) are taken into consideration. There exists no additional UE associated with the MBS in DL, $Q_{u,1}$ takes the maximum value that MBS can provide to a UE and it is stated as follows:

$$Q_{u,1} \leq N_{rf,M} * R_{u,1M} \tag{17}$$

However, UE is associated with SBS S in DL, therefore the maximum data rate SBS S can provide to UE is proportional to the residual capacity at SBS S, which is

$$Q_{u,1} \leq (N_{rf,S} - C_1) * R_{u,1S} \tag{18}$$

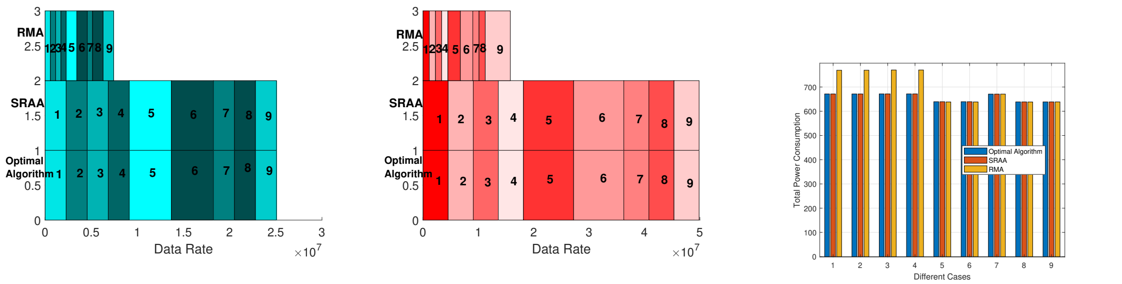
- To sum up, $Q_{u,1}$ value for RMA should be equal to the minimum of the $Q_{u,1}$ values stated in Equations (16), (17) and (18).

Based on the derivations explained above, for the network proposed in Figure 4, we demonstrate $Q_{d,j}$ and $Q_{u,j}$ values for the first case for the optimal algorithm, SRAA and RMA in Table 1. The remaining cases are derived similarly.

Optimal Algorithm	(a)	**(i)** $Q_{d,1} > R_{d,M1} * \left(\frac{P_S}{P_l}\right)$
		(ii) $Q_{d,1} \leq N_{rf,1} * R_{d,11}$
		(iii) $\beta_{d,11} < \beta_{d,21}$
	(b)	**(i)** $Q_{u,1} < min\Bigg((N_{RF,M} * R_{u,1M}), \big((N_{rf,1} - C_1) * R_{u,11}\big), Q_{u,1} < P_1 * \left(\frac{P_l + P_{u,1}}{R_{u,M}} - \frac{P_{u,1}}{R_{u,11}}\right)^{-1}$
		(ii) $Q_{u,1} < min\left((N_{RF,M} * R_{u,1M}), \big((N_{rf,2} - C_2) * R_{u,12}\big), P_2 * \left(\frac{(P_{u,M1} + P_l)}{R_{u,1M}} - \frac{P_{u,1}}{R_{u,12}}\right)^{-1}\right)$
SRAA	(a)	**(i)** $Q_{d,1} > R_{d,M1} * \left(\frac{P_S}{P_l}\right)$
		(ii) $Q_{d,1} \leq N_{rf,1} * R_{d,11}$
		(iii) $\beta_{d,11} < \beta_{d,21}$
	(b)	**(i)** $Q_{u,1} < min\Bigg((N_{RF,M} * R_{u,1M}), \big((N_{rf,1} - C_1) * R_{u,11}\big), Q_{u,1} < P_1 * \left(\frac{P_l + P_{u,1}}{R_{u,M}} - \frac{P_{u,1}}{R_{u,11}}\right)^{-1}$
		(ii) $Q_{u,1} < min\left((N_{RF,M} * R_{u,1M}), \big((N_{rf,2} - C_2) * R_{u,12}\big), P_2 * \left(\frac{(P_{u,M1} + P_l)}{R_{u,1M}} - \frac{P_{u,1}}{R_{u,12}}\right)^{-1}\right)$
RMA	(a)	**(i)** $Q_{d,1} > P_S * \left(c_{M1} * \left(\frac{P_l}{R_{d,M1}} + P_{fhl}\right) - c_{S1} * P_{fhl}\right)^{-1}$
		(ii) $Q_{d,1} \leq N_{rf,1} * R_{d,11}$
		(iii) $c_{11} * \beta_{d,11} < c_{21} * \beta_{d,21}$
	(b)	**(i)** $Q_{u,1} < min\Bigg((N_{RF,M} * R_{u,1M}), \big((N_{rf,1} - C_1) * R_{u,11}\big), P_1 * \left(\left(\frac{P_l + P_{u,1}}{R_{u,1M}} + P_{fhl}\right) * c_{M1} - P_{fhl} * c_{11} - \frac{P_{u,1} * c_{11}}{R_{u,11}}\right)^{-1}$
		(ii) $Q_{u,1} < min\Bigg((N_{RF,M} * R_{u,1M}), \big((N_{rf,2} - C_2) * R_{u,12}\big), P_2 * \left(\left(\frac{P_l + P_{u,1}}{R_{u,1M}} + P_{fhl}\right) * c_{M1} - P_{fhl} * c_{21} - \frac{P_{u,1} * c_{21}}{R_{u,12}}\right)^{-1}$

Table 1: 1)Association to SBS 1 in DL and MBS in UL

Figures 5a, 5b and 5c illustrate the similarities and differences of the optimal algorithm, SRAA and RMA for nine different cases. The optimal algorithm and SRAA tend to have same data rate demand threshold values for association to same BSs in UL and DL. On the other hand, since the



(a) UL Data Rate Thresholds (b) DL Data Rate Thresholds (c) Comparison of Association Cases

Figure 5: Algorithm Analysis

power consumption of each association in RMA is scaled, data rate demands for association with the same BS in UL and DL are less than the optimal algorithm and SRAA for all cases. Figures 5a and 5b demonstrate that the optimal algorithm and SRAA have same network topologies for same data rate demands both in UL and DL but RMA assigns the UE to the same corresponding BS with less data rate demand threshold value. In case of increase in data rate demand either in DL or UL, the frequency of the topology change is high for RMA compared to the optimal algorithm and SRAA.

Figure 5c illustrates the superiority of the optimal algorithm and SRAA over RMA in terms of minimizing the total power consumption. For the first four cases, the dynamic power consumption of MBS causes increase in the system power consumption and the association of UE to MBS is independent of UL or DL connection to MBS. The underlying difference between the power consumption values of the optimal algorithm, SRAA and RMA is that the power consumption values of each association pair in RMA are scaled with c_{Mj} and c_{ij}. However, in cases five, six, eight and nine (UE association to SBSs in UL and DL), the effect of c_{Mj} is eliminated and all of the algorithms consume the same power.

5. Simulations and Performance Evaluation

The goal of this section is to evaluate the performance of the SRAA for various data rate requirements, network topologies and weighting value w compared to previously proposed heuristic algorithms including RMA [36] and PMA [37]; demonstrate the importance of the incorporation of the dynamic power consumption of MBS and the FH link power consumption and decoupling UL and DL in UE association on the optimal power consumption. Besides RMA explained in Subsection 4.2, performance of SRAA is also compared to PMA, which is a heuristic algorithm previously proposed for matching problem [37]. This algorithm is based on using random walk to form minimum power consuming UE-BS pairs. Heuristic starts by associating a random UE with a random BS. Then, new association pairs are found by performing random walk based on the current matching. The random walk G, is updated in each turn based on the result of the previously realized random walk. The G is updated as follows: A random UE selects a random BS according to a Markov transition matrix, M. M is obtained by normalizing $\beta_{i,j}$ values with $N_{RF,i}$ values, $\left(\frac{\beta_{i,j}}{N_{RF,i}}\right)$, of the elements of the current matching. Each UE aims to associate with the minimum power consuming BS. This process is repeated for all UEs until each UE is assigned to the same BS, which it has been associated with in the r^{th} turn. The complexity of the algorithm is $\mathcal{O}(mn)$, where n is the number of BSs and m is the number of UEs.

Simulations are performed by using IBM CPLEX Optimization Studio with MathWorks MATLAB. Simulation results are obtained and averaged based on 5000 random configurations where SBSs and UEs are distributed randomly and uniformly within the circular area of radius within [50,700]m, with the additional restriction of minimum distance of 10m among SBSs.

Uniform random single-path (UR-SP) Channel Model in [41] is adopted capturing the dominant path in directional propagation environments. The links are assumed to be in Line-of-Sight (LoS), where no obstacles reside between transmit and receive antennas, or in Non-Line-Of-Sight ($NLoS$), where partial or full obstructions exist between the transmit and receive antennas, or in outage and we decide on the state randomly. The attenuation of the links are determined considering large scale statistics that arise primarily from the free space loss and the environment affecting the degree of refraction, diffraction, reflection and absorption. The dependence of the path loss on distance summarizing large scale statistics is modeled as $\alpha_{ij} = PL(d_0) + 10 * \eta_j * \log_{10}\left(\frac{d_{du,ij}}{d_0}\right) + X_{\sigma_j}$ in DL and as $\theta_{ij} = PL(d_0) + 10 * \eta_j * \log_{10}\left(\frac{d_{du,ij}}{d_0}\right) + X_{\sigma_j}$ in UL, where $PL(d_0)$ is the free space path loss in dB at a reference distance d_0, η_j is the path loss exponent, $d_{du,ij}$ is the distance between BS i and UE j and X_{σ_j} is a zero mean Gaussian random variable with σ_j^2 variance. However, path loss is modeled in LOS and NLOS scenarios differently by making use of different η_j and σ_j values in 28 GHz as demonstrated in [39]. We investigate the network with two different data rate demand ranges which are considered to be low and high as compared to maximum achievable data rates in LTE (300 Mbps in DL and 75 Mbps in UL) [52]. The internal hardware power consumption parameters of MBS and SBS are stated in [44] and FH link power consumption is stated in [46]. The simulation parameters follow Table 2.

5.1. Performance Comparison of Different Algorithms

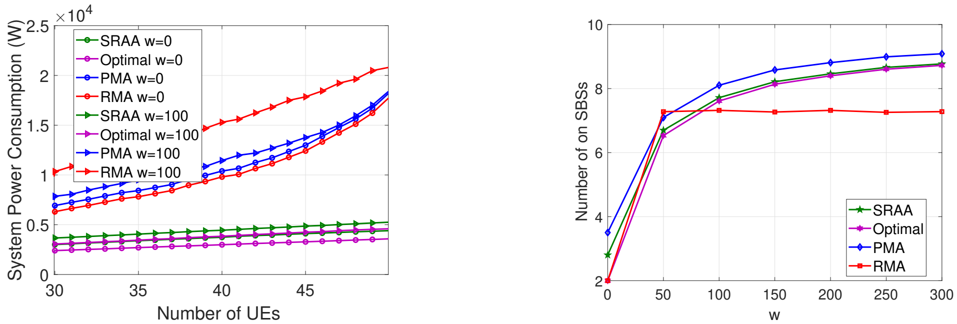
Figure 6a shows the system power consumption of different algorithms for increasing number of UEs. For w=0, as number of UEs increases, the power consumption of all algorithms increases. The SRAA performs very close to optimal, significantly better than previously proposed algorithms due to controlled on/off operation of SBSs with the repeated use of modified matching heuristic and GAP heuristic. The use of modified matching heuristic with the GAP heuristic reduces UE TX power by

Carrier Frequency	28 GHz SBSs 2.4 GHZ for MBS
Bandwidth	800 MHz for 28 GHz [39] 20 MHz for 2.4 GHz
PSD of Noise at each UE	-134 dBm/MHz [43]
Downlink tx power	MBS: 43 dBm SBS: 33 dBm UE : 23 dBm
BS antenna height	7 m [39]
SBS antenna height	2 m [39]
UE antenna height	1.5 m [39]
Number of RF Chains	MBS: 8 RF Chains SBS: 4 RF Chains
Path loss Model	ITU Urban Macro LOS-NLOS
Fading Model	σ_j MBS : 8.9 dB σ_j SBS-LOS : 1.1 dB [39] σ_j NLOS : 10 dB [39]
Path loss Exponent	η_j MBS : 2 η_j SBS-LOS : 1.9 [39] η_j SBS-NLOS : 4.5 [39]
Capacity of FH Link	4.45 Gbps
DL Data Demand (Low)	U(0.1-0.5)Mbps
DL Data Demand (High)	U(0.15-2)Gbps
UL Data Demand (Low)	U(0.01-0.03)Mbps
UL Data Demand (High)	U(0.2-0.5)Gbps

Table 2: System Parameters

finding optimal BSs for the minimum power consumption. Additionally, though high number of on SBSs provides decrease in UE TX power and the UL power consumption, the power consumption of PMA increases due to increase in the operational cost of on SBSs and exceeds the power consumption of RMA. Moreover, as w increases, the power consumption values of all algorithms increase and RMA has an unexpected increase in the total power consumption due to constant number of on SBSs as in Figure 6b, which causes increase in UE TX power.

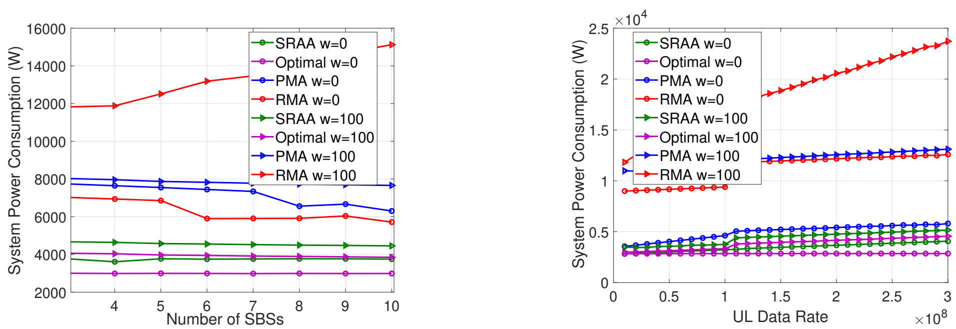
Figure 7a demonstrates the system power consumption of different algorithms for different number of SBSs.For 40 UEs, as the number of SBSs increases, the probability of selecting the minimum power consuming BSs increases, resulting in decrease in total power consumption for all algorithms. The SRAA demonstrates significant improvement compared to RMA and PMA. It minimizes the total power consumption by selecting the SBSs, which provide the most decrease in UE TX power and turning off the unused SBSs to further save operational cost of SBSs. Moreover, as w increases the power consumption of algorithms decreases except in RMA the power consumption increases due to low number of on SBSs as shown in Figure 6b. Fig. 7b illustrates the system power consumption of different algorithms for increasing UL data rate demand. The increase in UL data rate leads to an increase in the channel utilization from SBS S to UE j, $\beta_{u,jS}$, values and in the UL power consumption. For this purpose, the algorithms turn on additional SBSs to increase the probability of selecting the



(a) System power consumption of different algorithms for different number of UEs and w values

(b) Number of on SBSs of different algorithms for different w

Figure 6: Power Consumption Analysis and Number of on SBSs



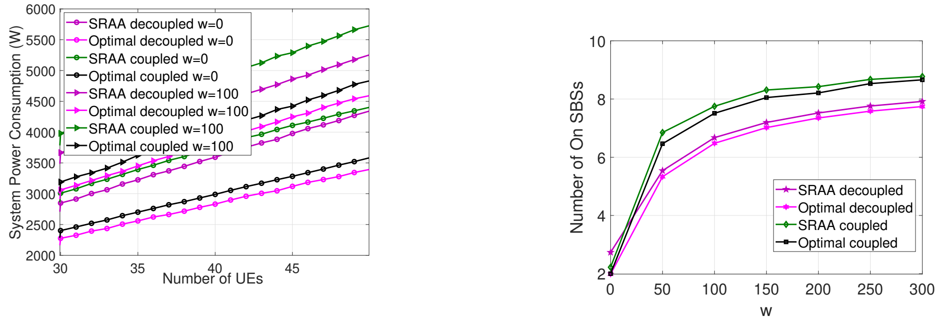
(a) System power consumption of different algorithms for different number of SBSs and w values

(b) System power consumption of different algorithms for different UL data rates and w values

Figure 7: Power Consumption Analysis

minimum power consuming BSs and decreasing $\beta_{u,jS}$, values and the UL power consumption. For w=0 and increasing UL data rate demand, the system power consumption increases for all algorithms and SRAA achieves significant improvements compared to PMA and RMA with the help of the controlled on/off operation of SBSs to decrease UE TX power. Moreover, as w increases, the power consumption of all algorithms increase noticeably because high number of on SBSs provides decrease in the UE power consumption but the static power consumption of on SBSs surpasses the power consumption of UEs in UL. At the same time, the number of on SBSs increases to compensate between the UL power consumption and the power consumption of on SBSs. However, the power consumption of RMA increases due to low number of on SBSs, which are not sufficient enough to compensate the power consumed in UL and the power consumption of on SBSs and cause increase in the UE TX power.

Figure 8a demonstrates the gain obtained from decoupling UL/DL for the optimal algorithm and the SRAA for different number of UEs. As the number of UEs increases, the system power



(a) System power consumption of coupled and decoupled association schemes of different algorithms for different number of UEs and w values

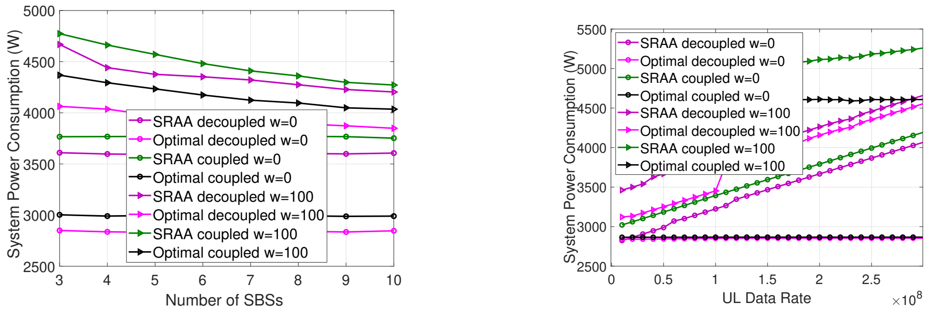
(b) Number of on SBSs of coupled and decoupled UE association schemes for different w

Figure 8: Power Consumption Analysis and number of on SBSs

consumption increases for all algorithms. The SRAA implemented for both coupled and decoupled UE association perform close to corresponding the optimal algorithms with approximation ratio less than 1.2, where approximation ratio is defined as ratio between the objective function values of proposed heuristic algorithm and the the optimal algorithm. Figure 8b illustrates number of on SBSs for different number of w values. As w value increases, the UL power consumption increases and the system turns on additional SBSs to further minimize the total power consumption by decreasing UE TX power.

Figure 9a compares the power consumption of the optimal algorithm and the SRAA for decoupled and coupled UL/DL UE association for different number of SBSs. The power consumption decreases as the number of SBSs increases mainly due to increasing number of UL and DL choices for UEs. UEs select the least power consuming SBS for the communication initiation. Moreover, as w increases, UL power consumption increases and additional SBSs are turned on to further minimize the total power consumption by decreasing UE TX power. On the other hand, decoupled UE association scheme outperforms coupled UE association scheme in terms of the power consumption. Coupled UE association causes UEs to transmit with high TX power as a result of the association to same BSs both in UL and DL though they can have different SINR values in UL and DL.

Figure 9b illustrates the effect of UL data rate on the system power consumption of the optimal algorithm and SRAA for decoupled and coupled UL/DL UE association. For increasing UL data rate, the power consumption of the heuristic algorithm with coupled and decoupled UE association is very close to that of the corresponding optimal algorithms with approximation ratio less than 1.2. Additionally, the increase in UL data rate causes an increase in the channel utilization values from MBS and SBS S to UE j in UL $\beta_{u,jM}$ and $\beta_{u,jS}$, respectively and in the UL power consumption of UEs. To decrease the effect of the increase in UE power consumption in UL, the algorithm turns on



(a) System power consumption of coupled and decoupled association schemes of different algorithms for different number of SBSs and w values

(b) System power consumption of coupled and decoupled association schemes of different algorithms for different UL data rate and w values

Figure 9: Power Consumption Analysis

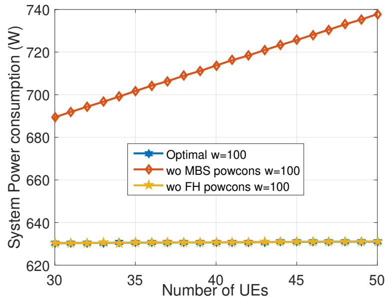
additional SBSs to reduce the TX power consumption of UEs as shown in Figure 8b.

5.2. *Effect of incorporating realistic power consumption models*

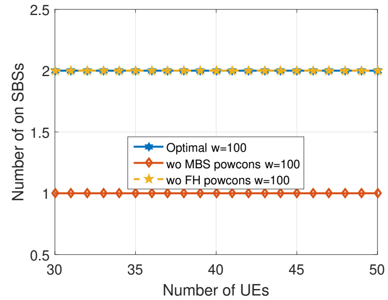
We demonstrate the importance of the incorporation of the dynamic power consumption of MBS and FH link power consumption on the optimal power consumption by investigating the network with the specified data rate demands both in UL and DL as in Table 2 for w=100.

5.2.1. *Low Data Rate Demand*

Figure 10a demonstrates the effect of the number of UEs on the system power consumption of the optimal algorithm, the optimal algorithm without the dynamic power consumption of MBS and the optimal algorithm without FH link power consumption for 10 SBSs. For increasing number of UEs, the optimal algorithm without the dynamic power consumption of MBS leads to higher power consumption than the power consumption of optimal algorithm and the power consumption of the optimal algorithm without FH link power consumption. Since the dynamic power consumption of MBS does not have an impact on increasing the power consumption in DL, the algorithm assigns all UEs to MBS in DL as long as it provides capacity to UEs and association to MBS causes increase in total power consumption. On the other hand, low data rate demands provide low $\beta_{u,jM}$ and $\beta_{u,jS}$ values in UL, where $\beta_{u,jM} \geq \beta_{u,jS}$, causing low power consumption on UL side. Therefore, the optimal algorithm without the dynamic power consumption of MBS assigns UEs to SBSs in UL as in Figure 11a. Additionally, the optimal algorithm and the optimal algorithm without FH link power consumption decide in favor of SBS in DL for UE association because in presence of low data rate demanding UEs, the power consumption of UE-SBS association in DL, which is the static power consumption of SBS, is less than the power consumption of UE-MBS association. Moreover, with the

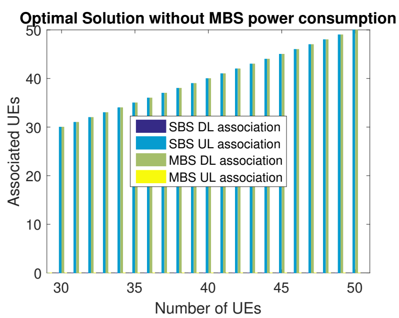


(a) Effect of the dynamic MBS power consumption on system power consumption for low data rate demands in DL

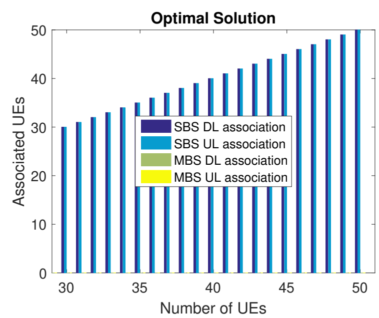


(b) Effect of the dynamic MBS power consumption on the number of on SBSs for low data rate demands in DL

Figure 10: Power Consumption and Number of on SBSs in a low data rate demanding scenario for increasing number of UEs



(a) Effect of the dynamic MBS power consumption on UE association pattern for low data rate demands in DL

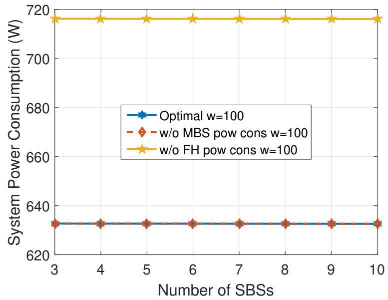


(b) UE Association for optimal algorithm for low data rate demands in DL

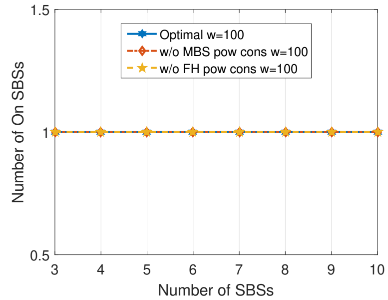
Figure 11: UE association in a low data rate demanding scenario for increasing number of UEs

same reason stated above for the optimal algorithm without the dynamic power consumption of MBS, on UL side, the power consumption of UE-MBS association is greater than the power consumption of UE-SBS association, UEs get also assigned to a SBS in UL as seen in Figure 11b and two SBSs are on as in Figure 10b.

Figure 12a represents the impact of the number of the SBSs on the system power consumption of the optimal algorithm, the optimal algorithm without the dynamic power consumption of MBS and the optimal algorithm without FH link power consumption for 40 UEs. For increasing number of SBSs, the power consumption of all algorithms decreases due to increasing number of UL and DL choices, resulting in decrease in UE TX power. Additionally, total power consumption of the optimal algorithm and the optimal algorithm without the dynamic power consumption of MBS outperform the optimal algorithm without FH link power consumption though the number of on SBSs is constant

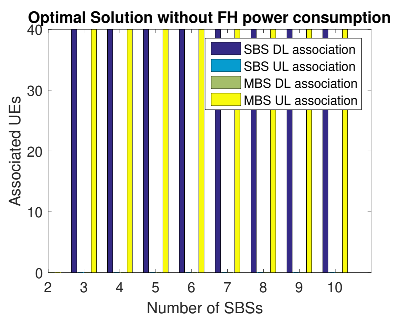


(a) Effect of FH link power consumption on system power consumption for low data rate demands in DL

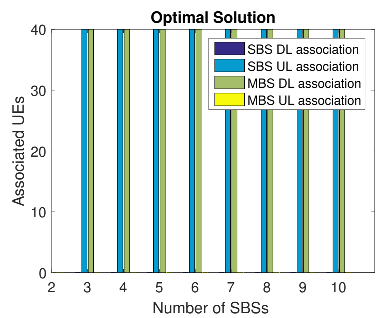


(b) Effect of FH link power consumption on the number of on SBSs for low data rate demands in DL

Figure 12: Power Consumption and Number of on SBSs in a low data rate demanding scenario for increasing number of SBSs



(a) Effect of FH link power consumption on UE association pattern for low data rate demands in DL



(b) UE Association for optimal algorithm for low data rate demands in DL

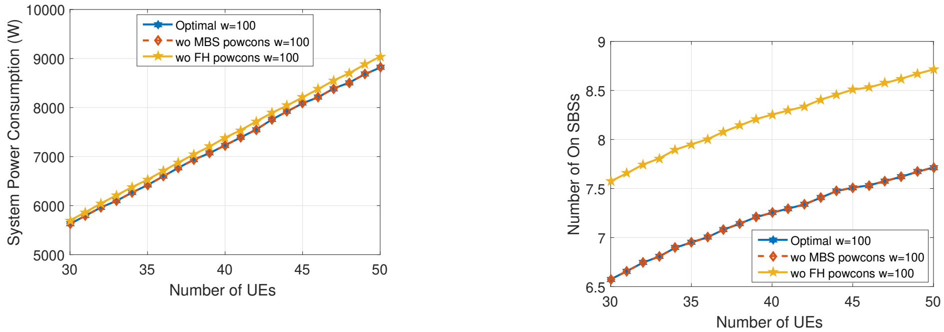
Figure 13: UE association in a low data rate demanding scenario for increasing number of SBS

and equal for all algorithms as shown in Figure 12b. The underlying reason is the difference between UE-BS association patterns as demonstrated in Figures 13a and 13b.

The optimal algorithm without FH link power consumption assigns UEs to SBSs in DL in Figure 13a, since the static power consumption of SBSs is less than the dynamic power consumption of MBS according to the DL power consumption part in objective function (1a). On the other hand, the optimal algorithm without FH link power consumption assigns UEs to MBS in UL due to only one on SBS as in Figure 12b and insufficient capacity of the corresponding on SBS.

5.2.2. High Data Rate Demand

Figure 14a represents the impact of increasing the number of UEs on the system power consumption of the optimal algorithm, the optimal algorithm without the dynamic power consumption of MBS and the optimal algorithm without FH link power consumption for 10 SBSs. For increasing number of UEs,

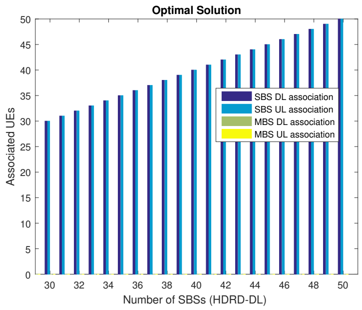


(a) Effect of the dynamic MBS power consumption on the system power consumption for high data rate demands in DL

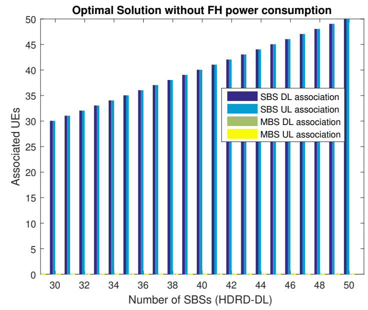
(b) Effect of the dynamic MBS power consumption on on SBS for high data rate demands in DL

Figure 14: Power Consumption and Number of on SBSs in a high data rate demanding scenario for increasing number of UEs

the power consumption of the optimal algorithm with and without the dynamic power consumption of MBS are equal to each other and increase linearly. They outperform the optimal algorithm without FH link power consumption due to the selection of optimal number and optimal location of SBSs. In the optimal algorithm without the dynamic power consumption of MBS, exclusion of the dynamic power consumption of MBS does not lead to the assigment of the UEs to MBS in DL, since $\beta_{d,Mj}$ values are so high such that the capacity requirements are not fulfilled. Therefore, the UEs associate with SBS in DL. The power consumption of DL association to SBS is dependent on power consumed on the FH links which is formulated as a function of data rate. As the number of UEs associated to SBS in DL increases and the number of on SBSs increases as in Figure 14b, the power consumption also increases. The dynamic power consumption also affects UL association to MBS. Sum of $\beta_{u,jM}$ values for high UL data rate demands exceeds the capacity of MBS, therefore UEs associate with different SBSs in UL than the SBSs in DL as shown in Figure 15a. Besides, the optimal algorithm also assigns UEs to SBSs in UL and DL due to high $\beta_{d,Mj}$ and $\beta_{u,jM}$ values. The power consumption of the optimal algorithm without FH link power consumption demonstrated in Figure 14a increases linearly as the number of UEs increases. In case of increase in the number of UEs, the number of on SBSs increases to provide more capacity to UEs as in Figure 14b. Furthermore, the capacity of MBS is not sufficient to provide service to high data rate demanding UEs both in UL and DL with high $\beta_{d,Mj}$ and $\beta_{u,jM}$. Therefore, UEs associate with different SBSs in UL and DL as illustrated in Figure 15b. Figure 16a demonstrates the effect of number of SBSs on the system power consumption of the optimal algorithm, the optimal algorithm without the dynamic power consumption of MBS and the optimal algorithm without FH link power consumption for 40 UEs. For increasing number of SBSs, the power consumption of the optimal algorithm with and without the dynamic power consumption of MBS

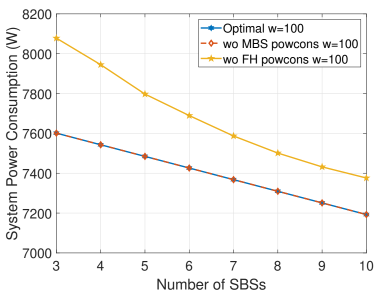


(a) Effect of the dynamic MBS power consumption on UE association pattern for high data rate demands in DL

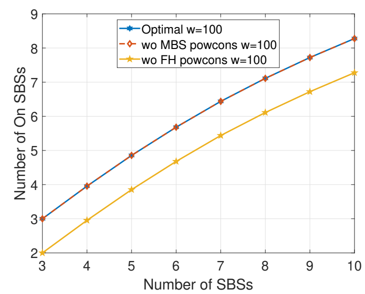


(b) UE Association for optimal algorithm for high data rate demands in DL

Figure 15: UE association in a high data rate demanding scenario for increasing number of UEs



(a) Effect of FH link power consumption on the system power consumption for high data rate demands in DL

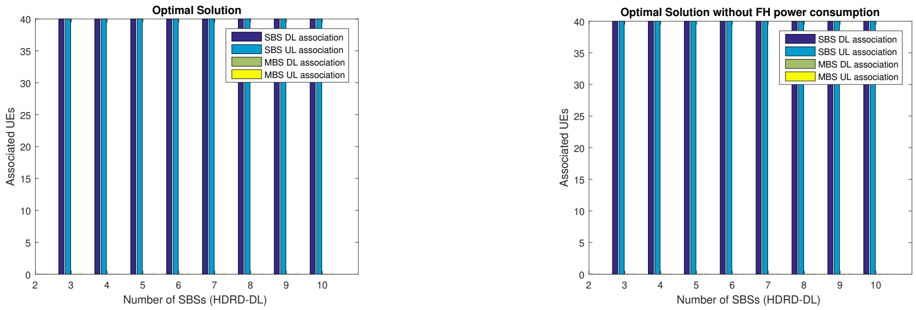


(b) Effect of FH link power consumption on the number of on SBSs for high data rate demands in DL

Figure 16: Power Consumption and Number of on SBSs in a high data rate demanding scenario for increasing number of SBSs

outperform the power consumption of the optimal algorithm without FH link power consumption. The uncontrolled increase in the static power consumption of SBSs and the negligible effect of the reduction in UE power consumption causes increase in the total power consumption of the optimal algorithm without FH link power consumption. Also, the power consumption values and association graphs of the optimal algorithm and the optimal algorithm without the dynamic power consumption of MBS are equal to each other as illustrated in Figure 16a. The power consumption values decrease unlike the decrease as in Figure 12a due to high data rate demand of UEs that forces the system to turn on more SBS to fulfill high capacity requirements. Moreover, the number of on SBSs increases with increasing number of SBSs. Increasing number of SBSs enables UEs to select different SBSs both in UL and DL, which reduce their power consumption as in Figure 16b.

As seen in Figure 17a and in Figure 17b, both algorithms assign UEs to SBSs both in UL and DL. Additionally, high data rate demands both in UL and DL generate large $\beta_{u,jM}$ and $\beta_{d,Mj}$ values



(a) UE Association for optimal algorithm for high data rate demands in DL

(b) Effect of FH link power consumption on UE association pattern for high data rate demands in DL

Figure 17: UE association in a high data rate demanding scenario for increasing number of SBSs

which exceed capacity limits of MBS. Therefore the system does not allow UEs to associate with MBS. As the number of SBSs increases, $\beta_{u,jM}$ and $\beta_{d,Mj}$ values are constant but the algorithms have the opportunity to select smaller $\beta_{u,jS}$ values from the on SBS set. For this reason, the UEs get associated with SBSs in UL and DL due to insufficient capacity of MBS.

5.3. Delay and Fairness Performance

We demonstrate the delay and fairness performance of different algorithms by investigating the network with the specified data rate demands both in UL and DL as in Table 2 for w=100.

5.3.1. Delay Performance

Figure 18 compares the end-to-end delay of different algorithms for different number of UEs. In HCRAN, end-to-end delay components comprise of BBU Pool processing delay [53], BBU Pool-BS transmission delay on FH link [54], BS on/off delay for control plane and data plane signaling [55], BS-UE wireless communication delay [54] and waiting time of UEs in the BBU Pool processing patch [56]. BBU Pool processing delay determines the delay performance of a protocol since the rest of the delay components are the same for different algorithms. BBU Pool processing delay increases as the number of UEs increases mainly due to long period of searching for the least power consuming UE-BS pairs, which enables UEs to transmit with less transmit power. SRAA performs very close to optimal, significantly better than previously proposed algorithms, repeated matching algorithm and perfect matching algorithm, since SRAA associates UEs with pre-selected on BSs and offloads UEs among them instead of turning on each BS and trying each candidate BS for association. However, selection of random BSs according to a Markov transition matrix by random UEs without constraining the number of on BSs in perfect matching algorithm exceeds the BBU Pool processing delay experienced by SRAA whereas BBU Pool processing delay experienced in repeated matching algorithm exceeds

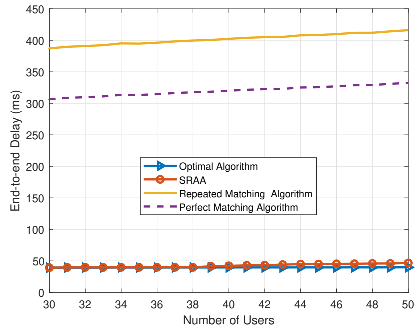


Figure 18: End-to-End Delay experienced in algorithms for increasing number of UEs

that experienced by the rest of the algorithms since brute force nature of the algorithm turns on each BS and forces UEs to get assigned with each candidate BS, repeatedly.

5.3.2. Fairness Performance

In the literature, Fairness index measures the equality of resource allocation among UEs and if not equal, it represents how far the resource allocation is from equality. It is defined as Jain's Formula [57] and [58], $Jain = \frac{\left(\sum_{i=1}^{N} \kappa_i\right)^2}{N*\left(\sum_{i=1}^{N} \kappa_i^2\right)}$, where κ_i represents a resource allocation that UE i receives in case when a network allocates resources to N UEs. However, in our case Jain's index corresponds to the satisfaction of the requirements of UEs, κ_j represents the satisfaction of the data rate requirement of UE j and Jain's index is formulated as $Jain = \frac{\left(\sum_{j=1}^{N_b} \kappa_j\right)^2}{N_b*\left(\sum_{j=1}^{N_b} \kappa_j^2\right)}$. If the rate provided to UE j by BS i is greater than demanded data rate, κ_j will be 1, indicating that the requirements of UE j is fulfilled, and UE j is associated with BS i. It is formulated as follows:

$$\kappa_j = \begin{cases} 1, & \text{if } \beta_{i,j} \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

Jains index equals to 1 if the optimization problem is feasible otherwise equals to zero.

Figure 19 demonstrates the fairness index of different algorithms for different number of UEs. SRAA outperforms previously proposed heuristic algorithms in terms of fairness because it allocates the resources to all UEs efficiently, such that each UE gets the corresponding demanding data rate. However, repeated matching algorithm cannot provide fair utilization in a pre-determined iteration number though all BSs are on. The reason is the selection of any UE-BS pair at the beginning and improvement of the feasibility of the pair by choosing another BS for the UE and testing the feasibility repeatedly until feasible and fair allocation is done. Last, perfect matching algorithm does not support fair allocation due to random formation of UE-BS pairs all along the algorithm which causes inefficient

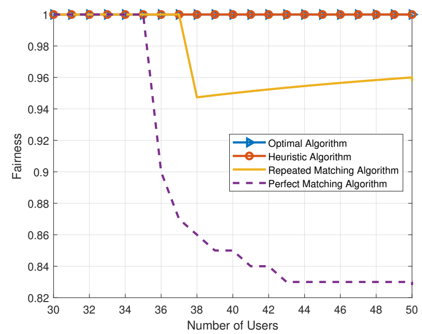


Figure 19: Jain's Index of the algorithms for increasing number of UEs

use of network resources for all UE-BS pairs in terms of providing demanding data rates.

6. Conclusion

In this paper, we present a holistic framework for energy efficient UE association incorporating UL/DL decoupling and the dynamic on/off scheme for SBSs and realistic power consumption models, including the static and the dynamic power consumption of MBS, the static power consumption of SBS, power consumption of the FH links as a function of data rate and the power consumption of UEs in UL. We formulate the optimization problem as a SSCFLP. We propose a heuristic algorithm that minimizes the total power consumption by improving the output of the primal heuristic composed of LP-Relaxation and rounding process with the help of repeated use of two heuristic algorithms, heuristic for modified matching and heuristic for re-allocation of UEs. As a result of LP relaxation and rounding, we obtain a set of assigned UE-BS pairs, unassigned UEs and off BSs as the input to the proceeding heuristic algorithms in contrast to RMA and PMA, that take all UEs as unassigned UEs and all BSs as off BSs as input. We demonstrate the superiority of the SRAA to previously proposed solution methods in terms of closeness to the optimality for different networks topologies and data rate requirements via extensive simulations. We also demonstrate that for decoupled UL/DL UE association scheme with SBSs operating on mmWave frequencies providing high bandwidth, high data rate demanding UEs get associated with SBSs in UL and DL, whereas low data rate demanding UEs get associated with SBSs in UL and MBS in DL.

As future work, we aim to concentrate on multi-connectivity in decoupled UL/DL UE association scheme in H-CRAN with the goal of minimizing the total network power consumption. We also aim to concentrate on minimizing the number of handovers in energy constrained UL/DL decoupled UE Association in H-CRAN and investigate the effect of multi-connectivity on decreasing the handover event. Last, we plan to investigate the effect of offloading UE traffic from the Cloud and BSs with the help of device-to-device communication on the total system power consumption.

References

[1] S. Rangan, T. S. Rappaport, E. Erkip, Millimeter-wave cellular wireless networks: Potentials and challenges, Proceedings of the IEEE 102 (3) (2014) 366–385 (2014). doi:10.1109/JPROC.2014.2299397.

[2] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, J. C. Zhang, What will 5g be?, IEEE Journal on Selected Areas in Communications 32 (6) (2014) 1065–1082 (2014). doi:10.1109/JSAC.2014.2328098.

[3] M. Peng, Y. Li, Z. Zhao, C. Wang, System architecture and key technologies for 5g heterogeneous cloud radio access networks, IEEE Network 29 (2) (2015) 6–14 (2015).

[4] A. Zakrzewska, V. B. Iversen, Resource sharing in heterogeneous and cloud radio access networks, in: 2012 IV International Congress on Ultra Modern Telecommunications and Control Systems, 2012, pp. 34–39 (2012).

[5] M. Peng, K. Zhang, J. Jiang, J. Wang, W. Wang, Energy-efficient resource assignment and power allocation in heterogeneous cloud radio access networks, IEEE Transactions on Vehicular Technology 64 (11) (2015) 5275–5287 (2015).

[6] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, P. Popovski, Five disruptive technology directions for 5g, IEEE Communications Magazine 52 (2) (2014) 74–80 (2014).

[7] M. Peng, Y. Li, J. Jiang, J. Li, C. Wang, Heterogeneous cloud radio access networks: a new perspective for enhancing spectral and energy efficiencies, IEEE Wireless Communications 21 (6) (2014) 126–135 (2014).

[8] P. Rost, G. Fettweis, Green communications in cellular networks with fixed relay nodes, Cooperative Cellular Wireless Networks (2011) 300 (2011).

[9] S. Bhaumik, G. Narlikar, S. Chattopadhyay, S. Kanugovi0, Breathe to stay cool: adjusting cell sizes to reduce energy consumption, in: Proceedings of the first ACM SIGCOMM workshop on Green networking, ACM, 2010, pp. 41–46 (2010).

[10] D. Brubaker, Optimizing performance and efficiency of pas in wireless base stations, White Paper, Texas Instruments (2009) 1 (2009).

[11] V. Mancuso, S. Alouf, Reducing costs and pollution in cellular networks, IEEE Communications Magazine 49 (8) (2011) 63–71 (2011).

[12] R. Mahapatra, Y. Nijsure, G. Kaddoum, N. Ul Hassan, C. Yuen, Energy efficiency tradeoff mechanism towards wireless green communication: A survey, IEEE Communications Surveys Tutorials 18 (1) (2016) 686–705 (Firstquarter 2016). doi:10.1109/COMST.2015.2490540.

[13] C. Liu, B. Natarajan, H. Xia, Small cell base station sleep strategies for energy efficiency, IEEE Transactions on Vehicular Technology 65 (3) (2016) 1652–1661 (2016).

[14] K. Son, E. Oh, B. Krishnamachari, Energy-efficient design of heterogeneous cellular networks from deployment to operation, Computer Networks 78 (2015) 95–106 (2015). doi:10.1016/j.comnet.2014.09.018.

[15] S. Zhou, J. Gong, Z. Yang, Z. Niu, P. Yang, Green mobile access network with dynamic base station energy saving, ACM MobiCom 9 (262) (2009) 10–12 (2009).

[16] N. Yu, Y. Miao, L. Mu, H. Du, H. Huang, X. Jia, Minimizing energy cost by dynamic switching on/off base stations in cellular networks, IEEE Transactions on Wireless Communications 15 (11) (2016) 7457–7469 (2016).

[17] S. Cai, Y. Che, L. Duan, J. Wang, S. Zhou, R. Zhang, Green 5g heterogeneous networks through dynamic small-cell operation, IEEE Journal on Selected Areas in Communications 34 (5) (2016) 1103–1115 (2016).

[18] Y. Shi, J. Zhang, K. B. Letaief, Group sparse beamforming for green cloud-ran, IEEE Transactions on Wireless Communications 13 (5) (2014) 2809–2823 (2014).

[19] P. Luong, F. Gagnon, C. Despins, L. Tran, Optimal joint remote radio head selection and beamforming design for limited fronthaul c-ran, IEEE Transactions on Signal Processing 65 (21) (2017) 5605–5620 (2017).

[20] W. Zhao, S. Wang, Traffic density-based rrh selection for power saving in c-ran, IEEE Journal on Selected Areas in Communications 34 (12) (2016) 3157–3167 (2016).

[21] C. Pan, H. Zhu, N. J. Gomes, J. Wang, Joint user selection and energy minimization for ultra-dense multi-channel c-ran with incomplete csi, IEEE Journal on Selected Areas in Communications 35 (8) (2017) 1809–1824 (2017).

[22] J. Zuo, J. Zhang, C. Yuen, W. Jiang, W. Luo, Energy efficient user association for cloud radio access networks, IEEE Access 4 (2016) 2429–2438 (2016). doi:10.1109/ACCESS.2016.2566338.

[23] H. Dahrouj, A. Douik, O. Dhifallah, T. Y. Al-Naffouri, M. Alouini, Resource allocation in heterogeneous cloud radio access networks: advances and challenges, IEEE Wireless Communications 22 (3) (2015) 66–73 (2015).

[24] Y. L. Lee, L. Wang, T. C. Chuah, J. Loo, Joint resource allocation and user association for heterogeneous cloud radio access networks, in: 2016 28th International Teletraffic Congress (ITC 28), Vol. 01, 2016, pp. 87–93 (2016).

[25] Y. Zhang, Y. Wang, W. Zhang, Energy efficient resource allocation for heterogeneous cloud radio access networks with user cooperation and qos guarantees, in: 2016 IEEE Wireless Communications and Networking Conference, 2016, pp. 1–6 (2016).

[26] J. G. Andrews, Seven ways that hetnets are a cellular paradigm shift, IEEE Communications Magazine 51 (3) (2013) 136–144 (2013).

[27] H. Elshaer, F. Boccardi, M. Dohler, R. Irmer, Downlink and uplink decoupling: A disruptive architectural design for 5g networks, in: 2014 IEEE Global Communications Conference, 2014, pp. 1798–1803 (2014).

[28] X. Sui, Z. Zhao, R. Li, H. Zhang, Energy efficiency analysis of heterogeneous cellular networks with downlink and uplink decoupling, in: 2015 IEEE Global Communications Conference (GLOBECOM), 2015, pp. 1–7 (2015).

[29] F. Boccardi, J. Andrews, H. Elshaer, M. Dohler, S. Parkvall, P. Popovski, S. Singh, Why to decouple the uplink and downlink in cellular networks and how to do it, IEEE Communications Magazine 54 (3) (2016) 110–117 (2016).

[30] H. Elshaer, M. N. Kulkarni, F. Boccardi, J. G. Andrews, M. Dohler, Downlink and uplink cell association with traditional macrocells and millimeter wave small cells, IEEE Transactions on Wireless Communications 15 (9) (2016) 6244–6258 (2016).

[31] M. Bacha, Y. Wu, B. Clerckx, Downlink and uplink decoupling in two-tier heterogeneous networks with multi- antenna base stations, IEEE Transactions on Wireless Communications 16 (5) (2017) 2760–2775 (2017).

[32] U. Bin Farooq, U. Sajid Hashmi, J. Qadir, A. Imran, A. N. Mian, User transmit power minimization through uplink resource allocation and user association in hetnets, in: 2018 IEEE Global Communications Conference (GLOBECOM), 2018, pp. 1–6 (2018).

[33] H. Elshaer, F. Boccardi, M. Dohler, R. Irmer, Load amp; backhaul aware decoupled downlink/uplink access in 5g systems, in: 2015 IEEE International Conference on Communications (ICC), 2015, pp. 5380–5385 (2015).

[34] S. Luo, R. Zhang, T. J. Lim, Downlink and uplink energy minimization through user association and beamforming in c-ran, IEEE Transactions on Wireless Communications 14 (1) (2015) 494–508 (2015).

[35] A. Alkhateeb, G. Leus, R. W. Heath, Limited feedback hybrid precoding for multi-user millimeter wave systems, IEEE Transactions on Wireless Communications 14 (11) (2015) 6481–6494 (2015).

[36] M. Rönnqvist, S. Tragantalerngsak, J. Holt, A repeated matching heuristic for the single-source capacitated facility location problem, European Journal of Operational Research 116 (1) (1999) 51–68 (1999). `doi:10.1016/S0377-2217(98)00045-9`.

[37] M. Kapralov, Algorithms for bipartite matching problems with connections to sparsification and streaming, Ph.D. thesis, Stanford University (2012).

[38] Y. Sano, A. Fukuda, S. Okuyama, Y. Kakishima, C. NA, Lte-advanced release 13 multiple antenna technologies and improved reception technologies, NTT Docomo Technical Jounal 18 (2) (2015) 62–71 (2015).

[39] T. S. Rappaport, G. R. MacCartney, M. K. Samimi, S. Sun, Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design, IEEE Transactions on Communications 63 (9) (2015) 3029–3056 (2015).

[40] Q. Xue, B. Li, X. Zuo, Z. Yan, M. Yang, Cell capacity for 5g cellular network with inter-beam interference, in: 2016 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), 2016, pp. 1–5 (2016).

[41] G. Lee, Y. Sung, J. Seo, Randomly-directional beamforming in millimeter-wave multiuser miso downlink, IEEE Transactions on Wireless Communications 15 (2) (2016) 1086–1100 (2016).

[42] S. Purnachand, R. Prasad, K. Sripath Roy, A. Suresh, Performance analysis of open loop and closed loop power control schemes for lte uplink, International Journal of Computer Science and Information Technologies 3 (2) (2012) 3683–3688 (2012).

[43] G. Athanasiou, P. C. Weeraddana, C. Fischione, L. Tassiulas, Optimizing client association for load balancing and fairness in millimeter-wave wireless networks, IEEE/ACM Transactions on Networking 23 (3) (2015) 836–850 (2015).

[44] M. Deruyck, W. Vereecken, W. Joseph, B. Lannoo, M. Pickavet, L. Martens, Reducing the power consumption in wireless access networks: Overview and recommendations, Progress In Electromagnetics Research 132 (2012) 255–274 (2012).

[45] S. Tombaz, P. Monti, K. Wang, A. Vastberg, M. Forzati, J. Zander, Impact of backhauling power consumption on the deployment of heterogeneous mobile networks, in: 2011 IEEE Global Telecommunications Conference - GLOBECOM 2011, 2011, pp. 1–5 (2011).

[46] R. Ramamonjison, A. Haghnegahdar, V. K. Bhargava, Joint optimization of clustering and cooperative beamforming in green cognitive wireless networks, IEEE Transactions on Wireless Communications 13 (2) (2014) 982–997 (2014).

[47] K. Smiljkovikj, L. Gavrilovska, P. Popovski, Efficiency analysis of downlink and uplink decoupling in heterogeneous networks, in: 2015 IEEE International Conference on Communication Workshop (ICCW), 2015, pp. 125–130 (2015).

[48] J. Krarup, P. M. Pruzan, The simple plant location problem: Survey and synthesis, European Journal of Operational Research 12 (1) (1983) 36 – 81 (1983). `doi:10.1016/0377-2217(83)90181-9`.

[49] L.-Y. Wu, X.-S. Zhang, J.-L. Zhang, Capacitated facility location problem with general setup cost, Computers and Operations Research 33 (5) (2006) 1226 – 1241 (2006). `doi:10.1016/j.cor.2004.09.012`.

[50] F. Chauvet, J.-M. Proth, A. Soumare, The simple and multiple job assignment problems, International Journal of Production Research 38 (14) (2000) 3165–3179 (2000). `doi:10.1080/002075400418207`.

[51] R. Cohen, L. Katzir, D. Raz, An efficient approximation for the generalized assignment problem, Information Processing Letters 100 (4) (2006) 162 – 166 (2006). `doi:10.1016/j.ipl.2006.06.003`.

[52] E. T. S. Institute, in: 3GPP TS 36.213 Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures, Vol. 14, 2017, pp. 1–456 (2017).

[53] S. Wang, C. Pan, C. Yin, Joint heterogeneous tasks offloading and resource allocation in mobile edge computing systems, in: 2018 10th International Conference on Wireless Communications and Signal Processing (WCSP), 2018, pp. 1–6 (Oct 2018). `doi:10.1109/WCSP.2018.8555725`.

[54] Q. Li, J. Lei, J. Lin, Min-max latency optimization for multiuser computation offloading in fog-radio access networks, in: 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018, pp. 3754–3758 (April 2018). `doi:10.1109/ICASSP.2018.8462297`.

[55] T. Blajić, D. Nogulić, M. Družijanić, Latency improvements in 3g long term evolution, in: Proc. Int. Conv. Inf. Commun. Technol., Electron. Microelectron., 2006, pp. 1–6 (2006).

[56] A. Arfaoui, S. Hamouda, L. Nuaymi, P. Godlewski, Minimization of delays in multi-service cloud-ran bbu pools, in: 2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC), 2017, pp. 1846–1850 (June 2017). doi:10.1109/IWCMC.2017.7986565.

[57] M. Amer, A. Busson, I. Guérin Lassous, Association optimization in wi-fi networks: Use of an access-based fairness, in: Proceedings of the 19th ACM international conference on modeling, analysis and simulation of wireless and mobile systems, ACM, 2016, pp. 119–126 (2016).

[58] R. K. Jain, D.-M. W. Chiu, W. R. Hawe, A quantitative measure of fairness and discrimination, Eastern Research Laboratory, Digital Equipment Corporation, Hudson, MA (1984).